University of London

THEORETICAL INVESTIGATIONS PROMPTED BY EXPERIMENTS WITH BAROCLINIC FLUIDS

Michael James Bell

Ph. D. Thesis

Imperial College, London

and

U. K. Meteorological Office, Bracknell

April 1989

ABSTRACT

The main part of this thesis presents analytical and numerical results which illuminate some aspects of the axisymmetric and regular wave regimes of laboratory experiments with rotating fluid annuli. The second part examines a method for the analysis of measurements of irregular waves.

The stability transition for simple baroclinic zonal flows with vertical and lateral shears and potential vorticity distributions similar to those of flows in differentially heated rotating annulus experiments is considered using the quasi-geostrophic equations. Neutral wave modes marking the transition are identified and the dependence of the transition on various features of the zonal flow investigated. The sensitivity of the transition's location to certain forms of lateral curvature is interpreted by an anti-cascade argument. Singular neutral modes found on laterally uniform flows subject to Ekman pumping and some very short wavelength trapped instabilities with similarities to tropospheric polar lows are also described.

The weakly non-linear development of waves near the transition is argued to depend crucially on the zonal flow's lateral shear. A model of a wave's development on a laterally sheared flow, involving a non-linear critical layer, is presented.

Numerical calculations of the stability of Rossby waves and other free modes in an f-plane channel show that waves of several azimuthal wavenumbers can be stable if the channel is long and narrow.

The justification for using the method of delays to reconstruct phase portraits of irregular low dimensional flows from experimental time series data is examined and the robustness and sensitivity of the method to various choices discussed.

ACKNOWLEDGEMENTS

This work was undertaken through the Public Research Institute Scheme with the co-operation and financial assistance of the U.K. Meteorological Office. I am particularly grateful to Raymond Hide for creating the opportunity for me to do this work.

I gratefully acknowledge the support, encouragement and stimulation I have received from many colleagues: Andy White's generosity with interesting suggestions, his critical supervision and balanced judgement have been particularly valuable; John Gibbon's and Raymond Hide's ready attention, encouragement and advice have also been greatly appreciated; I recognize too the value of the stimulating discussions I have had with Peter Read and Doug Johnson. I should also like to thank David Broomhead and David Chillingworth for their time, interest and advice, Stephen Cowley for drawing my attention to studies of non-linear critical layers and Pat Roberts for computational support.

Finally I thank Amelia, my wife, for both encouraging my efforts with this thesis and ensuring that they were occasionally relegated to second place.

To satisfy the regulations of the University, my supervisors hereby declare that I am solely responsible for the writing of this thesis and the new results which it contains

J. D. Gibbon

A. A. White.

CONTENTS LIST

Note

Equations are numbered according to their section. For example, the 11th numbered equation in section 3.2 is labelled (3.2.11). Within section 3.2 this equation is referred to simply as (11) whilst outside section 3.2 it is referred to explicitly as (3.2.11).

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

The weather over the U.K. is dominated by the cyclones which develop in the westerly winds over the N. Atlantic. These and similar cyclones in other longitudes play major roles in the transport of heat from the equator to the poles and in the maintenance of the mid-latitude westerly winds (Jeffreys 1926, Lorenz 1967). Their representation is consequently of primary importance both in weather forecasting and climate simulation.

It is fortunate that the essential dynamics of cyclones can be reproduced in the laboratory and studied there under a wide range of controlled conditions. This thesis is concerned with the interpretation of some of the most fundamental and well known results from the laboratory experiments. The following elementary accounts of the experiments and of the basic dynamics of their flows lead onto a summary of the thesis.

Laboratory Experiments

Each experiment is concerned with the motions of a fluid (e.g. water) in a rotating container. The density of the fluid is arranged to vary within the container either by maintaining some walls of the container at different temperatures (Hide 1953, Hide & Mason 1975) or by using two immiscible fluids of different densities (Hart 1972). The simplest experimental configuration which thermally forced laboratory rigs are currently designed to reproduce is illustrated in fig. 1.1. The container is a cylindrical annulus with conducting inner and outer sidewalls. The inner wall is held at temperature T_0 and the outer at $T_0 + \Delta T$; $|\Delta T|$ typically lies between 0.5°C and 10°C. The upper and lower lids of the annulus are rigid thermal insulators. Fluid (often water-glycerol) fills the annular cavity and the apparatus is rotated at a uniform rate of Ω radians per second (typically $0.1 \leq \Omega \leq 5$).

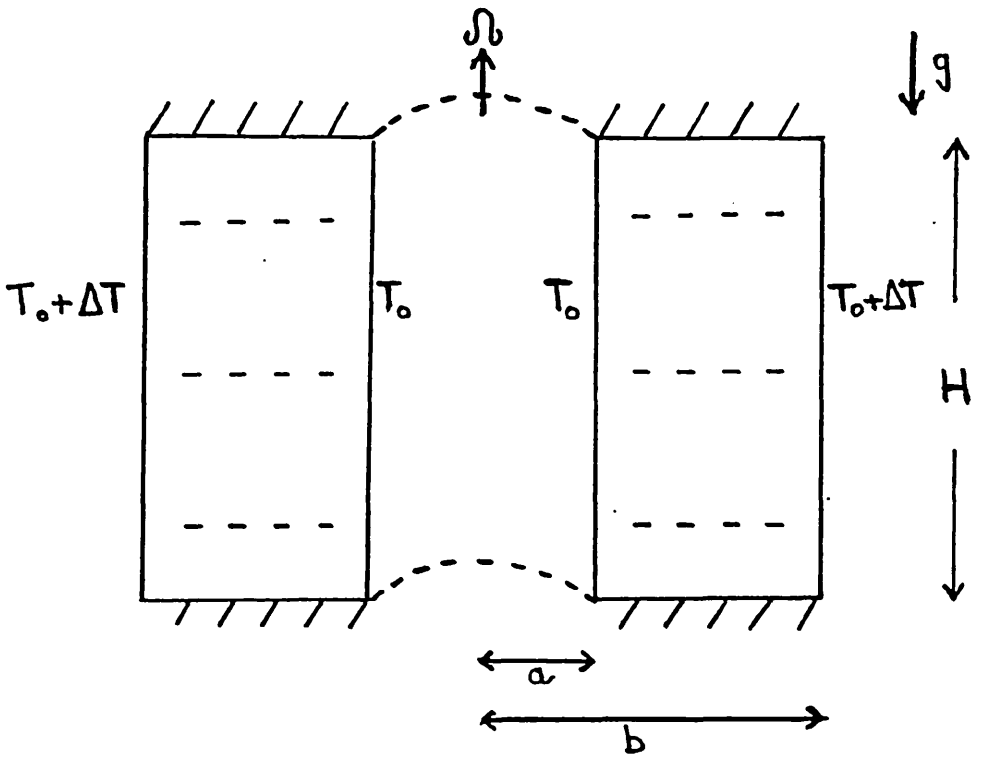


Figure 1.1

Cross-section illustrating the simplest configuration which laboratory (and numerical) experiments aim to reproduce. The cylindrical annulus of fluid is contained by rigid thermally insulating horizontal upper and lower boundaries and rigid conducting vertical side-walls. The whole of the inner wall is maintained at a uniform temperature T_0 and the whole of the outer at $T_0+\Delta T$. The annulus is rotated about its central axis at a constant rate of Ω rads^{-1}. The annulus to which figs. 1.2 - 1.4 apply has inner wall radius a=2.5cm, outer wall radius b=8.0cm and depth H=14.0cm.

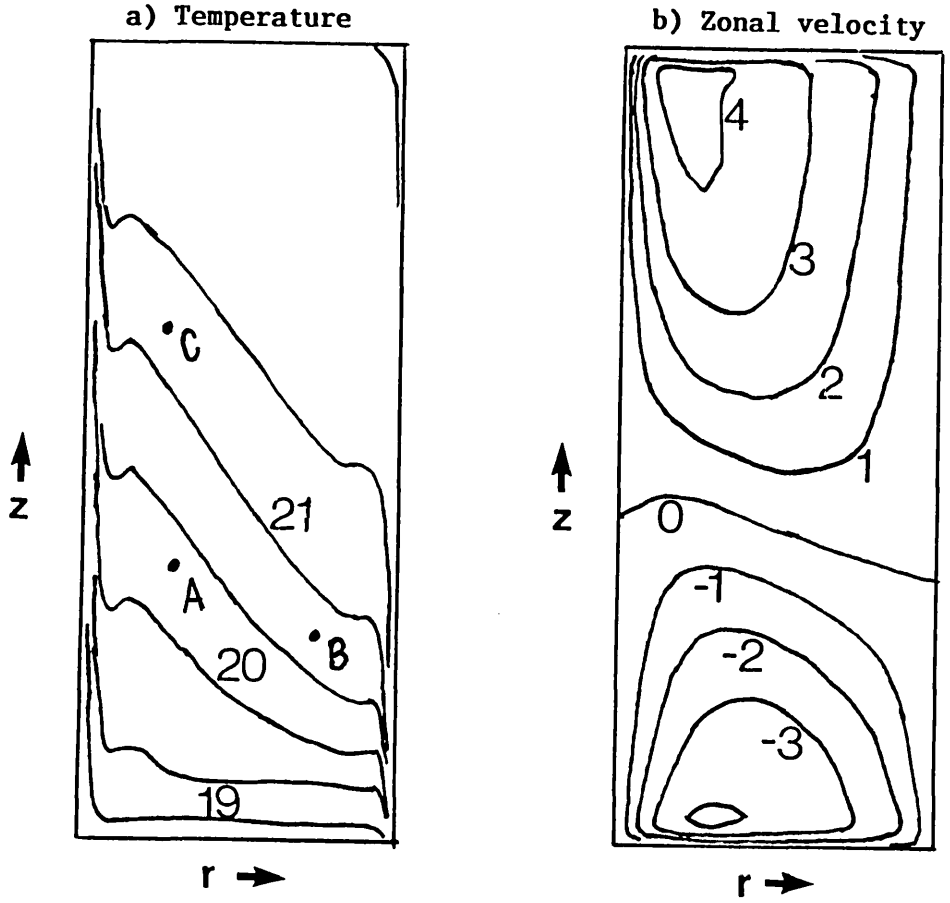


Figure 1.2

Height(z) / radius(r) cross-sections of (a) the temperature (^{o}C) and (b) the zonal velocity (mms^{-1}) fields of an axisymmetric flow. The fields are taken from a numerical simulation which used the annulus geometry of fig. 1.1 with Ω=0.499$rads^{-1}$ and ΔT=4.02^{o}C. The numerical model (Hignett et al. 1985) performs time integrations of the non-hydrostatic Navier-Stokes equations for an incompressible fluid discretized onto a grid with enhanced resolution near the fluid boundaries. The steady state fields presented were obtained using a 2D grid of 32(radial) x 32(vertical) points after 4900 seconds of evolution from a state of rest in thermal equilibrium. T_0 was set at 18.01^{o}C and the fluid had density ρ = 1.045 (1-0.000296(T-20)) gcm^{-3}, and coefficients of kinematic viscosity ν = 1.78 10^{-2} cm^2s^{-1} and diffusivity K = 1.29 10^{-3} cm^2s^{-1}. Points A, B and C are referred to on p 17.

For a particular apparatus and fluid the flows which can be obtained depend chiefly on the temperature difference, ΔT, and Ω. When $\Delta T/\Omega^2$ is large enough, the flow is symmetric about the axis of rotation; at smaller values of $\Delta T/\Omega^2$ non-axisymmetric waves are obtained.

An axisymmetric flow obtained in a numerical simulation with $\Delta T \simeq 4°C$ and $\Omega \simeq 0.5 rads^{-1}$ is illustrated in fig. 1.2. Fig. 1.2a presents a height-radius cross-section of the thermal field (units °C). It shows that at any given radius colder, denser fluid underlies warmer, lighter fluid (i.e. the flow is stably stratified) and that there are strong horizontal thermal gradients in the main body of the fluid particularly near mid-level. Fig 1.2b shows that the azimuthal velocity (units mms^{-1}) has a strong vertical shear and changes sign (i.e. direction) near mid-level.

Fig. 1.3 illustrates a non-axisymmetric wavenumber three flow obtained in a similar numerical simulation with $\Delta T = 4°C$ and $\Omega = 1.0 rads^{-1}$. The horizontal velocity field is shown at three horizontal levels: (a) 0.95 cm below the lid, (b) 5.73 cm above the base and (c) 0.95 cm above the base. It consists of an axisymmetric zonal flow which changes direction with height (as in fig. 1.2) and an azimuthally varying wave flow whose amplitude is relatively independent of height. Comparison of the orientation of the wave lobes in fig. 1.3 shows that the wave leans back slightly with height against the vertical shear of the axisymmetric flow.

The flow of fig. 1.3 is termed a steady wave (S) as its amplitude and shape do not change with time. Other flows with more complex temporal behaviour can also be obtained: the "amplitude" of some waves oscillates in a strictly periodic manner with time whilst the orientations of the troughs of other waves vary (though not with strict periodicity). These flows have been termed amplitude vacillations (AV) and shape vacillations (SV) respectively. It has also been found that several flows of different wavenumber or type can be obtained and retained for considerable periods under the same imposed conditions.

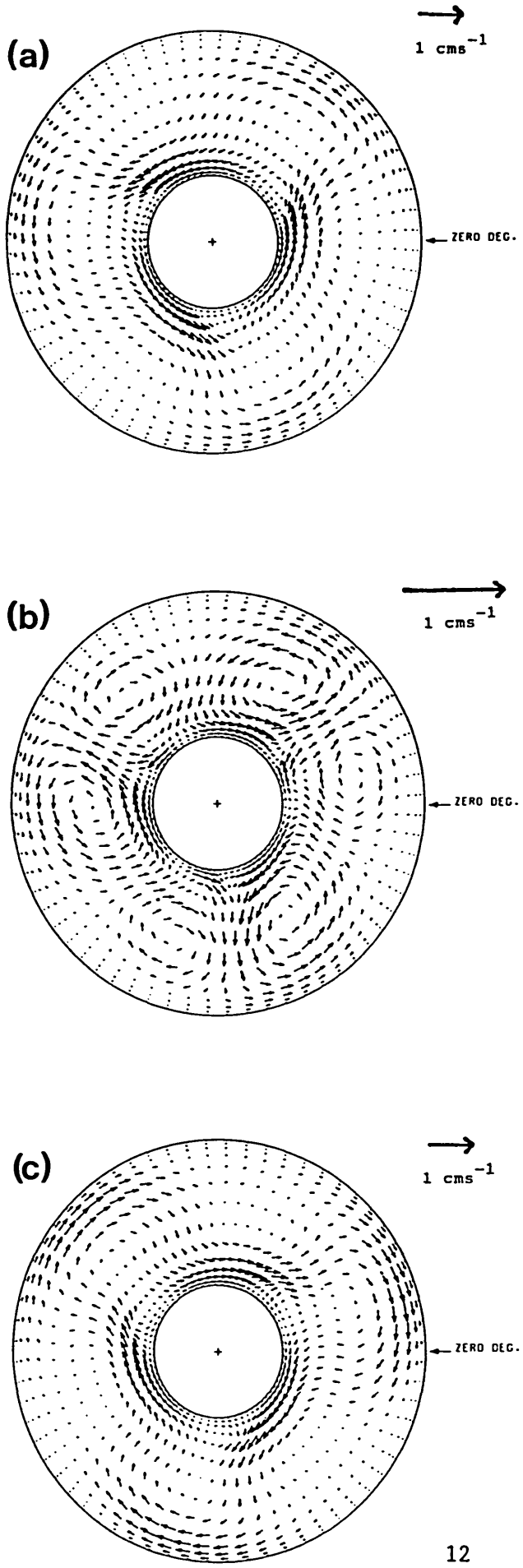


Figure 1.3

Horizontal cross - sections illustrating the horizontal velocity field of a steady wavenumber three flow at (a) 0.95 cm below the lid, (b) 5.73 cm above the base and (c) 0.95 cm above the base. The fields are taken from a numerical simulation using the annulus geometry of fig. 1.1 with Ω=1.0rads^{-1} and ΔT=4.0^{o}C. The simulation used a 3D version of the model described in fig. 1.2 with a grid of 64(azimuthal) x 16(radial) x 16(vertical) points. The fields were obtained after 1000 seconds of evolution from an initial state containing a small thermal wavenumber three perturbation to a steady axisymmetric flow. This simulation is the one used in the main comparison of Hignett et al. (1985); see that paper for more details.

An arrow length corresponding to a speed of 1 cms^{-1} accompanies each figure.

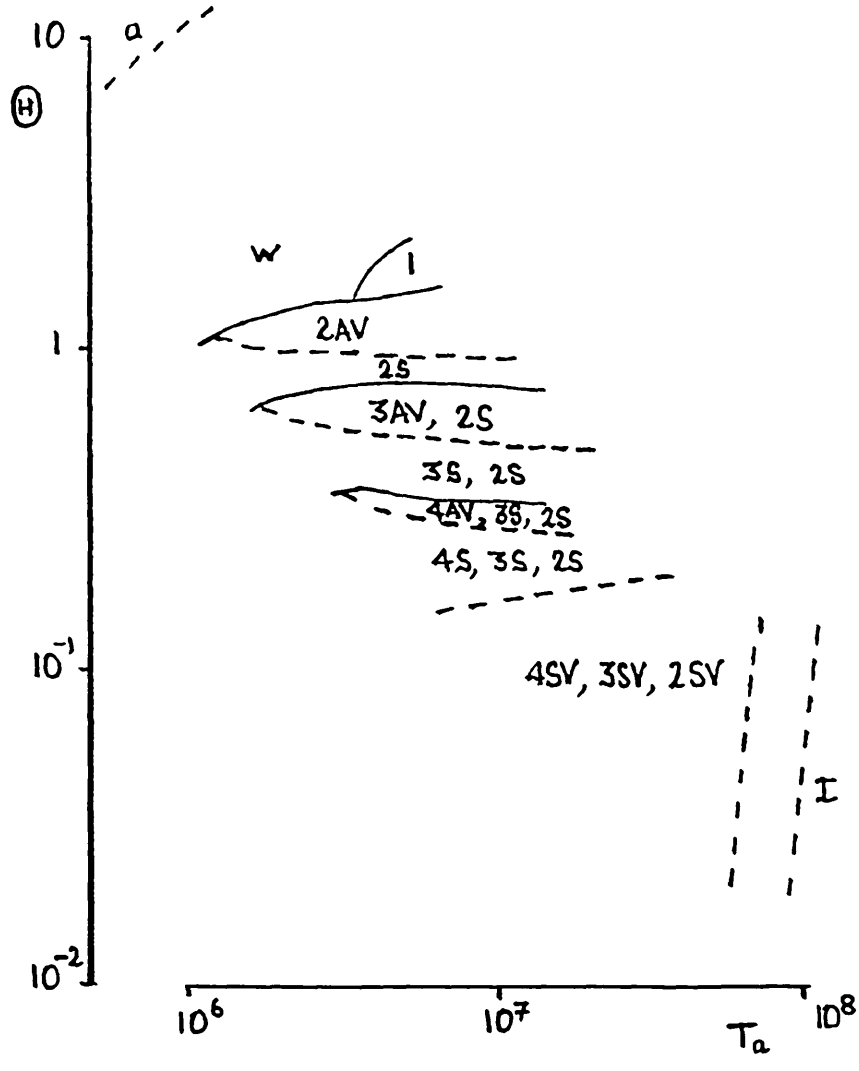


Figure 1.4

Regime diagram for a water-glycerol solution in the annulus of fig. 1.1 taken from Hignett (1985). The thermal Rossby number, $\Theta = g\alpha\Delta TH/(\Omega(b-a))^2$, and Taylor number, $Ta=4\Omega^2(b-a)^5/(\nu^2 H)$ are plotted on logarithmic scales. All flows which can be obtained and retained within the drawn regions are indicated: a denotes axisymmetric flow; w weak waves; S steady waves; AV amplitude vacillation; SV shape vacillation; and I irregular flows. Sharp transitions are indicated by continuous lines and transitions for which the precise location is in doubt by dashed lines. The mean of the side-wall temperatures was kept at 20°C throughout; the fluid has a coefficient of thermal expansion $\alpha=2.86.10^{-4}K^{-1}$, kinematic viscosity $\nu=1.66.10^{-2}cm^2s^{-1}$ and thermal diffusivity $K=1.27.10^{-3}cm^2s^{-1}$ at this temperature.

The flows which can be maintained in a particular experiment depend on ΔT and Ω, on the fluid's properties (e.g. its density ρ, coefficient of thermal expansion α, kinematic viscosity ν and conductivity κ) and the container's geometry (depth H, inner radius a, and outer radius b). The flows have been found to depend chiefly on two non-dimensional combinations of these parameters; namely the thermal Rossby number

$$\Theta = \frac{g \alpha \Delta T H}{\Omega^2 (b-a)^2} \qquad (1.1.1)$$

and the Taylor number

$$Ta = \frac{4 \Omega^2 (b-a)^5}{\nu^2 H} \qquad (1.1.2)$$

(Fowlis & Hide 1965). They also depend somewhat on the aspect ratio (b-a)/H and Prandtl number ν/κ (Jonas 1981).

The types of flow which can be obtained with a particular fluid and annulus geometry (Hignett 1985) over a wide range of thermal Rossby numbers and somewhat restricted range of (fairly high) Taylor numbers are summarised in the regime diagram of fig. 1.4. The flow is axisymmetric when $\Theta > 10$ and axisymmetric except for "weak" waves of small amplitude ($\simeq 0.01\ \Delta T$) when $2 \leq \Theta \leq 10$. The transition between weak waves and large amplitude waves, at high Taylor numbers, is referred to as the upper axisymmetric transition (UAT). Just below this transition only waves with one or two lobes can be obtained. Waves with more lobes can only be found at lower values of Θ.

Within the regular regime (of steady waves and amplitude vacillations) any given steady wave is retained if Θ is reduced sufficiently slowly. If Θ is increased, however, at a certain "transitional" value the wave amplitude starts to vacillate; the amplitude of the vacillation increases as Θ is increased further until the vacillation becomes unstable and gives way to a steady flow of lower wavenumber.

Elementary Geostrophic Dynamics

Fluid motions are governed by Newton's second law (in a form appropriate for a fluid) and the laws of thermodynamics and conservation of mass. Valuable accounts of the dynamics peculiar to fluids in (or on) rotating bodies are given by Greenspan (1968), Hide (1977), Pedlosky (1982a) and Gill (1982).

Let us consider Newton's law first for a fluid at rest relative to the rotating apparatus. A small "parcel" of fluid, distance r from the axis of rotation, accelerates inward in an inertial frame at rate $\Omega^2 r$. It is subject to the downward force of gravity ($\underset{\sim}{g}=-g\hat{\underset{\sim}{z}}$) and also to pressure (p) forces from the fluid surrounding it (Batchelor 1967 § 1.3). Newton's law, written in cylindrical co-ordinates, requires

$$\partial p/\partial z = -\rho g \quad ; \quad \partial p/\partial r = \rho \Omega^2 r \quad ; \quad \partial p/\partial \theta = 0, \qquad (1.1.3a)$$

where $\hat{\underset{\sim}{r}}$, $\hat{\underset{\sim}{\theta}}$ and $\hat{\underset{\sim}{z}}$ are unit vectors and $\hat{\underset{\sim}{r}}$ points radially outward and $\hat{\underset{\sim}{\theta}}$ anti-clockwise when viewed from above. (3a) may be simplified by re-defining the vertical direction to be opposite to that of apparent gravity $\underset{\sim}{g}'=\underset{\sim}{g} + \Omega^2 r\, \hat{\underset{\sim}{r}}$. Using z′ for this vertical ordinate (3a) becomes

$$\partial p/\partial z' = -\rho g' \quad ; \quad \partial p/\partial r = 0 \quad ; \quad \partial p/\partial \theta = 0. \qquad (1.1.3b)$$

A fluid whose pressure increases with depth as in (3b) is said to be in hydrostatic balance.

Moving fluid "parcels" suffer additional stresses due to the friction between oppositely moving fluid elements. The resulting viscous forces are important near the boundaries of the fluid but small in the fluid interior. The acceleration of a fluid parcel moving with velocity $\underset{\sim}{u} =(v\hat{\underset{\sim}{r}}, u\hat{\underset{\sim}{\theta}}, w\hat{\underset{\sim}{z}})$ relative to the apparatus is the sum of its acceleration relative to the apparatus, $D\underset{\sim}{u}/Dt$, its centrifugal acceleration, $-\Omega^2 r\, \hat{\underset{\sim}{r}}$, and its Coriolis acceleration, $2\underset{\sim}{\Omega}\Lambda\underset{\sim}{u}$, which results from changes in its position relative to the axis of rotation due to its relative velocity. The Coriolis

acceleration has magnitude $2\Omega\,(u^2+v^2)^{\frac{1}{2}}$ and a direction which is perpendicular to both the rotation axis and the direction of the horizontal motion.

The ratio of the magnitude of the "relative acceleration" $(D\underset{\sim}{u}/Dt)$ to the Coriolis acceleration may be estimated in two ways. For a motion which is doubling in amplitude every τ_d seconds the ratio is $R_d = \ln 2\ P\ /\ (4\pi\tau_d)$ where P is the period of rotation of the apparatus. The ratio for steady flows whose velocities vary by an amount U over distances of length L is approximately $R_a = U/(2\Omega L)$. R_a can be calculated to be 0.1 for the zonal flow in fig. 1.2 using $U = 5\text{mms}^{-1}$, $L = 50\text{mm}$ and $\Omega = 0.5\text{rads}^{-1}$. So the Coriolis acceleration dominates the "relative acceleration" in disturbances on the axisymmetric flow of fig. 1.2 if they take at least one rotation period to double their amplitude. The same domination of the Coriolis acceleration characterises a large fraction of the motions in atmospheric cyclones.

Away from the boundaries of the fluid Newton's second law for the "horizontal" motions hence yields the following approximate balance between the pressure gradient forces and the Coriolis acceleration

$$2\Omega\rho u = \partial p/\partial r \quad ; \quad 2\Omega\rho v = -1/r\ \partial p/\partial\theta . \qquad (1.1.4)$$

Flows satisfying (4) are said to be in geostrophic balance. The direction of their motion is parallel to the isobars (i.e. perpendicular to the pressure gradients). Differentiation of the first of equations (4) with respect to z′ and use of (3b) shows that

$$2\Omega\rho\,\partial u/\partial z' \simeq \partial/\partial r\ \partial p/\partial z' \simeq -g'\,\partial\rho/\partial r . \qquad (1.1.5)$$

The vertical shear of the zonal flow in fig. 1.2 and the radial gradient of the fluid's thermal and density fields are related by (5). Vertical wind shears related to density gradients as in (5) are termed thermal wind shears.

Conservation of mass requires the net flux of mass through the surface of any volume to equal the increase of mass within it. In

the rotating annulus experiments changes in a fluid parcel's density due to (conductive) diffusive heat fluxes are small enough for the fluid motion to be effectively incompressible;

$$1/r\ \partial/\partial r\ (rv) \quad + \quad 1/r\ \partial u/\partial \theta \quad + \quad \partial w/\partial z' \quad = 0. \qquad (1.1.6)$$

Using relation (4) (of geostrophic balance) in (6) (and neglecting the variations in ρ) one infers that

$$\partial w/\partial z' \quad = 0. \qquad (1.1.7)$$

Integrating (7) with respect to height from the rigid horizontal lower boundary (at which w is small), shows that for flows in geostrophic balance the vertical motions are small compared to the horizontal geostrophic motions and are associated with small departures from geostrophic motion.

The vertical motions are nevertheless of vital importance in the waves in the annulus experiments. The waves gain energy principally by vertical motions which extract gravitational energy from the axisymmetric density distribution associated with the zonal flow. The process is illuminated by consideration of the gravitational energy released on or required for the interchange of fluid parcels of equal volumes between various points in the fluid. Consider exchanges between the points labelled A, B and C in fig. 1.2a. The fluid at C lies directly above that at A and being warmer is less dense. On interchanging parcels at A and C there would be a net upward transport of mass and gravitational energy would increase. The fluid at A is, however, both denser and higher than that at B; so gravitational energy is released on exchange between these points. Interchange between B and C clearly does not release energy so only interchanges along lines of sufficiently shallow slope release potential energy.

To determine the vertical motions and time evolution of the geostrophic flow more accurate approximations to the governing equations than (4) are required. The quasi-geostrophic set of approximations and equations are appropriate for theoretical

investigations of annulus flows. They are summarised in section 2.1 and used throughout chapters two to four.

A summary of the thesis

The transition between the axisymmetric and regular wave regimes is the easiest feature of the experimental regime diagram to analyse and (perhaps for this reason) its interpretation has been the most keenly disputed. The main point at issue is whether the upper transition can be interpreted in terms of inviscid quasi-geostrophic theory. Chapter two clarifies this issue somewhat by identifying several (related) types of axisymmetric flows which do possess inviscid transitions and have distributions of vertical shears and potential vorticity which are similar to those of the annulus flows in several important respects. Several examples illustrating the (sometimes sensitive) dependence of the transition on the distribution of the flow's vertical and lateral shears are also presented and various interpretations of the transitions discussed. The final section of chapter two presents a varied selection of normal mode growth rate curves. A more detailed summary of the chapter is provided in section 2.1.

The evolution and equilibration of small amplitude waves on baroclinic zonal flows has been the subject of intense analytical study (Hart 1979). One aim of this work has been to provide insights into the dynamics of amplitude vacillating and/or steady waves just below the upper axisymmetric transition. Strong reasons are advanced in section 3.1 for thinking that the lateral shear of the axisymmetric flow has a crucial impact on the self-interaction of such waves. A detailed analysis in sections 3.2 & 3.3 of the inviscid development of a small amplitude wave on a laterally sheared flow confirms that the advection of the flow's potential vorticity in a strongly non-linear critical layer can determine the wave's evolution. The self-consistency of the analysis is examined in section 3.4. The calculations suggest that diffusive fluxes play an important modifying role in waves near the axisymmetric transition at moderate Taylor numbers.

It is remarkable that several large amplitude stable waves of different azimuthal wavenumbers can be obtained under the same imposed conditions over a large fraction of the regular wave regime. It is natural to wonder whether simple accounts of these waves and their stability can be given and whether or not the stability of a wave flow depends crucially on diffusive effects, on the confining effects of the boundaries or on the strength of the zonal mean component of the flow. The numerical investigations of chapter four provide some insight into these matters. It is shown that barotropic Rossby waves of several azimuthal wavenumbers are stable against linear quasi-geostrophic perturbations in narrow periodic f-plane channels. Steady state combinations of zonal flows and Rossby waves (White 1986) can also be stable when the zonal flow component is not too strong. These results suggest that the boundaries play a crucial role in stabilizing the waves but that diffusive effects are not essential to the waves' stability and that the waves are stable despite (rather than because of) the zonal mean component of the flow.

Modern advances in the understanding of simple non-linear dynamical systems might provide a useful qualitative framework for the classification and interpretation of shape vacillating flows and the transition from regular to irregular flows. A method (Takens 1981) for the investigation of a time series of experimental data generated by a simple (low dimensional) dynamical system is discussed in chapter five. The present author's interest in this method stemmed from problems encountered whilst trying to use it with experimental data. The discussion might be considered didactic but is included because experimenters have evidently found the method both difficult to understand and to implement soundly.

CHAPTER TWO

THE STABILITY OF ZONAL FLOWS AND THE UPPER AXISYMMETRIC TRANSITION

Section 2.1 INTRODUCTION

The sharpness of the upper axisymmetric transition is due to a strong dependence of the stability of the axisymmetric flow against small amplitude baroclinic waves on the imposed thermal Rossby number. Eady's (1949) theory gives a satisfying account of the transition , since it features a rapid transition between stability and healthy instability over Burger numbers consistent with the thermal Rossby number of the experimental transition.

The present work which explores similarly simplified but essentially realistic idealisations of the stability of zonal flows has two aims :

i) to clarify to which aspects of the zonal flow the transition is sensitive
ii) to improve understanding of the reasons for the existence of the transition.

The study was stimulated by the lack of solutions, other than Eady's and of two layer problems, which exhibit clear transitions and by two features of the axisymmetric flows found in wall heated rigid lid annulus experiments. These features are reproduced in numerical simulations of the flows using the Navier Stokes equations and illustrated in figure 2.1.

The first is the lack of a lateral thermal gradient in the zonal flow near the upper and lower boundaries (see fig. 2.1a). The vertical shear of the zonal flow (fig 2.1b) is also small at these boundaries as is consistent with the thermal wind relation.

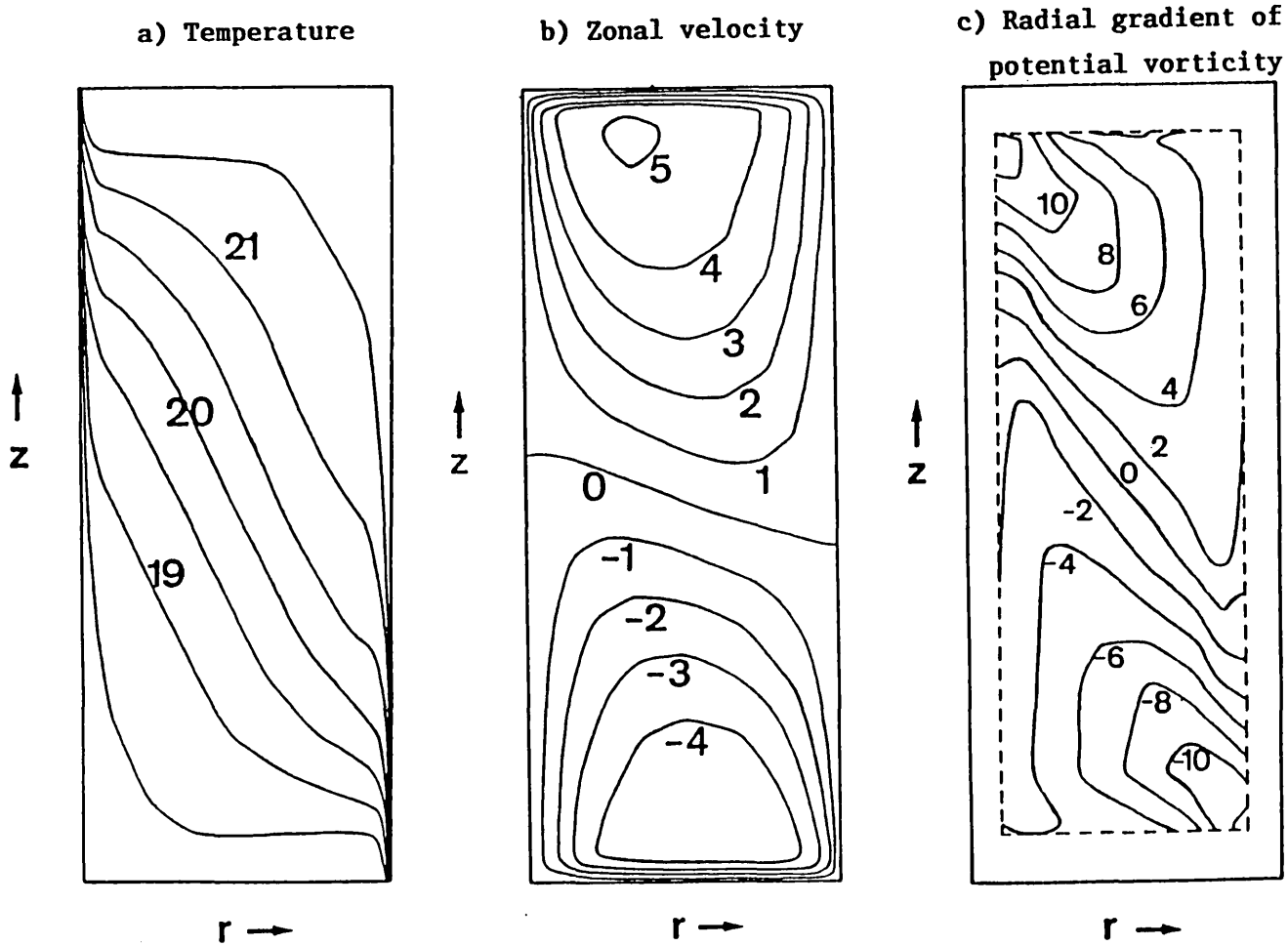


Figure 2.1

Height(z) / radius(r) cross-sections of the fields of (a) temperature ($^{\circ}C$), (b) zonal velocity (mm/s) and (c) the quasi-geostrophic potential vorticity gradient q_y ($0.2545 cm^{-1}s^2 = 2\Omega/(b-a)$) of an axisymmetric flow. The fields were derived from a 2D numerical simulation (see caption of fig. 1.2) with $\Delta T = 4^{\circ}C$ and $\Omega = 0.7$rad/s. in an annulus of inner radius a=49.75cm, outer radius b=55.25cm and depth H=14cm. The fluid has $\nu = 1.29 cm^2 s^{-2}$, $\kappa = 1.34.10^{-3}\ cm^2 s^{-1}$, $\alpha = 2.77.10^{-4}\ K^{-1}$ and $\rho = 1.026\ gcm^{-3}$ at $22^{\circ}C$, the temperature of the outer side-wall. q_y is constructed from the horizontal velocity and thermal fields using an area-averaged static stability at each level: the quantity actually plotted is $-\partial q/\partial r$ the cylindrical analogue of $\partial q/\partial y = q_y$. The flow presented is unstable to non-axisymmetric perturbations but similar in gross form to stable axisymmetric flows (see fig. 1.2).

Following Charney & Stern (1962) flows with no vertical shear at their boundaries will be termed internal baroclinic jets.

The second feature is that the potential vorticity gradient of the zonal flow, defined[1] (for constant N^2) by

$$\partial\bar{q}/\partial y \equiv \beta - \partial^2\bar{u}/\partial y^2 - f^2/N^2\,\partial^2\bar{u}/\partial z^2 \qquad (2.1.1)$$

and illustrated in fig. 2.1c, is of the same sign as the zonal velocity field, $\bar{u}$, throughout most of the annulus. The axisymmetric flow's distribution of $\partial\bar{q}/\partial y$ is related to its internal jet character which inclines its vertical shear, $|\partial\bar{u}/\partial z|$, to be a maximum at mid-level implying that $\partial^2\bar{u}/\partial z^2$ and $\bar{u}$ have opposite signs over most of the flow. Also, at a given height, the zonal velocity takes its maximum value at mid-channel so that $\partial^2\bar{u}/\partial y^2$ and $\bar{u}$ have opposite signs in most places. The combination of these points with (1) and the zero value of β in the experiments explains the distribution observed. This suggests that one study the idealised flows for which

$$T_s \equiv \partial\bar{q}/\partial y \,/\, (\bar{u} - \bar{u}_s) > 0 \qquad (2.1.2)$$

(i.e. T_s is positive definite throughout the flow) for some constant velocity $\bar{u}_s$ - which may be viewed as a Doppler shift velocity.

In sections 2.2 to 2.4 we examine the stability of flows which satisfy one or both of these idealisations. It is argued that such flows are stable to very short wave disturbances and that the short wave cut-off (SWC) to instability is marked by a neutral mode solution. Under some conditions the dependence of the neutral solution and hence of the SWC on the form of the zonal flow can be determined readily.

Section 2.2 presents an argument (adapted from Howard (1964)) concerning normal modes on internal baroclinic jets which satisfy (2). The argument establishes a Sturm Liouville problem for the neutral mode marking the SWC. This result is used to investigate

[1] $f = 2\Omega$ is the Coriolis parameter, $\beta = \partial f/\partial y$ the beta parameter and N the Brunt Vaisala frequency. An overbar indicates a zonally averaged quantity.

the dependence of the SWC on the degree and form of the lateral shear of the zonal flow and the tightness of the cylindrical curvature of the domain. It is found that lateral shear can affect the position of the cut-off considerably.

Normal modes on flows which satisfy (2) but are not internal jets are considered in section 2.3. The arguments locating inviscid SWCs are less direct than those of section 2.2 and involve three steps. Firstly an upper bound on the wavenumber of unstable solutions is found by making use of two integral constraints on the flow. Secondly it is argued that any unstable solution at a given wavenumber is bordered by solutions at both larger and smaller wavenumbers. Thirdly the gravest neutral modes bordering unstable solutions are identified. The arguments are presented first for laterally uniform flows. They are then extended to domains with slightly sloping endwalls and to laterally sheared flows. Two examples illustrate the dependence of the SWC on the height of the $\partial\bar{q}/\partial y = 0$ line for laterally uniform flows with thermal boundary gradients. Some, but not all, of the three steps outlined above apply to waves on laterally uniform flows in the presence of Ekman pumping; the neutral modes for most of these problems are shown to have a discontinuous thermal gradient at an internal critical level.

A three dimensional spectral model of internal jets is described in section 2.4. It is shown that the model has an anti-cascade property which applies at any truncation. This result provides an alternative interpretation of the sensitivity of the SWC to certain forms of lateral shear (found in section 2.2) and can be applied to flows which do not satisfy (2). It also forms the basis of the stability calculations for free modes reported in chapter four.

Section 2.5 provides a summary of the types of flows which possess SWCs and the arguments and interpretations of the cut-offs suggested by other authors.

The SWCs discussed in sections 2.2 to 2.5 are of primary importance for annulus flows only if they mark sharp transitions between stability and instability. Condition (2) is a judicious

criterion in this regard because it does not allow feeble unstable solutions trapped near $\partial\bar{q}/\partial y = 0$ to determine the cut-off position. Shooting methods for calculating unstable normal modes on laterally uniform and some laterally sheared flows are described in section 2.6 and some examples of their use presented in section 2.7. The first examples confirm that lateral shear can significantly affect the stability transition. An examination of a branch of solutions decaying under the influence of Ekman pumping follows. The chapter closes with two examples of flows with strongly unstable short waves. The first is Green's problem scaled in the manner most appropriate for annulus experiments. The second concerns vertically trapped jets similar to (but simpler than) those on which polar lows develop.

All of the studies reported in this and the following chapter are based on the quasi-geostrophic equations for an incompressible baroclinic fluid. These equations are particularly suited to analytical investigation because they express all velocity and thermal fields, to an adequate approximation, in terms of a single scalar field, namely Ψ, the horizontal geostrophic velocity streamfunction;

$$\underline{u}_g = \hat{\underline{k}} \wedge \underline{\nabla} \Psi \, . \qquad (2.1.3)$$

The governing equation states that the quasi-geostrophic potential vorticity, which is related to the streamfunction by

$$Q \equiv f + \nabla_h^2 \Psi + \partial/\partial z \left(f^2/N^2 \, \partial\Psi/\partial z \right) , \qquad (2.1.4)$$

is conserved following the geostrophic motion

$$D_g Q/Dt \equiv \partial Q/\partial t + (\underline{u}_g \cdot \underline{\nabla}) Q = 0 \, . \qquad (2.1.5)$$

Equation (5) may be derived by a systematic scale analysis from the Navier - Stokes momentum and thermodynamic equations using two assumptions. First that the Coriolis accelerations dominate accelerations apparent in the rotating frame; this requires the Rossby number to be small;

$$R_o \equiv U/(fL) \ll 1 \quad \text{and} \quad fT \gg 1 , \qquad (2.1.6)$$

where U, L and T are representative of the velocity, length and time scales of the motion and f is the Coriolis parameter. The second assumption is that

$$Ri \, . \, R_o = B / R_o \gg 1 , \qquad (2.1.7)$$

where

$$Ri \equiv N^2H^2/U^2 \quad \text{and} \quad B \equiv N^2H^2/(f^2L^2) , \qquad (2.1.8)$$

Ri being the Richardson number, B the Burger number, N the Brunt Vaisala frequency and H a typical depth scale of the motions. This ensures that $|w/H| \ll |v/L|$, so that the horizontal flow is non-divergent ($\nabla_h . \underset{\sim}{u}_h = 0$) to a first approximation[1], and that the basic thermal stratification is larger than variations induced in it by the motions. Approximations (6) and (7) are appropriate for many flows in both rotating annulus experiments and the atmosphere at mid or high latitudes. The derivation also assumes that the basic thermal stratification depends only on height. Zonal flows with both lateral and vertical shear will not in general be strictly consistent with this assumption and thermal wind balance. It may be interesting to investigate the effect of the neglected terms on the linear stability analyses of this chapter.

The governing equation (5) has the merit of being an analogue of Ertel's conservation law for an inviscid, adiabatic fluid (Pedlosky 1982a). It is just one of a set of approximations to Ertel's result, some of which are coarser (Phillips 1963) and others finer (Hoskins 1975) approximations (White 1987).

Boundary conditions on the normal flow are required to solve (5) for Ψ. The upper and lower boundary conditions are fairly firmly established for annulus flows. The interior vertical velocity matches onto Ekman layers which pump vertical motion according to the vorticity at their outer edge (Pedlosky 1982a). At the upper and lower boundaries the vertical velocities are given by

[1] The vertical velocity at the boundaries of the geostrophic interior must also be small enough for this to be true.

$$w = -f/N^2 \, (D_g/Dt) \, \partial\Psi/\partial z \quad \text{at } z=0,H$$

$$= +\tfrac{1}{2} \, (2\nu/f)^{1/2} \, \underline{k}.(\underline{\nabla} \wedge \underline{u}_g) \quad \text{at } z=0$$

$$= -\tfrac{1}{2} \, (2\nu/f)^{1/2} \, \underline{k}.(\underline{\nabla} \wedge \underline{u}_g) \quad \text{at } z=H \, . \qquad (2.1.9)$$

These conditions assume that $N^2 H / g << 1$ (White 1977).

The side boundary conditions are much less firmly established. The condition of no normal geostrophic flow, which seems reasonable, is commonly applied at the edge of a notional geostrophic interior;

$$\hat{\underline{n}} \cdot \underline{u}_g = 0 \quad \text{at} \quad y = 0, L \, . \qquad (2.1.10)$$

Conditions on the zonal flow at the side boundary are subject to even greater uncertainty. One which is widely used because it ensures energy conservation is

$$\partial \tilde{u}_g / \partial t = 0 \quad \text{at} \quad y = 0, L \, . \qquad (2.1.11)$$

Most of the calculations in this thesis are carried out in a periodic channel, rather than a cylindrical annulus, which has width L and height H. The x ordinate is chosen to be along the channel, y across it and z in the vertical direction. For reference we write out the equations for this domain in the non-dimensional notation used hereafter. The gap width L is used as the horizontal scale, H as the vertical and U as a velocity scale. Denoting dimensional variables with an asterisk superscript in the usual manner we take

$$u^* = Uu \, , \; v^* = Uv \, , \; x^* = Lx \, , \; y^* = Ly \, , \; z^* = Hz. \qquad (2.1.12a)$$

An appropriate non-dimensional time variable and streamfunction are

$$t^* = U/L \; t \quad , \quad \Psi^* = UL \, \Psi \, . \qquad (2.1.12b)$$

The non-dimensional form of the interior equation is then

$$D_g Q/Dt = \partial Q/\partial t + \partial\Psi/\partial x\, \partial Q/\partial y - \partial\Psi/\partial y\, \partial Q/\partial x = 0$$

$$Q = fL/U + \partial^2\Psi/\partial x^2 + \partial^2\Psi/\partial y^2 + \partial/\partial z\,(1/B\, \partial\Psi/\partial z)\,. \tag{2.1.13}$$

Many of the calculations in this chapter concern small perturbations to a horizontal zonal flow, $\bar{u}(y,z)$ with no variation in the x direction, which are not subject to Ekman pumping. Perturbations with a streamfunction ψ' of normal mode form

$$\psi' = \phi(y,z)\, \exp ik(x-ct) \tag{2.1.14}$$

are of particular (though not exclusive) interest. Such perturbations are governed by the set

$$(\bar{u}-c)(\partial^2\phi/\partial y^2 + \partial/\partial z\,(1/B\, \partial\phi/\partial z) - K^2\phi) + \bar{q}_y \phi = 0 \tag{2.1.15}$$

$$\bar{q}_y = \gamma - \partial^2\bar{u}/\partial y^2 - \partial/\partial z\,(1/B\, \partial\bar{u}/\partial z)\ ;\ \gamma = \beta L^2/U \tag{2.1.16}$$

$$\phi = 0 \qquad \text{at} \quad y = 0,1 \tag{2.1.17}$$

$$(\bar{u}-c)\, \partial\phi/\partial z - \partial\bar{u}/\partial z\, \phi = 0 \qquad \text{at} \quad z = 0,1 \tag{2.1.18}$$

as may be verified readily using (9),(10),(12) and (13). When dealing with this set $\bar{u}$ and $\partial\bar{q}/\partial y$ will be denoted by u and $\partial q/\partial y$, the overbar being superfluous.

Section 2.2 BAROCLINIC INTERNAL JETS

This section concerns only inviscid normal mode perturbations to internal baroclinic jets with $T_s > 0$ (see (2.1.2)). So the equations to be considered are

$$\partial^2\phi/\partial y^2 + \partial/\partial z\,(1/B\,\partial\phi/\partial z) - K^2\phi + \frac{q_y\phi}{u-c} = 0 \qquad (2.2.1)$$

$$\phi = 0 \quad \text{at } y = 0, 1 \;; \quad \partial\phi/\partial z = 0 \quad \text{at } z = 0, 1 \qquad (2.2.2)$$

$$q_y = T_s\,(u - u_s)\,. \qquad (2.2.3)$$

An argument adapted from an analysis of inviscid 2D disturbances to a parallel shear flow by Howard (1964) establishes that no unstable solutions of (1) - (3) have larger azimuthal wavenumbers than the neutral solution which is the gravest mode of the Sturm - Liouville problem defined by (4) and (5) below. The condition under which this neutral solution borders unstable ones and an expression for the growth rates of any bordering solutions are then established. The dependence of the SWC on the lateral shear of some separable velocity profiles is then discussed in some detail. The effect of cylindrical curvature on the SWC is also briefly noted. A discussion of the implications of these results for the interpretation of the UAT concludes the section.

HOWARD'S ARGUMENT

Consider the eigenvalue problem

$$\partial^2 f/\partial y^2 + \partial/\partial z\,(1/B\,\partial f/\partial z) + T_s f + \lambda f = 0 \qquad (2.2.4)$$

$$f = 0 \quad \text{at } y = 0, 1 \;; \quad \partial f/\partial z = 0 \quad \text{at } z = 0, 1 \qquad (2.2.5)$$

and let its eigenvalues be denoted by $\lambda_1, \lambda_2, \lambda_3, \ldots$ where $\lambda_1 < \lambda_2 \leqslant \lambda_3 \leqslant \ldots$

Result:

If $k^2 \geq -\lambda_n$ there are no more than $n - 1$ linearly independent unstable (growing) solutions of the stability problem (1) to (3).

The eigenfunction of the first eigenvalue λ_1 is a neutral mode solution of the stability problem with phase speed $c = u_s$ and $k^2 = -\lambda_1$. This gravest neutral mode eigenfunction will be denoted by ϕ_s and its wavenumber by k_s ($k_s^2 = -\lambda_1$). The above result proves that it marks the SWC if it is bordered at marginally smaller k by unstable solutions (see below).

The proof of the result makes use of the fact[1] that for any $n \geq 1$, λ_n is the minimum value of

$$\mathcal{J}(f) = \int_0^1 \int_0^1 \left(|\partial f/\partial y|^2 + 1/B\, |\partial f/\partial z|^2 - T_s |f|^2 \right) dy\, dz \qquad (2.2.6)$$

among complex-valued piecewise continuously differentiable functions satisfying (5) which are orthogonal to the eigenfunctions of λ_1, $\lambda_2, \ldots \lambda_{n-1}$ and of "unit magnitude", i.e.

$$\int_0^1 \int_0^1 |f|^2 \, dy\, dz = 1 .$$

Assume that (ϕ_1, c_1) and (ϕ_2, c_2) are unstable solutions of the stability problem with the same azimuthal wavenumber k. Integration by parts and application of (1) for ϕ_2 shows that

$$I_1 \equiv \iint \left(\partial\phi_1^*/\partial y\, \partial\phi_2/\partial y + 1/B\, \partial\phi_1^*/\partial z\, \partial\phi_2/\partial z + k^2 \phi_1^* \phi_2 \right) dy\, dz$$

$$= \iint \phi_1^* \phi_2 \, T_s (u - u_s)/(u - c_2) \, dy\, dz .$$

Similar manipulation using (1) for ϕ_1 gives

$$I_1 = \iint \phi_1^* \phi_2 \, T_s (u - u_s)/(u - c_1^*) \, dy\, dz .$$

Hence writing

$$F_K \equiv \phi_K/(u - c_K) \, , \qquad (2.2.7)$$

[1] Courant & Hilbert (1953) chapter VI §§ 1-3.

$$0 = \iint T_s \, F_1^* F_2 \, (u-u_s)(c_2 - c_1^*) \, dy \, dz .$$

The imaginary part of the last equality implies that

$$0 = \iint T_s \, F_1^* F_2 \, (u-u_s) \, dy \, dz . \qquad (2.2.8)$$

Now assume that (ϕ_1, c_1), (ϕ_2, c_2),.. (ϕ_n, c_n) are n linearly independent growing solutions of the stability problem at the same wavenumber k. Then there is a

$$\psi = \sum_{k=1}^{n} a_k \phi_k \quad ; \iint |\psi|^2 \, dy \, dz = 1 \qquad (2.2.9)$$

which is orthogonal to the first n - 1 eigenfunctions of (4) and (5). Applying (1) for each ϕ_k the definition of ψ implies that

$$\partial^2\psi / \partial y^2 + \partial/\partial z \, (1/B \, \partial\psi/\partial z) - K^2\psi = - \sum_{k=1}^{n} T_s (u-u_s) \, a_k F_k .$$

Simple rearrangement yields

$$\partial^2\psi / \partial y^2 + \partial/\partial z \, (1/B \, \partial\psi/\partial z) + T_s \psi - K^2\psi = \sum_{k=1}^{n} T_s (u_s - c_k) a_k F_k .$$

Multiplying this equation by ψ^*, integrating over the whole domain and using (8) reveals that

$$\iint \left(- |\partial\psi/\partial y|^2 - 1/B \, |\partial\psi/\partial z|^2 + T_s |\psi|^2 - K^2 |\psi|^2 \right) dy \, dz$$

$$= \iint T_s \sum_k \sum_\ell (u_s - c_k)(u - c_\ell^*) \, a_k F_k \, a_\ell^* F_\ell^* \, dy \, dz$$

$$= \iint T_s \sum_k \sum_\ell (u_s - c_k)(u - u_s + u_s - c_\ell^*) \, a_k F_k \, a_\ell^* F_\ell^* \, dy \, dz$$

$$= \iint T_s \left| \sum_k (u_s - c_k) \, a_k F_k \right|^2 dy \, dz$$

$$\geq 0 . \qquad (2.2.10)$$

Using the orthogonality of ψ to the first n - 1 minimising functions of (6) in (10) gives

$$-K^2 \geq \iint \left(|\partial\psi/\partial y|^2 + 1/B\, |\partial\psi/\partial z|^2 - T_s |\psi|^2 \right) dy\, dz \geq \lambda_n . \quad (2.2.11)$$

Hence $k^2 \leq -\lambda_n$ if there are n unstable solutions of the stability problem.

To prove the result it remains to exclude the possibility of n growing solutions being present when $k^2 = -\lambda_n$. In that case equality must hold in (10) and

$$T_s \sum_{k=1}^{n} a_k (u_s - c_k) F_k = 0 \qquad 0 \leq y, z \leq 1 .$$

Combining (1) with this equality shows that

$$\sum_k a_k (u_s - c_k)\left[\partial^2\phi_k/\partial y^2 + \partial/\partial z\,(1/B\, \partial\phi_k/\partial z) - K^2\phi_k \right]$$

$$= -T_s (u - u_s) \sum_k a_k (u_s - c_k) F_k = 0 \qquad 0 \leq y, z \leq 1 . \quad (2.2.12)$$

Defining

$$g \equiv \sum_{k=1}^{n} a_k (u_s - c_k) \phi_k , \quad (2.2.13)$$

(12) may be re-expressed as

$$\partial^2 g/\partial y^2 + \partial/\partial z\,(1/B\, \partial g/\partial z) - K^2 g = 0 \qquad 0 \leq y, z \leq 1 .$$

g satisfies the same boundary conditions, (2), as ϕ, so the last equation implies that

$$\iint \left(|\partial g/\partial y|^2 + 1/B\, |\partial g/\partial z|^2 + K^2 |g|^2 \right) dy\, dz = 0$$

and $g \equiv 0$ for $0 \leq y, z \leq 1$. $u_s - c_k \neq 0$, because c_k has a positive imaginary part, so (13) implies that the set $\{\phi_1, \ldots \phi_n\}$ is not linearly independent. Hence, if $k^2 = -\lambda_n$ there are less than n linearly independent unstable solutions of the stability problem and the proof of the result is complete.

Unstable modes bordering the gravest neutral mode

We confirm now that the gravest neutral mode ϕ_s does border unstable modes on its long wave side, provided that $c = u_s$ gives a critical level within the flow domain at which T_s / $\partial u/\partial z$ is of order 1. We explore solutions which border ϕ_s by substituting

$$\phi = \phi_s + \delta\phi \quad , \quad c = u_s + \delta c \quad , \quad K^2 = K_s^2 + \delta K^2 \qquad (2.2.14)$$

into (2.1.15), (2.1.17) and (2.1.18) and linearising with respect to the small perturbations

$$(\partial^2/\partial y^2 + \partial/\partial z\,(1/B\,\partial/\partial z) - K_s^2 + T_s)\delta\phi = -\frac{T_s \delta c \phi_s}{u-c} + \delta K^2 \phi_s \qquad (2.2.15)$$

$$\delta\phi = 0 \text{ on } y=0,1 \; ; \; (u-u_s)\partial/\partial z\,\delta\phi - \partial u/\partial z\,\delta\phi = \delta c\,\partial\phi_s/\partial z \text{ on } z=0,1.$$

Use of (2.1.18) in place of the internal jet boundary condition at this point saves needless repetition in section 2.3. Appendix A shows that the necessary and sufficient condition for (15) to have a solution for $\delta\phi$ is that

$$0 = \int_z\int_y \phi_s \left(-\frac{T_s \delta c \phi_s}{u-c} + \delta K^2 \phi_s\right) dy\,dz - \int_y \left[\frac{\partial c\,\partial u/\partial z\,\phi_s^2}{B(u-u_s)^2}\right] dy \quad . \qquad (2.2.16)$$

Consider

$$J \equiv \mathrm{Im}\int_z\int_y -\frac{T_s \phi_s^2}{u-c}\,dy\,dz = -\delta c_i \int_z\int_y \frac{T_s \phi_s^2}{(u-c_r)^2 + \delta c_i^2}\,dy\,dz \quad . \qquad (2.2.17)$$

Providing that the critical level lies within the flow domain use of the standard substitution (Maslowe 1981) of v for the variable z

$$u - c_r = \delta c_i \tan v$$

enables one to deduce that

$$J = -\pi\,\mathrm{sign}(\delta c_i)\int_0^1 \left(T_s \phi_s^2 / |\partial u/\partial z| \right)_{z=z_s} dy. \qquad (2.2.18)$$

So (16) may be rewritten in the schematic form

$$\left(-i\Gamma^2\,\mathrm{sign}(\delta c_i)+\Lambda\right)\delta c \;+\; T^2\delta k^2 = 0$$

in which Γ, Λ and T are real constants and Γ and T are of order 1. Hence

$$\delta c = -T^2\delta k^2\left(\Lambda + i\Gamma^2\,\mathrm{sign}(\delta c_i)\right)\Big/\left(\Gamma^4+\Lambda^2\right) . \qquad (2.2.19)$$

So there are unstable solutions on the long wave side of k_s (i.e. for $\delta k^2 < 0$) but not on the short wave side. It is worth noting that the gradient of the unstable branch dc_i / dk takes a finite value at these SWCs (at least when T_s / $\partial u/\partial z$ is of order 1 at the critical level) whereas the Eady growth rate curve has $dc_i/dk \to -\infty$ at its SWC.

Example 1: Sinusoidal Internal Jets

Let flows which satisfy

$$q_y = \alpha^2 u + \delta \qquad (2.2.20)$$

for real constants α and δ, be termed sinusoidal flows. α will be referred to as the total wavenumber of the flow. These flows satisfy (2.1.2) since choosing

$$u_s = -\delta / \alpha^2 \qquad (2.2.21)$$

and substituting (20) and (21) in (2.1.2) gives

$$T_s = \alpha^2 . \qquad (2.2.22)$$

According to Howard's argument sinusoidal flows which are also baroclinic internal jets have no unstable modes at azimuthal wavenumbers k with $k^2 \geq -\lambda_1 = k_s^2$, where k_s is the most positive eigenvalue of

$$\partial^2\phi_s/\partial y^2 + \partial/\partial z(1/B\,\partial\phi_s/\partial z) + (\alpha^2 - k_s^2)\phi_s = 0$$

$$\phi_s = 0 \quad \text{on } y = 0,1 \;;\; \partial\phi_s/\partial z = 0 \quad \text{on } z = 0,1\,. \tag{2.2.23}$$

((23) is obtained from (4) and (5) by setting $f = \phi_s$, $\lambda_1 = -k_s^2$ and using (22) for T_s.) Furthermore, this neutral mode (ϕ_s, k_s) marks the SWC provided u_s lies within the flow domain. k_s is particularly easy to calculate for sinusoidal internal jets when the stratification is independent of height and hence B a constant -

$$k_s^2 = \alpha^2 - \pi^2 \quad \text{and} \quad \phi_s = \sin \pi y\,. \tag{2.2.24}$$

One set of examples of sinusoidal internal jets is

$$u = -1/2 \cos \pi z \cos \eta\pi(y - 1/2)\,. \tag{2.2.25}$$

On an f plane these flows satisfy (20) with $\delta = 0$, whilst on a β plane $\delta = \gamma$ (see (2.1.16)). The parameter η, which may take any real value, is a measure of the lateral curvature of the axisymmetric flow. When $\eta = 0$ the flow is laterally uniform whilst when $\eta = 1$ the flow has a half sine lateral structure and is zero at $y = 0$ and 1.

The SWC wavenumber k_s is given by

$$k_s^2 = \pi^2/B + \pi^2(\eta^2 - 1)\,. \tag{2.2.26}$$

For a given value of B, the azimuthal wavenumber of the SWC clearly increases with η. There is a SWC for any value of η, including values substantially greater than 1. The dependence on η of the Burger number of the SWC, at a given fixed $k = k_1$, which is relevant to the thermal Rossby number of the upper axisymmetric transition, is entirely different as may be seen by re-arranging (26)

$$B_s = \pi^2/\{k_1^2 + \pi^2(1 - \eta^2)\} \tag{2.2.27}$$

This shows that for $\eta < 1$ there is a SWC corresponding to an upper axisymmetric transition and that B_s increases with η, but that $B_s \to \infty$ as $\pi^2\eta^2 \to \pi^2 + k_1^2$. For larger values of η the SWC wavenumber k_s according to (26) is larger than k_1 whatever the value of B so that all values of B lie on the long wavelength side of the SWC.

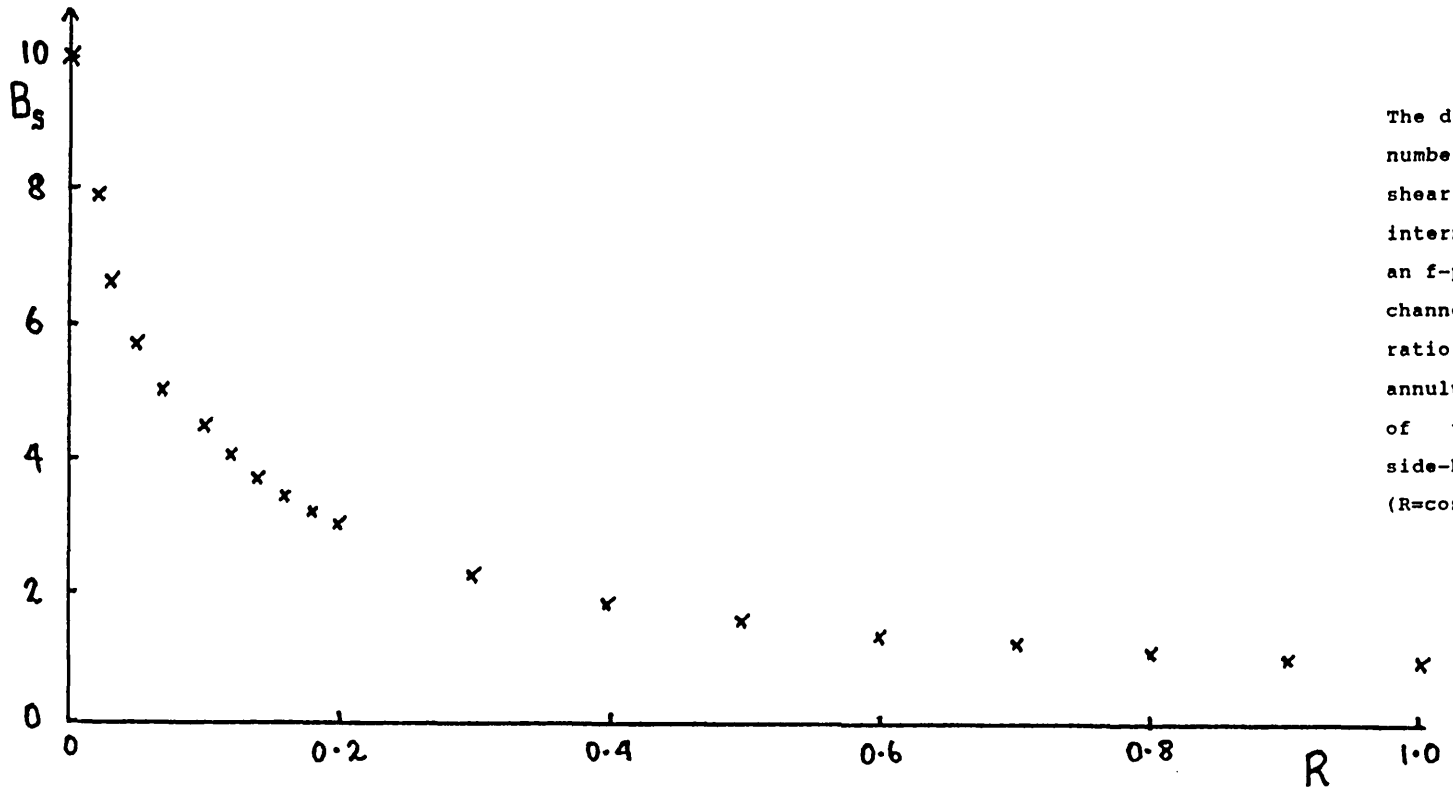


Figure 2.2

The dependence of the SWC Burger number (B_s) on a flow's lateral shear. The flow is the sinusoidal internal jet $u=-\frac{1}{4}\cos\pi z\,\cosh\eta\pi(y-\frac{1}{4})$ in an f-plane channel with $k_1=1$. This channel has a horizontal aspect ratio representative of the small annulus; see (29). R is the ratio of the zonal flow at the side-boundary to that in mid-channel ($R=\cosh\eta\pi/2$).

According to (27), the ratio of the Burger number at the SWC for a sinusoidal flow with half sine lateral shear ($\eta = 1$) to that with no lateral shear ($\eta = 0$) is

$$\frac{B_s(\eta=1)}{B_s(\eta=0)} = \frac{K_1^2 + \pi^2}{K_1^2} \quad . \tag{2.2.28}$$

From (2.1.12a) $k_1 = k_1^* L$, where k_1^* is the dimensional wavenumber of an azimuthal wavenumber one disturbance and L is the width of the channel. $k_1^* L_x = 2\pi$ where L_x is the repeat length of the channel. Modelling a cylindrical annulus with inner radius a and outer radius b by such a channel it is reasonable to take $L = b - a$ and $L_x = \pi (b + a)$. The "small annulus" investigated by Hignett (1985) and others has $b + a \simeq 2.(b - a)$. So

$$K_1 = K_1^* L = K_1^* L_x \, L/L_x \simeq \frac{2\pi(b-a)}{\pi(b+a)} \simeq 1 \, . \tag{2.2.29}$$

Combining (28) with (29) gives a Burger number for a flow with half sine lateral shear 11 times greater than that for a laterally uniform flow. The ratio (28) is clearly sensitive to $S \equiv (b-a)/(b+a)$, being proportional to S^{-2} when $S \ll \pi$.

The major impact of the half sine lateral profile on the Burger number, B_s, of the transition makes it appropriate to consider the dependence of B_s on the ratio, R, of the zonal flow at the geostrophic side-boundary to that in mid-channel. For flows of the form (25)

$$R \equiv \frac{u(z=1, y=1)}{u(z=1, y=1/2)} = \cos \eta \pi/2 \tag{2.2.30}$$

and B_s is given by (27). Figure 2.2 displays B_s plotted as a function of R for $0 \leq R \leq 1$ using $k_1 = 1$. Between R = 1 and R = 0.4, B_s varies by less than a factor of two, whilst for $R < 0.4$ B_s is very sensitive to R increasing rapidly as R decreases.

A second set of examples of sinusoidal internal jets is

$$u = -1/2 \cos \pi z \sin \eta \pi (y - 1/2) \tag{2.2.31}$$

for which k_s is again given by (26). These flows resemble quite

closely the axisymmetric flows found in internally heated annuli (Read 1985).

A final set of examples of sinusoidal flows is

$$u = -1/2 \cos \pi z \cosh \eta \pi (y-1/2). \qquad (2.2.32)$$

For these flows

$$\alpha^2 = \pi^2/B - \pi^2\eta \; ; \; K_s^2 = \pi^2/B - \pi^2(1+\eta^2). \qquad (2.2.33)$$

The hyperbolic shear stabilises the flow resulting in SWCs at smaller wavenumbers k_s, or smaller Burger numbers B_s, as η is increased. The flows (32) have minimum values at mid-channel and maximum values at the geostrophic boundaries. White (1986) has used them to represent the mean zonal flow obtained in the large amplitude wave regime (see chapter four).

An interpretation of the SWCs for the sinusoidal internal jets based on an anti-cascade argument is presented in section 2.4.

Example 2: Separable flows on an f plane

Consider axisymmetric flows which are separable in the sense that

$$u(y,z) = G(y)\,H(z). \qquad (2.2.34)$$

On an f plane the choice $u_s = 0$ in (3) gives

$$T_s = -1/G \; d^2G/dy^2 - 1/H \; d/dz \,(1/B \; dH/dz) \qquad (2.2.35)$$

Solutions of (4), (5) and (35) may be sought as linear combinations of separable solutions $f = \sigma(y).\chi(z)$ for which

$$d^2\sigma/dy^2 - \sigma/G \; d^2G/dy^2 - \mathcal{B}\,\sigma = 0$$
$$\sigma = 0 \quad \text{at } y = 0,1 \qquad (2.2.36)$$

and

$$d/dz\,(1/B \; d\chi/dz) - \chi/H \; d/dz\,(1/B \; dH/dz) - \nu\chi = 0$$
$$d\chi/dz = 0 \quad \text{at } z = 0,1 \qquad (2.2.37)$$

with

$$K_s^2 = \nu + \xi \,. \qquad (2.2.38)$$

(36) - (38) could be explored in some detail but we limit ourselves to two investigations into the influence of lateral curvature on the transition.

First let $H(z) = -\frac{1}{2}\cos \pi z$ and $B^2(z) = B_s^2$, as before, and

$$G(y) = 1 - a + a \sin \pi y \,. \qquad (2.2.39)$$

Then $k_s^2 = \pi^2 / B_s + \xi$ where ξ is the first eigenvalue of (36) and (39). We have calculated B_s as a function of the ratio of the flow at the side to the middle, $R = 1 - a$, at fixed $k_s = 1$ for comparison with fig. 2.2. The two curves are barely distinguishable, suggesting that fig. 2.2 is insensitive to the details of the lateral curvature assumed.

Further support for this suggestion derives from the second example in which we examine (36) for flows which vanish at the side boundaries

$$G = 0 \quad \text{at} \quad y = 0,1 \qquad (2.2.40a)$$

and have

$$G > 0 \quad \text{for} \quad 0 < y < 1 \,. \qquad (2.2.40b)$$

Substitution of $\sigma = G$ and $\xi = 0$ in (36) reveals this combination to be a solution of (36) for flows which satisfy (40a). $\sigma = G$ is the gravest mode solution when the flow satisfies (40b). So it is apparent that separable flows (34) with $T_s > 0$ which are zero only at the side boundary (40b) have SWCs at wavenumbers $k_s = \sqrt{\nu}$, which do not depend on the details of the lateral structure of the zonal flow.

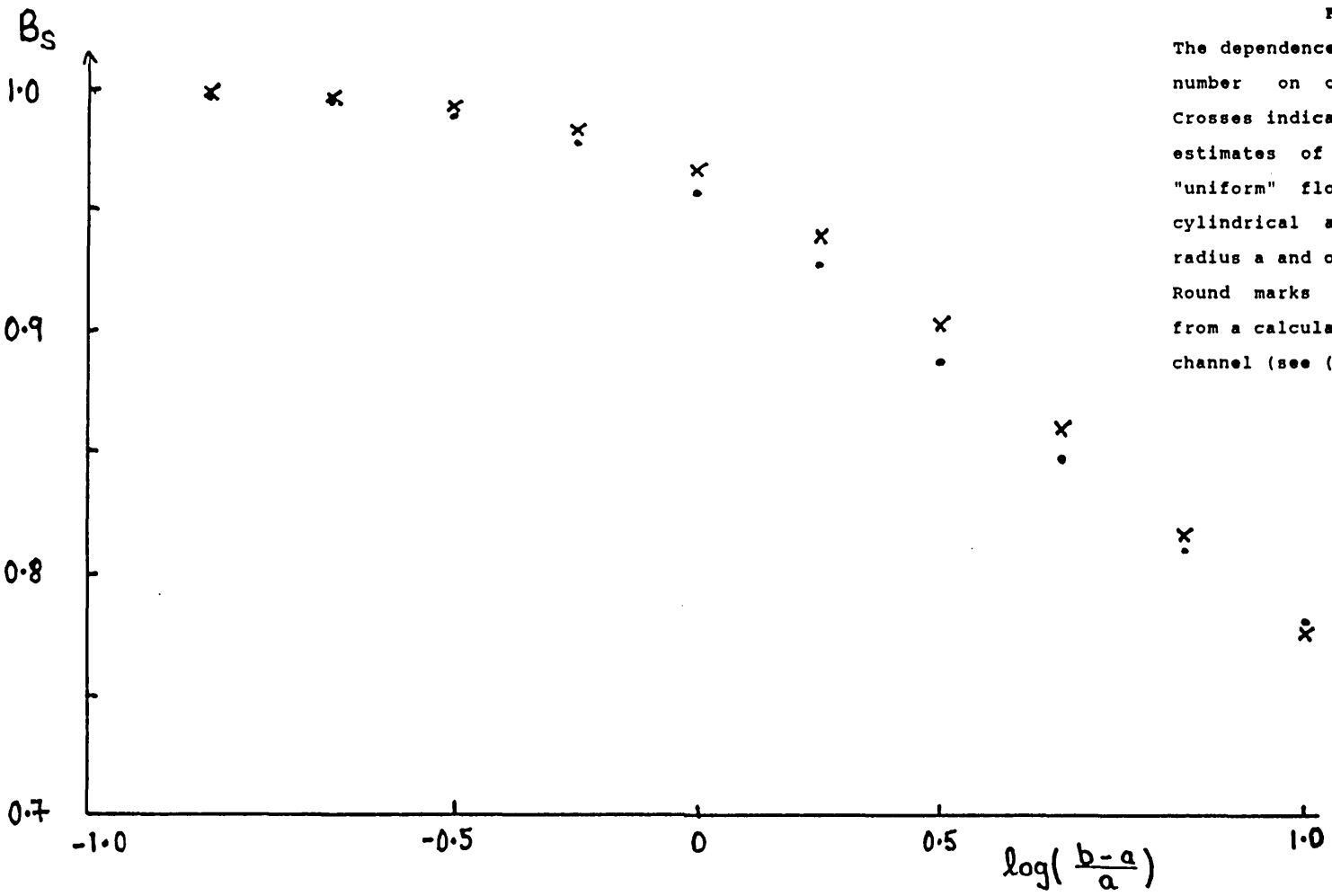


Figure 2.3

The dependence of the SWC Burger number on cylindrical curvature. Crosses indicate accurate numerical estimates of B_s for the laterally "uniform" flow $u=-\frac{1}{4}r\cos\pi z$ in a cylindrical annulus with inner wall radius a and outer wall radius b. Round marks indicate B_s estimated from a calculation for a Cartesian channel (see (47)).

Example 3: Laterally uniform flows in cylindrical domains

The equations corresponding to (1) and (2) for a cylindrical domain with side boundaries at $r^* (= a\, r) = a, b$ are

$$(u-rc)\left(1/r\, \partial/\partial r (r\, \partial\phi/\partial r) + \partial/\partial z \left(f^2 a^2/(N^2 H^2)\, \partial\phi/\partial z \right) - m^2\phi/r^2 \right) = -q_r \phi$$

$$\phi = 0 \quad \text{on } r = 1, b/a \; ; \quad \partial\phi/\partial z = 0 \quad \text{on } z = 0, 1 \tag{2.2.41}$$

in which ϕ and $\partial q / \partial r$ are related to ψ' and $u(r,z)$ by

$$\begin{aligned} \psi' &= \phi(r,z) \exp im(\theta - ct) \\ q_r &= -\partial/\partial r \left(1/r\, \partial/\partial r (r u) \right) - \partial/\partial z \left(f^2 a^2/(N^2 H^2)\, \partial u/\partial z \right) . \end{aligned} \tag{2.2.42}$$

Howard's argument can be repeated for these equations without difficulty.

The SWC problem for the analogue of the laterally uniform sinusoidal internal jet

$$u = -1/2\, r \cos \pi z \tag{2.2.43}$$

with uniform stratification is

$$\begin{aligned} & 1/r\; d/dr \left(r\, d\phi_s/dr \right) - m^2 \phi_s / r^2 = -\pi^2 \phi_s\, f^2 a^2/(N^2 H^2) \\ & \phi = 0 \quad \text{at} \quad r = 1, b/a . \end{aligned} \tag{2.2.44}$$

It is convenient to re-express (44) in terms of a normalised radial ordinate y, the ratio, g, of the gap width to the inner radius and the Burger number B:

$$y = \frac{r-1}{b/a - 1} \quad ; \quad g = \frac{b-a}{a} \quad ; \quad B = \frac{N^2 H^2}{f^2 (b-a)^2} \tag{2.2.45}$$

In these terms (44) is

$$\begin{aligned} & d^2\phi_s/dy^2 + (y + 1/g)^{-1}\, d\phi_s/dy - m^2 \phi_s (y+1/g)^2 = -\pi^2 \phi_s / B \\ & \phi_s = 0 \quad \text{at} \quad y = 0, 1 . \end{aligned} \tag{2.2.46}$$

The cross marks on figure 2.3 display the dependence of the SWC Burger number of wavenumber one (m=1) solutions of (46) over the range $-1 < \log g < 1$. These results were derived using the shooting method described in section 2.6 with a 100 point resolution. The

round marks on the same figure are the corresponding cut-off values for disturbances in a Cartesian channel of length $\pi(b+a)$ and width $b-a$. The plotted values are inferred from (27) with $\eta=0$, (29) and (45)

$$B = \pi^2/(\pi^2 + K_1^2) \quad ; \quad K_1 = 2g/(g+2). \tag{2.2.47}$$

The two curves correspond quite closely showing that the crude representation of the cylindrical annulus as a Cartesian channel does not introduce large errors in the calculation of the upper axisymmetric transition.

Implications for the interpretation of the Upper Axisymmetric Transition

The volume average of the Burger number of an axisymmetric flow is closely related to the imposed thermal Rossby number. The relationship can be made precise by defining the volume average value of $\partial T/\partial z$ to be $\sigma_z\ \Delta T/H$. $\sigma_z \simeq 0.67$ for many axisymmetric flows (Hide 1967). Then

$$B = N^2H^2/(f^2L^2) \simeq \frac{g\alpha\sigma_z \Delta T H}{4\Omega^2 L^2} = \sigma_z \Theta/4 \tag{2.2.48}$$

and $\Theta \simeq 6B$. The UAT for the jet $u = -\frac{1}{2}\cos\pi z$ thus occurs at $\Theta \simeq 6$. This value is much larger than that of the UAT at moderate Taylor numbers (see fig. 1.4) but in excellent agreement with the main transition at higher Taylor numbers according to figs. 4 & 10 of Hide & Mason (1978).

The main transition for $u = -\frac{1}{2}\cos\pi z \sin\pi y$ occurs for values of $\Theta > 60$, which is at loggerheads with experimental findings. The calculation for separable flows which vanish at the side boundary (see (40)) suggests that this high value of Θ is a feature of zonal flows which satisfy a no-slip condition at the side boundary.

The dependence of the cut-off Burger number, B_s, on the ratio, R, of the axisymmetric flow at the side to the middle of a supposed

geostrophic region is illustrated by fig. 2.2. The fact that B_s is highly sensitive to R and "unrealistic" according to the experiments when $R < 0.4$ and insensitive and "realistic" when $R > 0.4$ suggests that $R > 0.4$ in the laboratory experiments. To substantiate this claim it would be necessary to understand the nesting of the side boundary layers of the wave perturbations and of the axisymmetric flow itself. This is not attempted here (though see p. 117).

Section 2.3 FLOWS WITH BOUNDARY THERMAL GRADIENTS

The flows studied in this section have

$$\partial u/\partial z > 0 \qquad 0 \leq y, z \leq 1 \tag{2.3.1}$$

and $T_s > 0$ (see (2.1.2)) for some value of u_s within the flow domain. (1) ensures that there is no more than one steering level (where $u=c_r$) within the flow. We have chosen $\partial u/\partial z > 0$ for definiteness; any solution (ϕ, c) of (2.1.15) - (2.1.18) for given fields of u and $\partial q/\partial y$ implies a solution (ϕ^*, $-c^*$) for the fields -u and $-\partial q/\partial y$. So flows satisfying $\partial u/\partial z < 0$ everywhere and with $T_s>0$ for u_s within the flow domain, may also be considered to belong to the class under consideration.

We establish first that laterally uniform flows with

$$T_s > 0 \quad \text{for} \quad u(z=0) < u_s < u(z=1) \tag{2.3.2}$$

have inviscid SWCs (i.e. finite upper bounds on the azimuthal wavenumber, k, of their unstable inviscid normal modes) marked by neutral modes. These flows also have large Burger number cut-offs to instability provided $Bq_y = B\gamma - \partial^2 u/\partial z^2$ is bounded as $B\to\infty$. This last condition requires $B\gamma$ to be bounded as $B\to\infty$, which must be the case if (2) is to be satisfied as $B\to\infty$. The Burger number and azimuthal wavenumber cut-offs are found simultaneously by determining the cut-off value of the non-dimensional wavenumber, p:

$$p^2 = B\,(K^2 + \pi^2). \tag{2.3.3}$$

Three steps replace the single argument presented in section 2.2. The first establishes a rigorous upper bound on the wavenumber (p) of unstable disturbances. It is argued in the second that any normal mode solution with no singular critical level at wavenumber p_0, say, must be neighboured by solutions at $p_0+\delta p$ and $p_0-\delta p$ for some, sufficiently small, δp. It would follow that any unstable

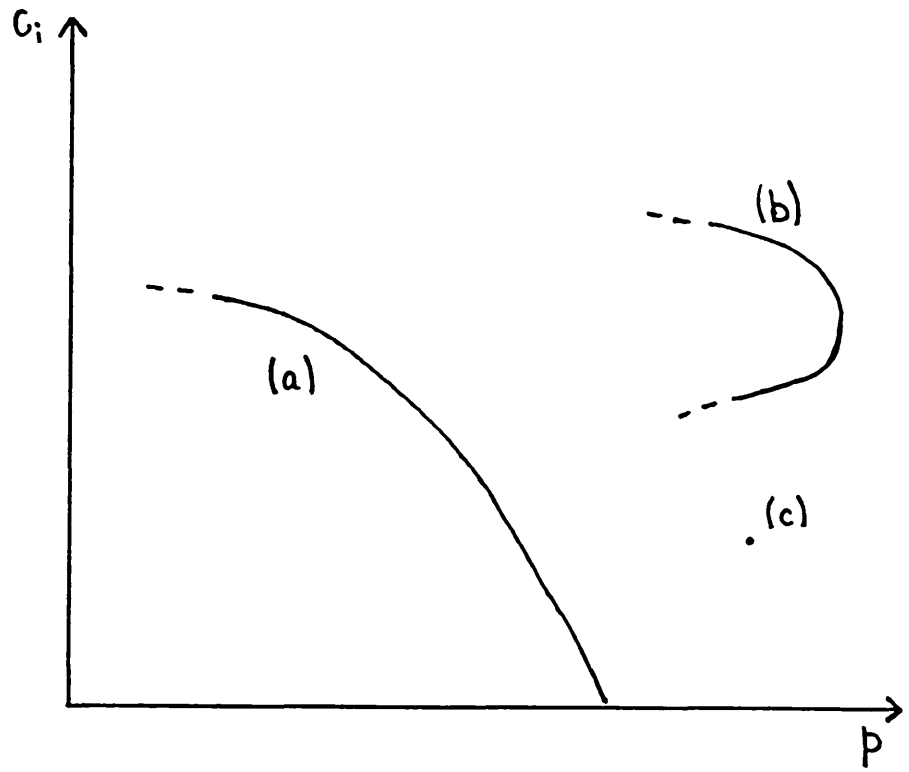


Figure 2.4

Schematic depiction of solution branches in the wavenumber p, (3), growth rate (c_i) plane. It is proved, in the second part of the main argument of section 2.3, that solution branches cannot turn in the manner of branch (b). Branches terminate, like (a), at neutral solutions (c_i=0). Isolated solutions like (c) are argued to be rare.

normal mode solution would be joined by an uninterrupted string of solutions in the (p, c_i) plane to a neutral solution of larger wavenumber, as in curve (a) on figure 2.4, because there is a finite bound on the wavenumber of unstable modes. Furthermore strings of unstable or neutral solutions could only terminate at neutral solutions with critical levels within the flow domain. The third step establishes that the neutral solution with the largest azimuthal wavenumber is given by the gravest mode solution of

$$\partial^2\chi/\partial z^2 + \{ \beta T_s - p^2 \}\chi = 0 \qquad 0 \le y, z \le 1$$

$$(u-u_s)\partial\chi/\partial z - du/dz\,\chi = 0 \quad \text{on } z=0,1 \; ; \quad \chi = 0 \quad \text{on } y=0,1. \tag{2.3.4}$$

These neutral modes are shown in section 2.2 to border marginally longer wavelength instabilities.

The argument submitted for the second step succeeds in proving that turning points like that on curve (b) of fig 2.4 are not possible but does not absolutely exclude isolated solutions like point (c) on fig 2.4. It does, however, establish that such isolated solutions can only occur for extremely rare flows (of infinite codimension).

Some examples of the dependence of the SWC on the proximity of the $\partial q/\partial y=0$ level to boundary thermal gradients then follow. The key points involved in extending the arguments to domains with slightly sloping endwalls and the few laterally sheared flows which satisfy (2) are then indicated. The section concludes with a discussion of the problem of normal modes subject to Ekman pumping which is continued in section 2.7.

A bound on the wavenumber of unstable normal modes

Multiplication of (2.1.15) by B ϕ^* / (u-c) and use of the boundary conditions in integration by parts yields a complex integral constraint on unstable normal modes whose real and imaginary parts are presented overleaf. (6) expresses a well known constraint on the meridional flux of generalised potential vorticity (Bretherton 1966a and Eliassen 1983).

$$\int_z\int_y \; B\,|\partial\phi/\partial y|^2 + |\partial\phi/\partial z|^2 + BK^2|\phi|^2\,dy\,dz$$

$$= \int_z\int_y \frac{Bq_y(u-c_r)|\phi|^2}{|u-c|^2}dy\,dz + \int_y\left[\frac{\partial u/\partial z\,(u-c_r)|\phi|^2}{|u-c|^2}\right]_{z=0}^{1} dy \qquad (2.3.5)$$

$$0 = \int_z\int_y \frac{Bq_y|\phi|^2}{|u-c|^2}\,dy\,dz + \int_y\left[\frac{\partial u/\partial z\,|\phi|^2}{|u-c|^2}\right]_{z=0}^{1} dy \; . \qquad (2.3.6)$$

(5) may be derived from a conservation theorem for the "quasi-energy" of a small (Andrews 1983) or finite amplitude (McIntyre & Shepherd 1987) disturbance to a steady flow. Constraints similar to (5) and (6) are familiar from inviscid shear flow stability (Drazin & Reid 1981 section 22).

An upper bound on the wavenumber p of unstable normal modes on such flows can be derived using (5). By Poincare's inequality

$$\text{l.h.s. (2.3.5)} \geqslant p^2 I \quad ; \quad I = \int_z\int_y |\phi|^2\,dy\,dz \; . \qquad (2.3.7)$$

All contributions to the first term on the r.h.s. of (5), other than those near the steering level, may be bounded by a finite multiple of I. The boundary terms on the r.h.s. of (5) may give a significant contribution if the steering level lies close to one of the horizontal boundaries. The key to this sub-section is that (6) may be used to bound the contributions to the r.h.s. of (5) from a region containing the steering level in terms of contributions well away from it. This point is made rigorously by (8) below. Not surprisingly, boundary terms in (5) which are not close to the steering level can be bounded by a finite multiple of I when p is large enough (see (13) & (17) below). The value of the upper bound on p hence established, (19) below, is not particularly significant, because the SWC is located precisely in the next sub-section; tightness of the bound has been sacrificed in favour of clarity of presentation.

To proceed with the proof we divide the flow domain into regions A_R, B_R, D_R and E_R, see fig. 2.5, whose boundaries lie on isotachs of the zonal flow - which are horizontal when the flow is laterally uniform. To abbreviate notation let u_0=u(z=0), u_1=u(z=1),

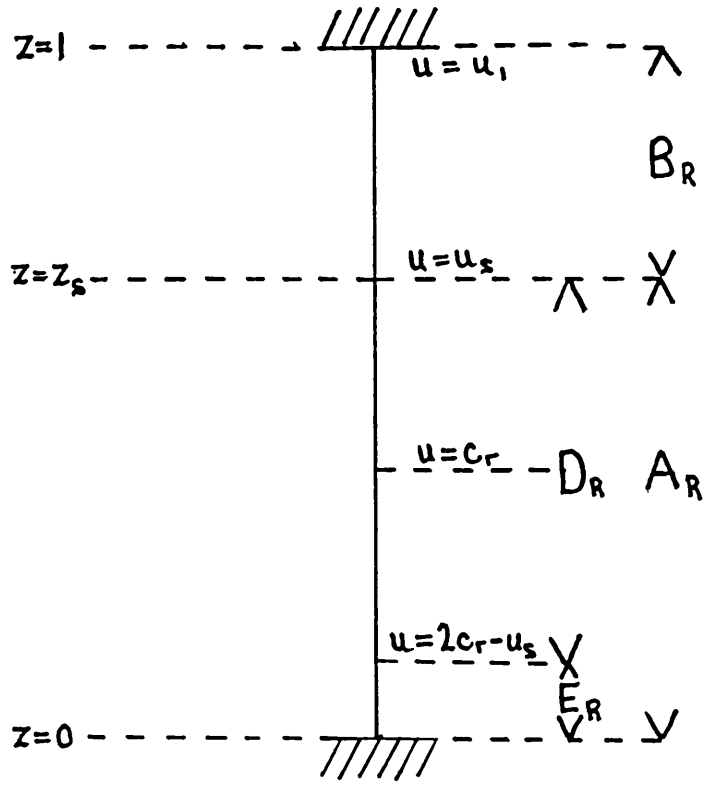


Figure 2.5

The division of the flow domain used to infer an upper bound on the wavenumber of unstable normal modes (see inequalities (10) and (19)). Regions A_R and B_R, which are separated by the isotach $u=u_s$ on which $\partial q/\partial y=0$, cover the whole of the flow domain and its boundaries. Region A_R lies on the same side of $u=u_s$ as the steering level $u=c_r$. Region D_R covers the part of A_R which lies between $u=u_s$ and $u=2c_r-u_s$. Region E_R covers any part of A_R left uncovered by D_R.

$u_{z0}=\partial u/\partial z(z=0)$, $u_{z1}=\partial u/\partial z(z=1)$ and $u_s=u(z=z_s)$. For a given complex phase speed c, region D_R extends between $z = z_s$ (where $\partial q/\partial y = 0$) and the level where $u = 2c_r - u_s$. Region A_R contains region D_R and extends from z_s to the lower boundary if $c_r < u_s$, or to the upper boundary if $c_r > u_s$. Region B_R is the complement of A_R w.r.t. the whole domain whilst region E_R is the part of A_R which is not in D_R; if $2c_r - u_s < u_0$ or $2c_r - u_s > u_1$ then region E_R will be empty. Fig. 2.5 is drawn for a case with $c_r < u_s$ and a non-empty region E_R. In the above informal discussion, D_R is the region containing the steering level and B_R and E_R the regions in which the first term on the r.h.s. of (5) is easily bounded by a finite multiple of I.

The regions have been chosen so that

(i) all contributions to the r.h.s. of (6) from B_R are of one sign and all from region A_R, and hence D_R, of the opposite sign

(ii) all contributions to the r.h.s. of (5) from B_R are of one sign

(iii) $|u-c_r|$ in $B_R \geq |u_s-c_r| \geq |u-c_r|$ in D_R.

Denoting the upper and lower boundaries of B_R and D_R by B_{bdy} and D_{bdy}, the following inequalities result from points (i) - (iii) and (6)

$$\left| \iint_{B_R} \frac{Bq_y(u-c_r)|\phi|^2}{|u-c|^2}\,dy\,dz \;+\; \int_{B_R}\left[\frac{\partial u/\partial z\,(u-c_r)|\phi|^2}{|u-c|^2}\right]_{B_{bdy}} dy \right|$$

$$\geqslant |u_s-c_r| \left| \iint_{B_R} \frac{Bq_y|\phi|^2}{|u-c|^2}\,dy\,dz \;+\; \int_{B_R}\left[\frac{\partial u/\partial z\,|\phi|^2}{|u-c|^2}\right]_{B_{bdy}} dy \right|$$

$$\geqslant |u_s-c_r| \left| \iint_{D_R} \frac{Bq_y|\phi|^2}{|u-c|^2}\,dy\,dz \;+\; \int_{D_R}\left[\frac{\partial u/\partial z\,|\phi|^2}{|u-c|^2}\right]_{D_{bdy}} dy \right|$$

$$\geqslant \left| \iint_{D_R} \frac{Bq_y(u-c_r)|\phi|^2}{|u-c|^2}\,dy\,dz \;+\; \int_{D_R}\left[\frac{\partial u/\partial z\,(u-c_r)|\phi|^2}{|u-c|^2}\right]_{D_{bdy}} dy \right| \qquad (2.3.8)$$

The first inequality follows from (i)-(iii), the second from (6) & (i) and the third from (iii). So the contribution to the r.h.s. of (5) from D_R is less than that from B_R. Contributions to the first terms on the r.h.s. of (5) from B_R and E_R are easily bounded in terms of I ;

$$\iint_{B_R} \frac{B q_y (u-c_r)|\phi|^2}{|u-c|^2}\,dy\,dz \le \max_{0\le y,z\le 1}\{BT_5\} \iint_{B_R} |\phi|^2\,dy\,dz$$

$$\iint_{E_R} \frac{B q_y (u-c_r)|\phi|^2}{|u-c|^2}\,dy\,dz \le 2 \max_{0\le y,z\le 1}\{BT_5\} \iint_{E_R} |\phi|^2\,dy\,dz \quad . \tag{2.3.9}$$

Substitution of (7) - (9) in (5) shows that

$$p^2 \int_z \int_y |\phi|^2\,dy\,dz \le 2\,\mathrm{Max}\{BT_5\} \int_{B_R+E_R} |\phi|^2\,dy\,dz + \left[\int_y \frac{\partial u/\partial z (u-c_r)|\phi|^2\,dy}{|u-c|^2}\right]_{E_R dy} + 2\left[\int_y \frac{\partial u/\partial z (u-c_r)|\phi|^2\,dy}{|u-c|^2}\right]_{B_R dy} \tag{2.3.10}$$

To bound the horizontal boundary contributions by a finite multiple of I note first that region D_R covers the boundary belonging to A_R if $2c_r - u_s < u_0$ or if $2c_r - u_s > u_1$. So only lower boundary terms with $c_r > (u_s + u_0)/2$ and upper boundary terms with $c_r < (u_s + u_1)/2$ need further consideration; boundary terms on the r.h.s. of (5) involving almost singular critical layer contributions will originate from D_R and can be bounded by the contribution from region B_R, as demonstrated above (see (8)).

Consider the lower boundary first. We will show that when p is large enough ((13) below) $|\phi|$ does not decay rapidly with height near the boundary ((14) below) and hence that the lower boundary contribution is smaller than a certain constant times I ((17) below). From (2.1.18)

$$\left(d/dz \int_y |\phi|^2\,dy\right)_{z=0} = -2\left(\int_y \frac{(c_r-u)\partial u/\partial z\,|\phi|^2\,dy}{|u-c|^2}\right)_{z=0} \ge \frac{-2u_{z0}}{c_r-u_0}\left(\int_y |\phi|^2\,dy\right)_{z=0} \quad \text{when } c_r > u_0 \tag{2.3.11}$$

Addition of $\phi^*/(u-c)$ times (2.1.15) and $\phi/(u-c^*)$ times $(2.1.15)^*$, integration from 0 to 1 w.r.t. y and from 0 to Z w.r.t. z yields

$$\left[d/dz \int_0^1 |\phi|^2\,dy\right]_0^Z = 2\int_0^Z \int_0^1 \left(|\partial\phi/\partial z|^2 + p^2|\phi|^2 - \frac{B q_y (u-c_r)|\phi|^2}{|u-c|^2}\right) dy\,dz \tag{2.3.12}$$

Let z_F be the level of the isotach $u = u_0 + (u_s - u_0)/4$ and region F_R be confined by $z = 0$ and that isotach. When $c_r > (u_s + u_0)/2$, $3|u-c_r| > |u-u_s|$ within F_R and $|Bq_y(u-c_r)|/|u-c|^2 < 3BT_s$. So the r.h.s. of (12) is positive within F_R if

$$p^2 > 3 \max_{0\le y,z\le 1} BT_s \tag{2.3.13}$$

in which case

$$\left(d/dz \int_0^1 |\phi|^2 dy \right)_{z\in F_R} \geq \left(d/dz \int_0^1 |\phi|^2 dy \right)_{z=0} \tag{2.3.14}$$

and using (11)

$$\left\{ \int_0^1 |\phi|^2 dy \right\}_{z\in F_R} \geq \{ 1 - 4u_{z0} z/(u_s-u_0) \} \left\{ \int_0^1 |\phi|^2 dy \right\}_{z=0} . \tag{2.3.15}$$

By integration of (15) and manipulation of simple integration formulae one finds that

$$\iint_{E_R} |\phi|^2 dy dz \geq 1/2 \, \mathrm{Min}\{ z_F, (u_s-u_0)/4u_{z0} \} \left(\int_y |\phi|^2 dy \right)_{z=0} \tag{2.3.16}$$

and hence

$$-\left(\int_y \frac{\partial u/\partial z (u-c_r)|\phi|^2 dy}{|u-c|^2} \right)_{\substack{z=0 \\ \text{in } F_R}} \leq H_0 \iint_E |\phi|^2 dy dz \tag{2.3.17}$$

$$H_0 \equiv 4u_{z0}/(u_s-u_0) \cdot \mathrm{Max}\{ 1/z_F, 4u_{z0}/(u_s-u_0) \} .$$

An entirely similar argument for $c_r < (u_s + u_1)/2$ at the upper boundary yields

$$\left(\int_y \frac{\partial u/\partial z (u-c_r)|\phi|^2 dy}{|u-c|^2} \right)_{\substack{z=1 \\ \text{in } F_R}} \leq H_1 \iint_E |\phi|^2 dy dz \tag{2.3.18}$$

$$H_1 \equiv 4u_{z1}/(u_1-u_s) \cdot \mathrm{Max}\{ 1/z_G, 4u_{z1}/(u_1-u_s) \}$$

where z_G is the level of the isotach $u = u_1 + (u_s - u_1)/4$. The same arguments repeated for the boundary terms of B_R easily yield bounds corresponding to (17) and (18) in which $H_0/2$ and $H_1/2$ take the place of H_0 and H_1. Finally, use of (10), (13) (17) and (18) and elimination of the integral I shows that

$$p^2 \leq 3 \max_{0\leq y,z\leq 1} \{BT_\zeta\} + H_0 + H_1 . \qquad (2.3.19)$$

In conclusion, unstable modes on flows satisfying (1) and (2) must satisfy (19).

Solution branches do not terminate at unstable modes

Assume that Φ is a solution of (2.1.15) - (2.1.18) with eigenvalue C at wavenumber k = K, which does not have a singular critical level. For convenience we look for neighbouring solutions at $K^2 + \delta k^2$ with B unchanged. Setting

$$\phi = \Phi + \delta\phi \quad , \quad c = C + \delta c \quad , \quad k^2 = K^2 + \delta k^2 \qquad (2.3.20)$$

and substituting into (2.1.15) - (2.1.18), without any linearisation one finds that

$$\begin{aligned} L\delta\phi &\equiv \left(\partial^2/\partial y^2 + \partial/\partial z(1/B\,\partial/\partial z) - K^2 + q_y/(u-C) \right)\delta\phi \\ &= \delta k^2(\Phi+\delta\phi) - q_y(\Phi+\delta\phi)\sum_{i=1}^{\infty} \frac{\delta c^i}{(u-C)^{i+1}} . \end{aligned} \qquad (2.3.21)$$

This equation must be solved subject to

$$\begin{aligned} &\delta\phi = 0 \quad \text{on} \quad y=0,1 \\ &(u-C)\partial\delta\phi/\partial z - \partial u/\partial z\,\delta\phi = \delta c(\partial\Phi/\partial z + \partial\delta\phi/\partial z) \text{ on } z=0,1 . \end{aligned} \qquad (2.3.22)$$

Posing a power series expansion for $\delta\phi$ in terms of δk^2 and δc

$$\begin{aligned} \delta\phi &= \phi_{1K}\delta k^2 + \phi_{1c}\delta c + \phi_{2KK}(\delta k^2)^2 + \phi_{2Kc}\delta k^2\delta c + \phi_{2cc}\delta c^2 + \ldots . \\ &= \delta\phi_1 + \delta\phi_2 + \ldots . \end{aligned} \qquad (2.3.23)$$

and assuming that δk^2 and δc are of the same order of magnitude, $\delta\phi_1$ must satisfy

$$\begin{aligned} &L\delta\phi_1 = \delta k^2\Phi - \delta c\, q_y\Phi/(u-C)^2 \equiv W \\ &\delta\phi_1 = 0 \text{ on } y=0,1 \quad ; \quad (u-C)\partial/\partial z(\delta\phi_1) - \partial u/\partial z\,\delta\phi_1 = \delta c\,\partial\Phi/\partial z \text{ on } z=0,1 . \end{aligned} \qquad (2.3.24)$$

As shown in appendix A, the necessary and sufficient condition for there to be a solution for $\delta\phi_1$ is that

$$\int_z\int_y W\Phi\, dydz \;-\; \int_y \left[\frac{\Phi\,\delta c\,\partial\Phi/\partial z}{B(u-C)}\right]_0^1 dy = 0 \qquad (2.3.25)$$

Substitution of (24) into (25) yields a condition for solution of the form

$$\alpha_{11}\,\delta c \;+\; \alpha_{12}\,\delta K^2 = 0 \qquad (2.3.26)$$

in which α_{11} and α_{12} are complex valued constants which depend on Φ, C and p (but not on δk^2 or δc). These constants will typically only be zero at isolated wavenumbers K on solution branches and will be non zero for all unstable modes for a generic subset of zonal flows. Unless $\alpha_{11} = 0$ a complex value δc will satisfy (26) for any choice of δk^2 and solutions will neighbour Φ on both long and short wave sides. So only cases with $\alpha_{11} = 0$ need further attention. For these cases one must solve (21) and (22) to higher orders in δk^2 and δc. Substituting $\delta\phi_1$ given by (23) into the r.h.s of (21) and (22) the condition for $\delta\phi_1 + \delta\phi_2$ to have a solution is of the form

$$\alpha_{11}\delta c \;+\; \alpha_{12}\delta K^2 \;+\; \alpha_{21}\delta c^2 + \alpha_{22}\delta c\,\delta K^2 + \alpha_{23}(\delta K^2)^2 = 0\,. \qquad (2.3.27)$$

The condition for solution with $\delta\phi_1+\delta\phi_2+..\delta\phi_n$ may be written as a polynomial in δc (of order $\leq n$) whose coefficients involve δk^2, α_{11}, α_{12}, etc. For any chosen (positive or negative) value of δk^2 this polynomial will have as many solutions for δc as the order of the polynomial in δc. The order will be non-zero provided at least one coefficient of a (non-zero) power of δc is non zero. Hence quadratic turning points, like that on curve (b) of figure 2.4, which have no neighbours on the short wave side, are not possible on any zonal flow. Only isolated solutions and turning points like

$$\delta K^2 = \exp(\delta c^2) \qquad (2.3.28)$$

are not ruled out.

Such turning points may be described as rare in a special sense. Let μ be an independent parameter and $(\Xi(y,z;\mu), C(\mu))$ be pairs of

smooth functions involving non-singular phase velocities. α_{11} may be viewed as a function of (Ξ,C) and hence of μ determined by (24) and (25);

$$-\alpha_{11} = \int_z\int_y \frac{\Xi^2 q_y}{(u-C)^2}\,dy\,dz + \int_y \left[\frac{\Xi^2 \partial u/\partial z}{B(u-C)^2}\right]'_{z=0} dy .$$

α_{11} may be differentiated w.r.t. μ an arbitrary number of times. All the derivatives vanish together only for pairs of (Ξ,C) of infinite codimension in an appropriate space of pairs of functions. So the exceptional points (28) are of infinite codimension in this sense. This formal description of the rareness of the isolated points is subject to the characteristic weakness of such arguments, namely that the function space considered is not confined to solutions of the problem (i.e. of (2.1.15) - (2.1.18) in this case).

Comparison of neutral modes

Consider the value of the r.h.s. of (6) as a neutral solution, whose critical level lies within the fluid, is approached. The contribution from the critical layer becomes of order $(q_y|\phi|^2)(u=c_r) / c_i$ as $c_i \to 0$. If the critical level is near a horizontal boundary the area integral and boundary contribution have the same sign there by assumptions (1) and (2). Furthermore since $\partial u/\partial z > 0$ there is only one critical level. So (6) can only be satisfied as $c_i \to 0$, and by the neutral solution itself, if $q_y|\phi|^2 = 0$ at the critical level.

The argument above is applicable only to neutral solutions which are neighbours to unstable ones because (6) was derived on the assumption that $c_i \neq 0$. But Bretherton (1966a) has shown, by consideration of the dispersion of fluid particles near the critical level, that $q_y|\phi|^2 = 0$ at the critical level of any neutral solution. Hence only three types of neutral modes are possible:

Type 1: modes with critical levels at $\partial q/\partial y = 0$ (i.e. $c_r = u_s$)
Type 2: modes with steering levels outside the flow ($c_r < u_0$ or $c_r > u_1$ as in external Rossby waves in Green's problem)
Type 3: modes with $\phi = 0$ at the critical level.

Let the gravest mode solution of type 1 be χ_s with $p = p_s$. To prove that there are no neutral modes of types 2 or 3 with wavenumber $m > p_s$ let us assume that ψ is such a solution of type 2 or 3 with real phase velocity c_r

$$\partial^2\psi/\partial z^2 + \left(Bq_y/(u-c_r) - m^2 \right)\psi = 0 \qquad 0 \le z \le 1$$

$$(u-c_r)\partial\psi/\partial z - \partial u/\partial z\,\psi = 0 \quad \text{on } z=0,1 \; ; \quad \psi = 0 \quad \text{on } y = 0,1 . \tag{2.3.29}$$

We choose the phases of ψ and χ_s so that both are real, multiply (29) by χ_s, subtract ψ multiplied by (4) (with $\chi = \chi_s$ and $p^2 = p_s^2$) and integrate from $z = z_1$ to $z = z_2$ using $q_y = T_s(u-u_s)$ to obtain

$$\left[\chi_s\partial\psi/\partial z - \psi\partial\chi_s/\partial z\right]_{z_1}^{z_2} = \int_{z_1}^{z_2} \psi\chi_s\left\{ m^2 - p_s^2 + BT_s\frac{(u_s-c_r)}{(u-c_r)} \right\} dz. \tag{2.3.30}$$

For neutral modes of type 2, $T_s(u_s-c_r)/(u-c_r) > 0$ for $0 \le z \le 1$ and the factor multiplying $\chi_s\psi$ in the integrand on the r.h.s. of (30) is positive. If ψ is of one sign throughout $0 \le z \le 1$ we take $z_1 = 0$, $z_2 = 1$ and choose ψ and χ_s to be positive so that the r.h.s. of (30) is positive. The l.h.s. of (30) is then

$$\left[\chi_s\partial\psi/\partial z - \psi\partial\chi_s/\partial z\right]_0^1 = \left[\frac{\partial u/\partial z\;\chi_s\psi(c_r-u_s)}{(u-c_r)(u-u_s)} \right]_0^1 \tag{2.3.31}$$

which, for solutions of type 2 ($c_r < u_0$ or $c_r > u_1$), is negative. If ψ changes sign within the flow domain we integrate from $z = 0$ to the first level at which ψ changes sign. Then the r.h.s. of (30) is still positive and the l.h.s. is

$$\left[\chi_s\partial\psi/\partial z - \psi\partial\chi_s/\partial z\right]_0^{z_1} = \chi_s\partial\psi/\partial z\big|_{z_1} - \left(\chi_s\partial\psi/\partial z - \psi\partial\chi_s/\partial z\right)_{z=0}. \tag{2.3.32}$$

$\partial\psi/\partial z < 0$ at $z=z_1$ so the l.h.s. of (30) is again negative. We must conclude that the sought solutions of type 2 with $m > p_s$ cannot satisfy (30) and so do not exist.

For neutral modes of type 3, $T_s(u_s-c_r)/(u-c_r) > 0$ either (a) above $u=c_r$ if $u_s > c_r$ or (b) below $u=c_r$ if $u_s < c_r$. Taking case (a) first, consider (30) with $z_2 = 1$ and z_1 at the highest level at which ψ changes sign. Since $u(z_1) \ge c_r$ and χ_s and ψ may be taken to be positive in the chosen interval the r.h.s. of (30) is positive.

On the other hand the upper boundary term on the l.h.s. of (30) is negative (by (31)) as is the lower boundary term $-\chi_s(z_1)\,\partial\psi/\partial z(z_1)$. For case (b) setting $z_1 = 0$ and z_2 at the lowest level where ψ changes sign results in another contradiction.

In conclusion, neutral modes of types 2 and 3 above with wavenumber $m \geq p_s$ are not possible and the gravest mode solution of (4) must be the shortest wave neutral mode solution of (2.1.15) - (2.1.18).

SWCs of laterally uniform flows with boundary thermal gradients

The location of the SWC of internal jets with $T_s > 0$ is shown in section 2.2 and fig. 2.2 to be sensitive to the lateral curvature of the zonal flow. The following examples show that the cut-off location is also sensitive to the separation of the $\partial q/\partial y=0$ contour from a boundary with a thermal gradient.

Consider first

$$u = -\tfrac{1}{2}\cos\alpha\pi z \;;\; \tfrac{1}{2} < \alpha \leq 1\,,\; \gamma = 0. \qquad (2.3.33)$$

When $\alpha=1$ the flow is an internal jet with $\partial q/\partial y=0$ at midlevel ($z=\frac{1}{2}$). As α is decreased, $\partial u/\partial z$ at $z=1$ becomes increasingly positive and the $\partial q/\partial y=0$ level moves up towards the upper boundary. For arbitrary α, however, $u_s=0$, $T_s=\alpha^2\pi^2$ and $\partial u/\partial z=0$ at $z=0$, so the SWC solutions of (4) have

$$\chi = \cosh\eta z \;;\; \eta\tanh\eta = -\alpha\pi\tan\alpha\pi \;;\; p^2=\eta^2+\alpha^2\pi^2. \qquad (2.3.34)$$

When $\alpha=1$, the solution of (34) has $\eta=0$ and $p=\pi$, in agreement with previous solutions. According to (34) as $\alpha \to \frac{1}{2}$ from above, $\eta \to \infty$. The values of $\log(p^2)$ for the SWC as a function of α are marked by crosses on fig. 2.6. The amplitude of the streamfunction as a function of height, normalised so that $\phi=1$ at $z=1$, is also displayed by crosses on fig. 2.7 for two values of α. The neutral modes with large values of p obtained with α close to 0.5 are trapped at the upper boundary.

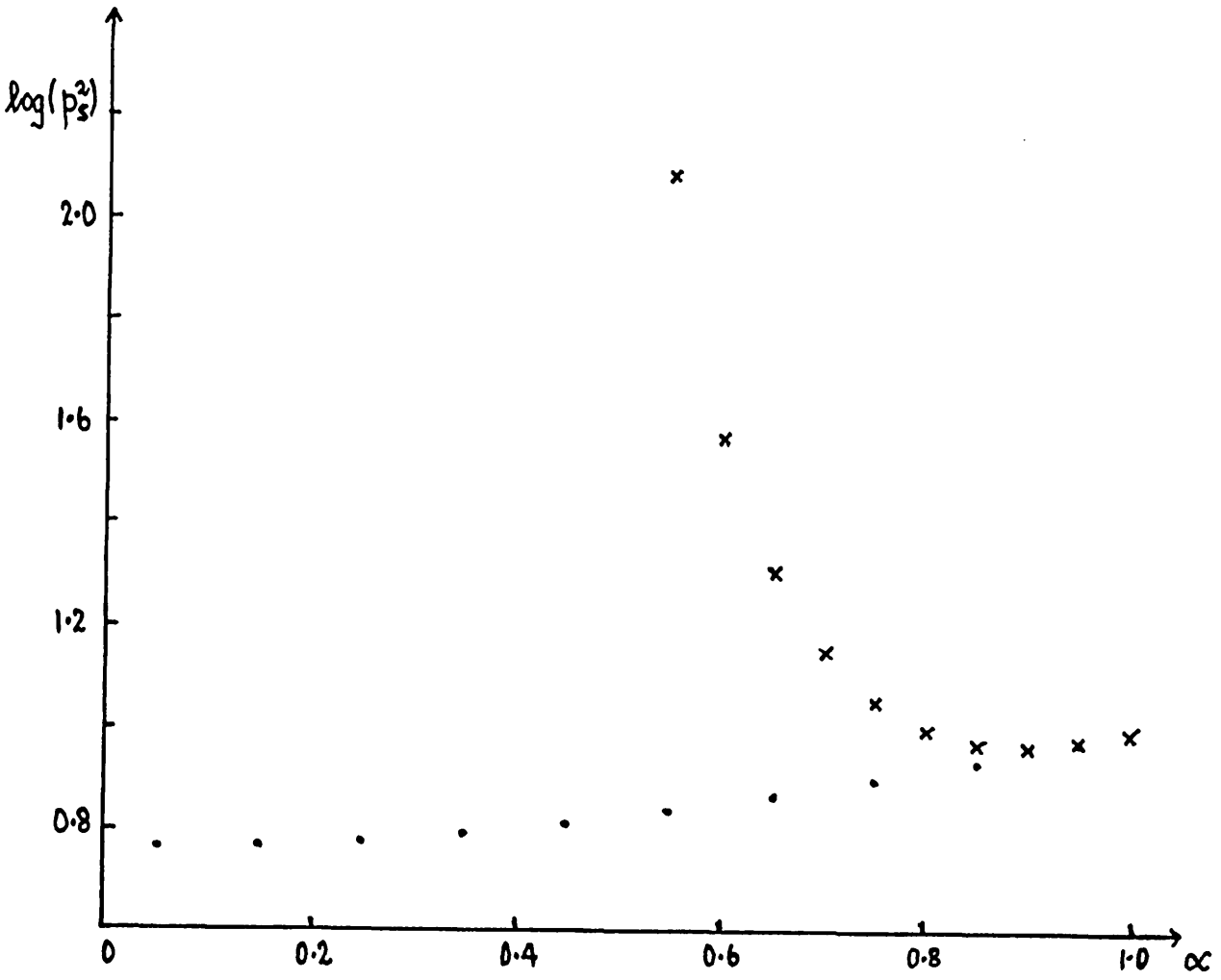


Figure 2.6

The dependence of the SWC wavenumber on the height of the $\partial q/\partial y=0$ line. The SWC wavenumber p_s is shown for two f-plane flows:
(i) $u=-\frac{1}{2}\cos\alpha\pi z$ (crosses)
(ii) $u= 1/(\alpha\pi)\ \sin\alpha\pi(z-\frac{1}{2})$ (circles).
The $\partial q/\partial y=0$ level tends to the upper boundary as $\alpha\downarrow\frac{1}{2}$ in case (i) but remains at mid-level for all α in case (ii).

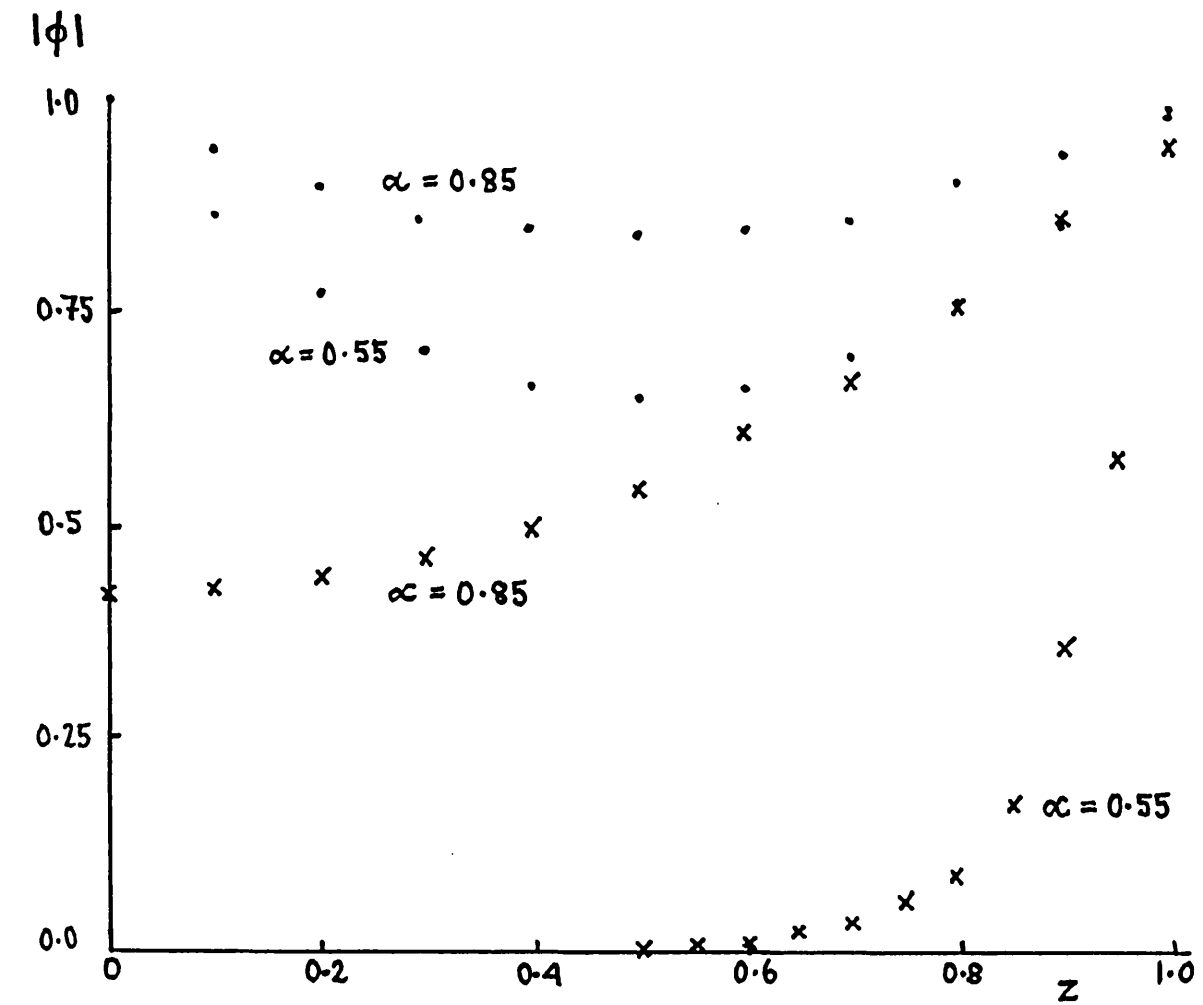


Figure 2.7

The height dependence of the amplitudes of the normal modes of fig 2.6. The amplitudes are only displayed for α=0.55 and α=0.85. The normal mode solution for case (i) (crosses) with α=0.55 has wavenumber p=11 (see (34)) and is clearly trapped near the upper boundary.

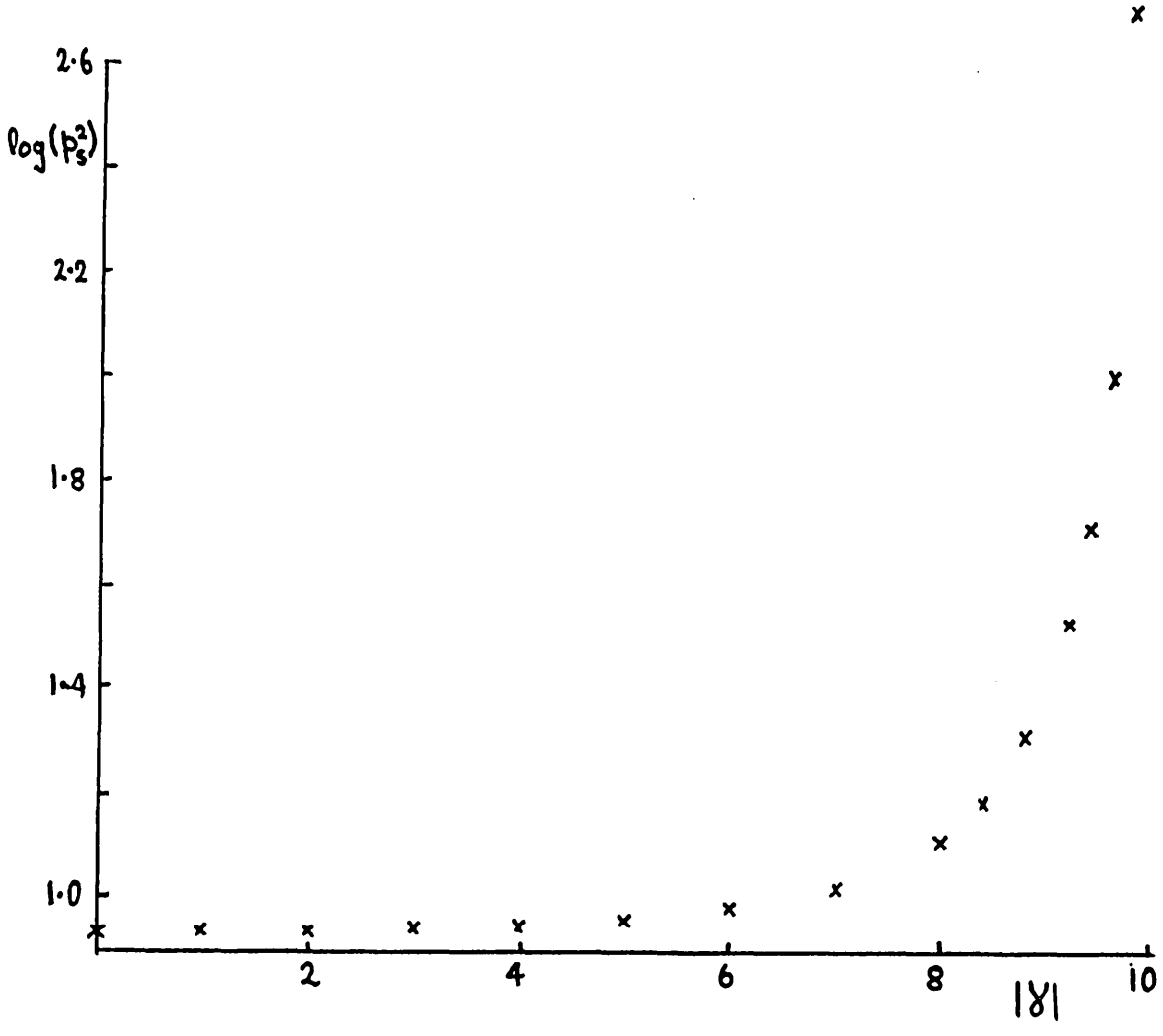


Figure 2.8

A second example of the dependence of the SWC on the height of the $\partial q/\partial y=0$ line. The SWC wavenumber p_s for the flow $u=z-\cos\pi z$ is shown as a function of the magnitude of the beta parameter γ defined in (2.1.16). The cut-off wavenumber is the same for $-\gamma$ as $+\gamma$. The flow has thermal boundary gradients and its $\partial q/\partial y=0$ level tends to the upper or lower boundary as $|\gamma|\uparrow\pi^2$.

A code named WAVENO which solves (4) by the shooting method has been checked against these analytical results and used, with a resolution of 100 points in the vertical, to investigate several other flows . The SWC wavenumber p_s and streamfunction structure for

$$u = \frac{1}{\alpha\pi} \sin\alpha\pi(z-\tfrac{1}{2}) \quad ; \quad 0 < \alpha \leq 1 \quad , \quad \gamma = 0 \qquad (2.3.35)$$

are plotted as circles on figs. 2.6 & 2.7 respectively for comparison with the results for (33). Apart from an arbitrary scale factor, the flow (35) is identical with (33) when $\alpha=1$ in both formulae, but the $\partial q/\partial y=0$ level remains at $z=\frac{1}{2}$ independent of α in (35) and the Eady flow $u=z$ is obtained in the limit as $\alpha \to 0$. It is clear from figs. 2.6 & 2.7 that for flow (35) the SWC is relatively insensitive to α and that the neutral modes are symmetric about mid-level and not trapped near the boundaries.

Fig. 2.8 displays the dependence of the wavenumber of the SWC on γ for

$$u = z - \cos\pi z \quad , \quad 0 \leq |\gamma| < \pi^2 \qquad (2.3.36)$$

When $\gamma=0$, $\partial q/\partial y=0$ lies at mid-level, but as γ increases to π^2, $\partial q/\partial y=0$ moves down to reach the lower boundary where there is a boundary thermal gradient which provides a generalised potential vorticity gradient of the opposite sign to the gradient in the bulk of the interior. The limiting behaviour of flows (33) and (36) is consonant with the lack of SWCs when boundary thermal gradients provide generalised potential vorticity gradients opposite in sign to those in the interior, which was anticipated by Bretherton (1966a) and partly confirmed by McIntyre (1970). The growth rates near the SWCs with large wavenumbers, which could be studied using the codes developed for section 2.7, might be of some interest but have not been investigated.

Slightly sloping boundaries

An important variant on the standard annulus experiments is obtained by sloping the upper and/or lower boundaries of the container (Mason 1975). (2.1.15) - (2.1.17) are still applicable and when the slope is shallow (2.1.18) can be adapted to capture the vertical motion of a fluid parcel sliding along the sloping boundary; the appropriate boundary condition is

$$(u-c)\partial\phi/\partial z - (\partial u/\partial z - \frac{B}{Ro}\, dh/dy)\,\phi = 0 \qquad \text{on } z=0,1 \qquad (2.3.37)$$

for upper and lower boundaries lying on z = h(y,0) and z = 1 + h(y,1) respectively with Ro denoting the Rossby number (2.1.6). The SWC argument given in this section can be repeated if

$$(\partial u/\partial z)' \equiv \partial u/\partial z - \frac{B}{Ro}\, dh/dy > 0. \qquad (2.3.38)$$

$(\partial u/\partial z)'$ determines the sign of the "effective" potential vorticity gradient on the boundaries. In the horizontal wall experiments $\partial u/\partial z$ is small at the boundaries. If $dh/dy > 0$ and $\partial u/\partial z = 0$ then the generalised potential vorticity gradient on the boundary would be of the opposite sign to that by the boundary in the interior. This would almost certainly result in short wavelength trapped instabilities resembling those found by Green (1960) (see also Bretherton 1966a and equation (5.5) of McIntyre 1970). These instabilities would not appear in Eady's problem augmented by sloping boundaries because $u = z$ has no interior potential vorticity gradients. Short wave instabilities were not reported by Mason (1975); perhaps $\partial u/\partial z$ near the boundary is affected by the boundary slope, or (38) only holds for very shallow slopes, or the waves eluded detection by the thermocouple array.

Zonal flows with lateral shear

There are two points that make the baroclinic stability problem considerably more difficult for zonal flows with vertical and lateral shear than for the laterally uniform flows considered thus far.

Firstly laterally uniform flows which satisfy (2) as $B\to\infty$ have $B\gamma$ bounded as $B\to\infty$. So their stability depends chiefly on p; the dependence of the growth rates on B at constant k (which is relevant to the interpretation of the rotating annulus experiments) and on k at constant B can be considered simultaneously. This is no longer the case for laterally sheared flows. For the SWC argument the most serious aspect of this difference is that

$$Bq_y = B\gamma - B\,\partial^2u/\partial y^2 - \partial^2u/\partial z^2 \qquad (2.3.39)$$

does not remain bounded as $B\uparrow\infty$ if $\partial^2u/\partial y^2 \neq 0$ or if γ rather than $B\gamma$ is held constant as B varies. The argument that the equality (5) cannot be satisfied at large values of B is hence difficult. So the argument of section 2 is repeated here only for problems with k variable and B fixed. In practice, however, the results can be used to retrieve information about the Burger number of the SWC at fixed azimuthal wavenumber for given flows, by making use of evaluations of the azimuthal wavenumber of the SWC at fixed Burger number performed over a wide range of Burger numbers. Plotting the locus of the SWC on a phase diagram (see Bell & White 1988a section 5) in which B k is plotted on the abscissa and B π on the ordinate enables the Burger number of the SWC for a chosen k to be deduced.

Secondly whereas laterally uniform flows have horizontal and hence parallel isotachs and contours of $\partial q/\partial y$ in the meridional plane, isotachs and $\partial q/\partial y$ contours of most laterally sheared flows intersect each other. The arguments using (5) and (6) require the $\partial q/\partial y = 0$ contour to be coincident with an isotach. Only a minority (non-generic subset) of zonal flows satisfy this condition. Nevertheless it seems worth extending the arguments already presented to flows of this type. We will show that a laterally and vertically sheared zonal flow u(y,z) has a SWC if

$$\text{(a)}\quad \mu \le \partial u/\partial z \le M\ ;\quad 0\le y,z\le 1\ ,\ \mu, M > 0 \qquad (2.3.40)$$

$$\text{and (b)}\quad T_s \equiv q_y/(u-u_s) > 0 \quad \text{in } 0\le y,z\le 1 \text{ for a } u_s \qquad (2.3.41)$$

$$\text{such that}\quad u_{ox} \equiv \text{Max } u(y,z=0) < u_s < \text{Min } u(y,z=1) \equiv u_{IN}\ .$$

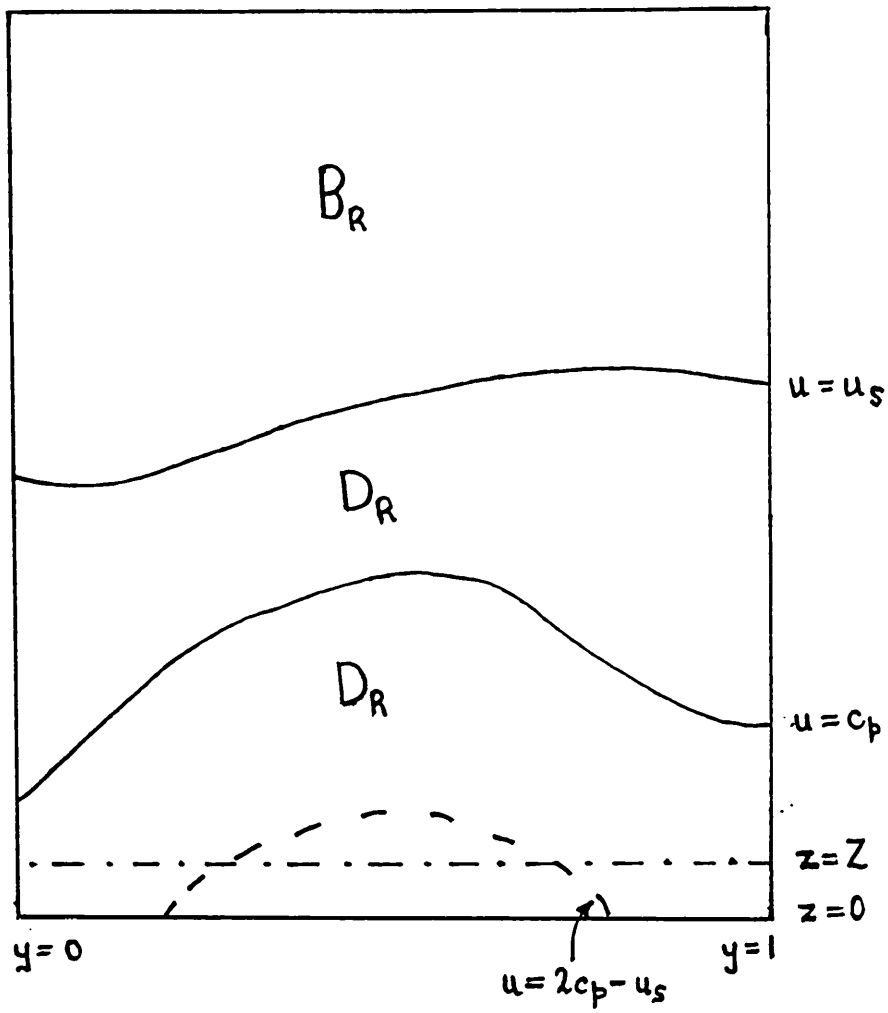


Figure 2.9

Re-definition of the regions of figure 2.5 for a laterally sheared flow. As before A_R and B_R are separated by the isotach $u=u_s$ and together cover the whole domain. A_R again lies on the same side of $u=u_s$ as the steering level, $u=c_p$. The previous definition of D_R, that it cover all of A_R in which $2c_p-u_s \leq u \leq u_s$ (assuming $c_p < u_s$), suffers from the fact that for some values of c_p (as in the figure) D_R would cover some but not all of the lower boundary. D_R and E_R are re-defined so that either i) D_R covers all of the lower boundary or ii) E_R covers a region $0 \leq z \leq Z$ (where Z is independent of y).

Furthermore if the gravest mode solution ϕ_s with eigenvalue k_s of

$$\partial^2\phi/\partial z^2 + B\,\partial^2\phi/\partial y^2 + B(T_s - K^2)\phi = 0$$
$$\phi = 0 \quad \text{on} \quad y = 0,1 \;;\; (u-u_s)\partial\phi/\partial z - \partial u/\partial z\,\phi = 0 \quad \text{on} \quad z=0,1 \qquad (2.3.42)$$

is of one sign throughout $0 \leq y,z \leq 1$ the cutoff is marked by the neutral mode $\phi = \phi_s$ with wavenumber k_s and phase speed $c = u_s$. The proof of these statements requires three amendments to the arguments already given.

1) The definitions of the boundaries of regions A_R to E_R need a little adjustment. The problem, see figure 2.9, is that for some phase speeds, such as $c = c_p$ in the figure, region D_R covers some of the (lower) boundary but not all of it. The contribution to the r.h.s. of (5) from the portion of the boundary not covered by D_R is not easy to bound using (12) which involves an area integral over a rectangle in the y,z plane between y=0 and y=1 and z=0 and some value Z. One way out of this difficulty is to extend region D_R so that it covers the whole of the (lower) boundary. Let $u_{0N} = \text{Min } u(y,z=0)$. If for some finite P

$$P\,|u_s - c_p| \geq |u_{0N} - u_s|$$

then re-defining D_R to cover all of region A_R we find that

$$P\,|u-c_p| \text{ in } B_R \geq P\,|u_s - c_p| \geq |u-c_p| \text{ in } D_R$$

and, repeating the derivation of (8), that the contribution to the r.h.s. of (5) from D_R is less than P times that from B_R. An alternative treatment may be applied when the real phase speed is larger than $u_{0x} = \max(u(y,z=0))$ (as is c_p in figure 2.9). Then for some finite Q,

$$Q\,|u_{0x} - c_p| \geq |u_{0x} - u_s|.$$

Denoting by Z the height for which

$$\max_{0\leq y\leq 1} u(y,Z) = (u_{0x} + c_p)/2,$$

E_R can be defined to be the rectangular region between z=0 and Z. In this region $2Q\,|u - c_p| > |u - u_s|$ and when

$p^2 > 2\,Q\,\mathrm{Max}_E(BT_s)$ (cf (13)), (14) can be derived and the contribution to the r.h.s. of (5) from the lower boundary bounded much as before.

One or other of these treatments will be possible for any phase speed provided there is a $c_p < u_s$, and also a $c_p > u_s$, for which both treatments are possible. This will clearly be the case given (40) and (41). The bound on k (which we do not give in detail) will be very large in some cases but, as before, it is not the size of the bound but the fact that there is one which is important.

2) The discussion of the continuity of solution branches needs some amendment because flows with lateral shear may have more than one solution of (2.1.15) - (2.1.18) with the same wavenumber k and complex phase speed c. This is not possible for the 1-dimensional problem which arises for laterally uniform flows. If there are N solutions Φ_i, $1 \le i \le N$ one must search for solutions of the form

$$\phi = \sum_{i=1}^{N} a_i \Phi_i + \hat{\delta}\phi \quad ; \quad c = C + \hat{\delta}c \quad ; \quad k^2 = K^2 + \hat{\delta}k^2$$

cf (20). The solutions must satisfy N solvability conditions (one for each Φ_i) so at each order the condition for solution may be written as a matrix equation

$$A_{ij}\, a_j = 0 \qquad (2.3.43)$$

in which the elements A_{ij} of the N by N square matrix A are polynomials in δc. (43) has solutions if and only if $\det(A) = 0$, so the condition for solution is again a polynomial in δc and the conclusions previously reached concerning the continuity of solution branches are also valid for laterally sheared flows.

3) If the gravest mode solution ϕ_s of (42) is of one sign throughout the domain the comparison arguments with other neutral modes are easily repeated. The only new feature arises when the neutral modes ψ to be compared with ϕ_s are not of the same sign for all y ($0 \le y \le 1$) at a given value of z. Integration as in (30) over any area, Ω say, in which $\psi > 0$ and on whose boundaries $\psi = 0$ yields

$$\int_y \left[\phi_s \partial\psi/\partial z - \psi \partial\phi_s/\partial z\right]_{z_1(y)}^{z_2(y)} dy = \int_z \left[B\psi \partial\phi_s/\partial y - B\partial\psi/\partial y\, \phi_s\right]_{y_1(z)}^{y_2(z)} dz$$
$$+ \iint_\Omega \psi\phi_s \left\{m^2 - \beta^2 + \frac{BT_s(u_s - c_r)}{(u-c_r)}\right\} dy\, dz \qquad (2.3.44)$$

If $y_1(z) = 0$, $\psi(y_1) = \phi_s(y_1) = 0$ and the y_1 boundary contribution to (44) is zero. Otherwise at the boundary of Ω, $\psi=0$ and $\partial\psi/\partial y(y_1)>0$, so the y_1 boundary terms make a null or positive contribution to the r.h.s. of (44). The y_2 boundary terms are similarly non-negative and the comparison arguments may be repeated without difficulty.

The arguments given above fail for laterally sheared flows which satisfy (41) well above and well below the $\partial q/\partial y=0$ curve but not in its vicinity. But even if contributions to (5) were dominated by the vicinity of a near critical steering level, (5) would still imply that

$$B(K^2 + \pi^2) < B q_{yc} / c_i \qquad (2.3.45)$$

in which q_{yc} is the maximum value of q_y on any isotach which is touched by the $q_y=0$ line. Hence the maximum growth rate kc_i associated with a critical layer could be no larger than q_{yc} / k at large wavenumbers. Arguments which apply to baroclinic internal jets of this type are presented in section 2.4.

Ekman pumping and singular neutral solutions

The influence of viscosity on the stability of zonal flows is of considerable interest; most annulus experiments, for example, are performed at moderate Taylor numbers for which diffusive effects on the waves are not entirely negligible; diffusive transports are, of course, of primary importance in the determination of the zonal flow at all Taylor numbers. The main influence of viscosity on the stability problem is expected to appear first in the upper and lower boundary layers through which it induces vertical motion at the edge of the geostrophic interior. The side boundary layers are believed to be relatively passive and less important. Zonal flows with lateral curvature require axisymmetric vertical motions in the

geostrophic interior to match the Ekman pumping at the boundaries implied by (2.1.9). It would be interesting to attempt to use the quasi-geostrophic potential vorticity conservation equation to assess the linear stability of a basic flow $(\bar{u}, \bar{w})$ with $\partial\bar{w}/\partial z = 0$.

The effect of Ekman pumping on laterally uniform flows can be investigated using (2.1.15) - (2.1.17) supplemented by

$$(u-c)\,\partial\phi/\partial z - \partial u/\partial z\,\phi + i\beta^2 \epsilon X \phi = 0 \quad \text{on } z=0,1 \tag{2.3.46}$$

$$X = E^{1/2}/(kR_o) \quad , \quad E = \nu/(2fH^2)$$
$$\epsilon = 1 \text{ on } z=0 \quad , \quad \epsilon = -1 \text{ on } z=1 \tag{2.3.47}$$

which may be derived using (2.1.9). Repeating the derivation of (6) but using (46) in place of (2.1.18) yields

$$0 = \int_z\int_y \frac{B\,q_y|\phi|^2}{|u-c|^2}\,dy\,dz + \int_y\left[\frac{|\phi|^2}{|u-c|^2}\{\partial u/\partial z - \epsilon\frac{(u-c_r)}{c_i}\beta^2 X\}\right]_{z=0}^{1} dy. \tag{2.3.48}$$

The extra term in (48) becomes unbounded as $c_i \to 0$. It foils attempts to bound the contribution to the constraint corresponding to (5) from region D_R in terms of that from region B_R. This difficulty highlights the special relationship between constraints (5) and (6) which enabled the inequalities leading to (8) to be deduced; the Ekman pumping's realignment of the phase of the streamfunction degrades the relationship. So the upper bound on unstable wavenumbers (19) need not apply to problems involving Ekman pumping. The comparisons of neutral modes also fail because Bretherton's argument that $\partial q/\partial y\,|\phi|^2 = 0$ at the critical level of any neutral solution does not hold in the presence of Ekman pumping; the boundary term in (48) can annul a singular meridional flux of potential vorticity at the critical level.

As demonstrated in appendix A, however, the arguments concerning the continuity of solution branches continue to apply when Ekman pumping is present. Furthermore one may consider perturbations to the Ekman parameter, X, rather than the wavenumber, k^2, and search for the conditions on δc for neighbouring solutions to exist. The conclusions found for perturbations of k^2 apply mutatis mutandis for perturbations of X when p is finite; except for flows of infinite

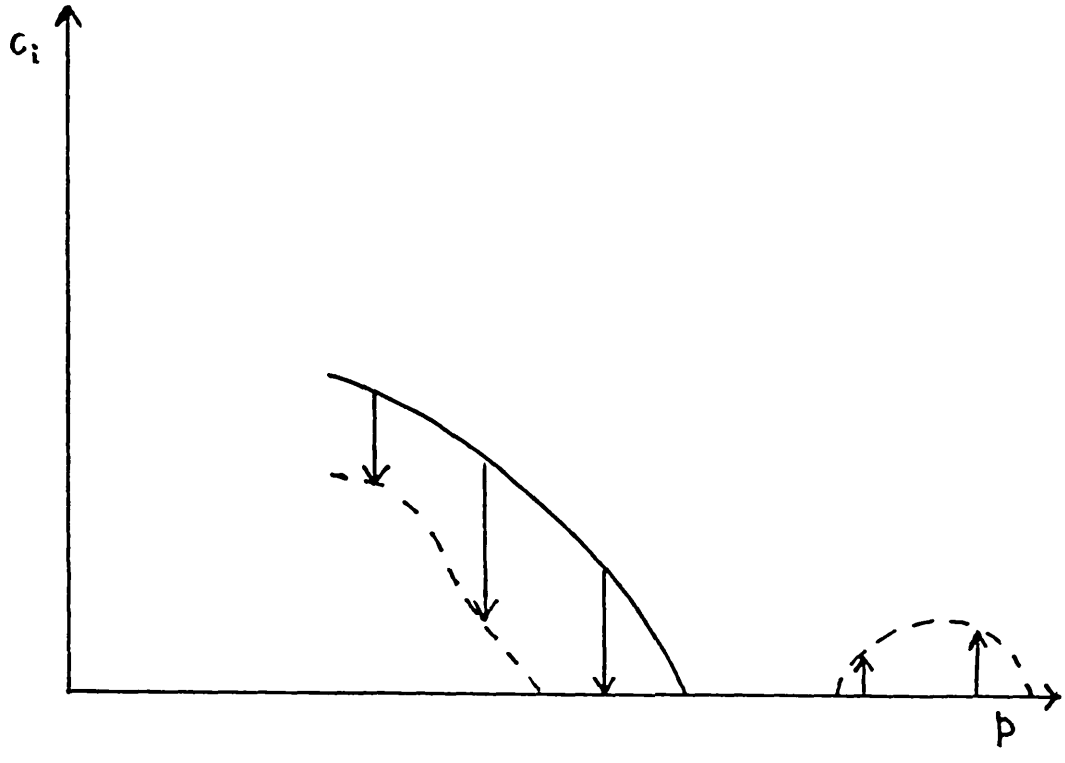


Figure 2.10

A schematic depiction of the dependence of solution branches on Ekman pumping. The solid curve represents a branch of inviscid (X=0) solutions in the (p,c_i) plane. The dashed curve depicts a possible branch of solutions for $X \equiv E^{\frac{1}{2}}/(kR_o)=X_0$. The vertical arrowed lines indicate how the X_0 branch could evolve from the inviscid solutions.

co-dimension, solution branches in the (X, c_i) plane only terminate at neutral solutions with interior critical levels. Our knowledge of the inviscid solutions limits the possibilities when Ekman pumping is present as shown on fig. 2.10. Solutions for $X = 0$ are indicated by continuous lines and solutions for an $X = X_0 > 0$ by dashed lines. The vertical lines and arrows represent the paths of the solutions as X is increased from zero to X_0 (whilst p is held constant). Values of p with solutions at $X = X_0$ but no solution at $X = 0$ (or vice versa) must have neutral solutions (with interior critical levels) at intermediate values of X. Note that it is possible that there are short wave solutions when $X = X_0$, but if there are any they must originate from neutral modes for $0 < X < X_0$.

The main issues which merit attention are

(a) the effect of Ekman pumping on the most unstable inviscid branch and its terminating neutral mode

(b) whether singular neutral modes can occur

(c) whether growing or decaying solution branches exist beyond the inviscid SWC.

Some insight into points (a) and (b) can be gained by studying the neutral solutions possessed by the sinusoidal internal jet

$$u = -1/2 \cos \pi z \tag{2.3.49}$$

on a β plane. When $\beta = 0$ there are non-singular solutions, with $c_r = 0$, of the form

$$\phi_s = \{ \cos \mathfrak{s}(z-1/2) + i\mu \sin \mathfrak{s}(z-1/2) \} \sin \pi y \tag{2.3.50}$$

which satisfy (2.1.15) - (2.1.17) and (46) when

$$\mathfrak{s}^2 = \pi^2 - p^2 \; ; \; \mathfrak{s} = 2p^2 X\mu \; ; \; \mathfrak{s}\mu = 2p^2 X . \tag{2.3.51}$$

It is clear from (51) that $\mu^2 = 1$ and hence that $|\phi_s|^2 = \sin^2 \pi y$, independent of height. The solutions tend to the inviscid solution marking the SWC as $X \to 0$ and satisfy (48) with $c_i \to 0$ since the boundary terms in (48) are of equal magnitude and opposite sign. When $\beta \neq 0$ the corresponding neutral solutions are singular for most

values of X (as may be confirmed by searching for non-singular solutions). The solutions may be found by solving the equations with $c_i \to 0$. For (49)

$$B q_y / (u-c) = \pi^2 \frac{(u + B\gamma/\pi^2)}{u-c} = \pi^2 + \pi^2 \frac{(c + B\gamma/\pi^2)}{u-c} \qquad (2.3.52)$$

Integrating (2.1.15) across a critical level z_r, $u(z_r) = c_r$, one finds that in the limit as $c_i \to 0$

$$\begin{aligned} \left[\partial\phi/\partial z \right]_{z_r^-}^{z_r^+} &= \int_{z_r^-}^{z_r^+} \partial^2\phi_s/\partial z^2 \, dz \\ &= -\pi^2 (c_r + B\gamma/\pi^2) \phi(z_r) \int_{z_r^-}^{z_r^+} dz/(u-c) \\ &= -\pi^2 (c_r + B\gamma/\pi^2) \phi(z_r) \, i\pi \, \mathrm{sign}(\delta c_i) / \partial u/\partial z(z_r) \end{aligned} \qquad (2.3.53)$$

Denoting the neutral mode below z_r by ϕ and above z_r by ψ, it must satisfy

$$\begin{aligned} -(U/2 + c_r) \partial\phi/\partial z &= -i\beta^2 X \phi \quad \text{at } z=0 \\ \partial^2\phi/\partial z^2 &= \{ \beta^2 - \pi^2 - \pi^2 (c_r + B\gamma/\pi^2)/(u-c_r) \} \phi \quad ; \quad 0 \le z \le z_r \end{aligned} \qquad (2.3.54)$$

below the critical level,

$$\begin{aligned} (U/2 - c_r) \partial\psi/\partial z &= i\beta^2 X \psi \quad \text{at } z=1 \\ \partial^2\psi/\partial z^2 &= \{ \beta^2 - \pi^2 - \pi^2 (c_r + B\gamma/\pi^2)/(u-c_r) \} \psi \quad , \quad z_r \le z \le 1 \end{aligned} \qquad (2.3.55)$$

above the critical level, and

$$\begin{aligned} \phi &= \psi \quad \text{at } z=z_r \\ \partial\psi/\partial z - \partial\phi/\partial z + i\pi^3 \, \mathrm{sign}(\delta c_i)(c_r + B\gamma/\pi^2)\phi(z_r)/\partial u/\partial z(z_r) &= 0 \quad \text{at } z=z_r \end{aligned} \qquad (2.3.56)$$

across the critical layer.

The following iterative method has been used to search for solutions of (54) - (56) with a specified value of $B\gamma$ and X by iterative adjustments to p and c_r. The error in the trial solution, with a resolution of N points, for trial values of c_r and p is calculated as follows. z_r is calculated from c_r and a uniform grid with N_1 = NINT(N.z_r) points set up below the critical level (NINT means the nearest integer). The grid's N_1th point is arranged to be at height z_r and the grid straddles the lower boundary with its 0th point at $-\Delta z/2$ where $\Delta z = z_r/(N_1 + \frac{1}{2})$. Setting ϕ=1 at the 0th point, (54) is integrated, using the standard finite difference formulae

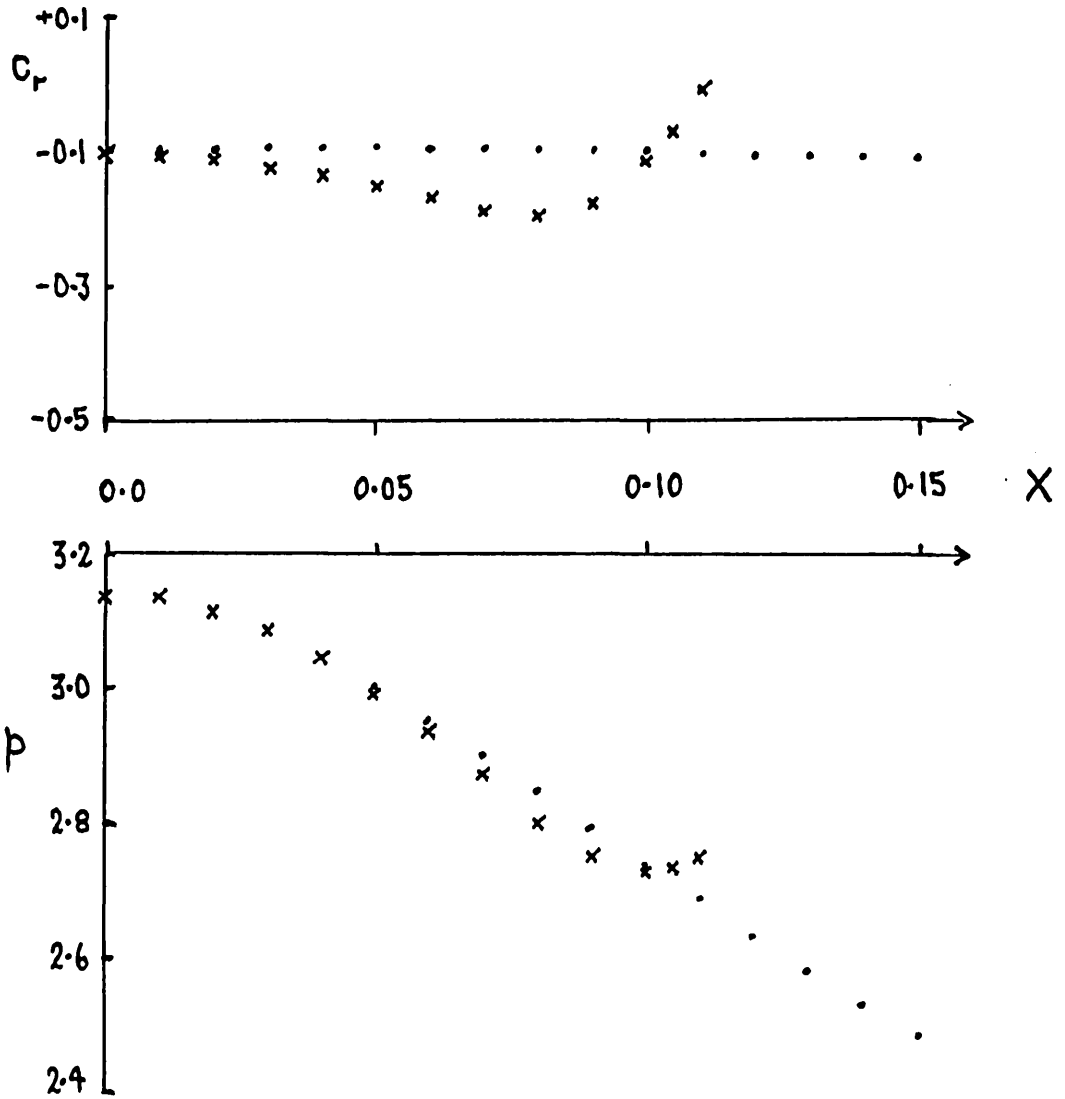


Figure 2.11

The dependence of the phase speed (c_r) and wavenumber (p) of some singular neutral normal modes on the Ekman parameter. The modes are calculated for the internal jet $u=-1/2\cos\pi z$ on a beta plane ($\gamma=1$). Growing solutions ($kc_i \downarrow 0$) are plotted as dots and decaying solutions ($kc_i \uparrow 0$) as crosses. All values are taken from finite difference numerical calculations (see text).

described in section 2.6, to give $\phi(N_1)$ as an estimate of ϕ at z_r. An estimate of $\partial\phi/\partial z$ at z_r^- is afforded by $\{\phi(N_1)-\phi(N_1-1)\} / \Delta z$. A similar grid with N_2 = NINT{ N.$(1-z_r)$} points is used to make a similar estimate of ψ and $\partial\psi/\partial z$ at z_r^+. Retaining $\phi = 1$ at $z = 0$, ψ at $z = 1$ may be determined by the condition that $\phi = \psi$ at z_r. The value of the l.h.s. of (56b) is viewed as the error in the solution for given values of p and c_r. The dependence of the error on p and c_r is determined by repeating the process with slightly different values of p and c_r. Iteration by linear extrapolation as described in section 2.6 is used to search for solutions.

Program NEUSING which performs these calculations has been checked against the non-singular solution (51) for X = 0.1. The code yields p = 2.748 using $N_1 + N_2 = 100$ and p=2.750 for $N_1+N_2 = 200$ in acceptable agreement with the analytic value of p = 2.7523. Solutions for $B\gamma = 1.0$ and $0.0 \leq X \leq 0.15$ produced by NEUSING using $N_1 + N_2 = 100$ are shown in figure 2.11. Both p and c_r are presented for solutions with $c_i \downarrow 0$ and $c_i \uparrow 0$. Solutions with $N_1 + N_2 = 200$ differ from those presented by only 0.1 %. The figure shows that the neutral mode marking the inviscid SWC moves to lower wavenumbers as X is increased; there are examples in which the cut-off moves to slightly higher wavenumbers when Ekman pumping is introduced (Holopainen 1961). The phase speeds of the modes with $c_i > 0$ tend to differ from $-B\gamma / \pi^2$ ($\simeq -0.101$) by smaller amounts than those of the "decaying" neutral modes with $c_i < 0$. It appears that there is a non-singular neutral mode with X≃0.1 at which the two solution branches cross. Unfortunately it is difficult to follow the $c_i<0$ modes beyond X=0.11 using NEUSING. These results nevertheless establish that singular neutral modes are possible on laterally uniform flows in the presence of Ekman pumping and suggest that only neutral modes with special symmetries (as in (50)) or at isolated Ekman mumbers (as in fig. 2.11) will be non-singular. Investigations of neutral and decaying solutions beyond the inviscid SWC (point (c) above) are discussed in section 2.7.

Section 2.4 A SPECTRAL MODEL OF BAROCLINIC INTERNAL JETS

Numerical simulations (using finite difference approximations to the Navier-Stokes equations) produce axisymmetric, steady wave and amplitude vacillating flows which are almost isothermal at the upper and lower boundaries (Hignett et. al. 1985, White 1986). This fact leads one to define internal baroclinic jets, more broadly than in section 2.1, as flows which satisfy

$$\partial\Psi / \partial z = 0 \quad \text{on} \quad z = 0,1. \tag{2.4.1}$$

This boundary condition is consistent with the inviscid boundary condition of no normal motion, see (2.1.9), and implies that $\partial\bar{u}/\partial z = 0$ at the boundaries.

At the side boundaries the condition of no normal flow in a Cartesian channel is

$$\partial\Psi / \partial x = 0 \quad \text{on} \quad y = 0,1. \tag{2.4.2}$$

This is trivially satisfied by the axisymmetric flow which must be constrained by an auxillary condition. It is difficult to satisfy the energy conserving boundary condition (2.1.11) in a spectral model without requiring $\bar{u}_g = 0$ or $\bar{\Psi} = 0$ at the sides. This is undesirable since it results in an unrealistic estimate of the upper axisymmetric transition (see section 2.2). A convenient, though ad hoc, condition on the the axisymmetric part of the streamfunction, $\bar{\Psi}$, is given by

$$\bar{\Psi} + \eta \, \partial\bar{\Psi} / \partial y = 0 \quad \text{on} \quad y = 0,1 \tag{2.4.3}$$

in which η is independent of both time and height (a less severe restriction is noted in (9) below). This Sturm - Liouville (S-L) type condition (Kreyszig section 4.8 1979) has merit in that it allows Ψ to be expanded in terms of the eigenfunctions of the potential vorticity operator P which satisfy the S-L type boundary conditions (1) - (3);

$$\Psi = \sum_{i=1}^{N} a_i \psi_i \,, \tag{2.4.4}$$

$$P\psi_i \equiv \partial^2\psi_i/\partial x^2 + \partial^2\psi_i/\partial y^2 + \partial/\partial z\,(1/B\,\partial\psi_i/\partial z) \tag{2.4.5}$$
$$= -\lambda_i \psi_i$$

$$\partial\psi_i/\partial x = 0 \quad \text{on } y=0,1 \;;\; \bar{\psi}_i + \eta\,\partial\bar{\psi}_i/\partial y = 0 \quad \text{on } y=0,1 \tag{2.4.6}$$
$$\partial\psi_i/\partial z = 0 \quad \text{on } z=0,1 \;;\; \psi_i(x,y,z) = \psi_i(x + L_x/L, y, z).$$

The final condition above, that ψ_i be periodic in x, takes the channel's dimensional length to be L_x. The approach to be outlined can be used with B(z) and in cylindrical geometry, but we will use Cartesian geometry and $B(z) = B_0$. The vertical dependence of the eigenfunctions is then cos mπz, the lateral dependence of the wave components sin lπy and the azimuthal dependence cos rπkx or sin rπkx where k,l and m are integers and $r = 2\,L\,/\,L_x$. When $\gamma \neq 0$ the cos rkπx and sin rkπx functions with the same vertical and lateral structure are best handled in pairs. If $a_i\psi_i$ is one of a pair let $b_i\ \phi_i$ be the other. The alternative notation which uses exp irkπx as the azimuthal dependence of the basis functions has some advantages (see (15) below).

The S-L nature of (5) - (6) ensures that the basis functions ψ_i are orthogonal in the sense that

$$\langle \psi_i\, \psi_j \rangle \equiv \iiint \psi_i\, \psi_j \, dx\, dy\, dz = \langle \psi_i\, \psi_i \rangle \delta_{ij}. \tag{2.4.7}$$

The rather severe restriction that η be independent of height in (3) can be alleviated by allowing η to depend on the vertical wavenumber. When $B(z) = B_0$, for example, the vertical eigenfunctions are cos mπz and $\bar{\Psi}$ may be expressed as

$$\bar{\Psi} = \sum_m \bar{\Psi}^{(m)}(y) \cos m\pi z \,. \tag{2.4.8}$$

The choice of $\eta = \eta^{(m)}$ may be made separately for each vertical wavenumber (of form cos mπz)

$$\bar{\psi}_i^{(m)} + \eta^{(m)}\,\partial\bar{\psi}_i^{(m)}/\partial y = 0 \quad \text{on } y=0,1 \,. \tag{2.4.9}$$

The orthogonality of the basis functions (7) can be proved using (9)

in place of (6b).

Inserting (4) into (2.1.13) and using (5) one finds that

$$\sum_{j=1}^{N} \lambda_j \, da_j/dt \, \psi_j = -\sum_{j=1}^{N}\sum_{k=1}^{N} \lambda_k \, a_j \, a_k \, J(\psi_j, \psi_k) + \sum_{j=1}^{N} \gamma \, a_j \, \partial\psi_j/\partial x \qquad (2.4.10)$$

where $J(a,b) = \partial a/\partial x \, \partial b/\partial y - \partial a/\partial y \, \partial b/\partial x$. Multiplication of (10) by ψ_i, integration over all x,y and z and use of (7) yields

$$\lambda_i \frac{da_i}{dt} \langle \psi_i \psi_i \rangle = -\sum_{j=1}^{N}\sum_{k=1}^{N} \lambda_k \, a_j \, a_k \langle \psi_i \, J(\psi_j, \psi_k) \rangle + \gamma \, b_i \langle \psi_i \, \partial\phi_i/\partial x \rangle . \qquad (2.4.11)$$

Conservation Laws and an Anti-cascade property

For arbitrary ψ_1, ψ_2, ψ_3 integration by parts and use of the b.c.s shows that

$$\begin{aligned}
\langle \psi_1 J(\psi_2 \psi_3) \rangle &= \int_z\int_y\int_x \psi_1 (\partial\psi_2/\partial x \, \partial\psi_3/\partial y - \partial\psi_2/\partial y \, \partial\psi_3/\partial x) \, dx\,dy\,dz \\
&= \int_z\int_y \left[\psi_1 \psi_2 \partial\psi_3/\partial y\right]_{x=0}^{Lx/L} dy\,dz - \int_z\int_y\int_x \psi_2 \partial\psi_1/\partial x \, \partial\psi_3/\partial y \, dx\,dy\,dz \\
&\quad - \int_z\int_y\int_x \psi_2\psi_1 \partial^2\psi_3/\partial x\partial y \, dx\,dy\,dz - \int_z\int_x \left[\psi_1\psi_2 \partial\psi_3/\partial x\right]_{y=0}^{1} dx\,dz \\
&\quad + \int_z\int_y\int_x \psi_2 \partial\psi_1/\partial y \, \partial\psi_3/\partial x \, dx\,dy\,dz + \int_z\int_y\int_x \psi_2\psi_1 \partial^2\psi_3/\partial x\partial y \, dx\,dy\,dz \\
&= \langle \psi_2 \, J(\psi_3, \psi_1) \rangle . \qquad (2.4.12)
\end{aligned}$$

Furthermore, $J(a,b) = -J(b,a)$, so $\langle \psi_1 \, J(\psi_2, \psi_3) \rangle$ has the same associative properties as the triple vector product $\underset{\sim}{a}.(\underset{\sim}{b}\wedge\underset{\sim}{c})$. (12) is useful for establishing that analogues of the energy and enstrophy are conserved by interactions between an arbitrary "triad" of components. First using (11) and (12) one finds that

$$\begin{aligned}
dE_A/dt &\equiv d/dt \; 1/2 \sum_{i=1}^{3} \lambda_i \, a_i^2 \langle \psi_i \psi_i \rangle \\
&= a_1 a_2 a_3 \{ (\lambda_2 - \lambda_3) \langle \psi_1 J(\psi_2, \psi_3) \rangle + (\lambda_3 - \lambda_1) \langle \psi_2 J(\psi_3, \psi_1) \rangle \\
&\quad + (\lambda_1 - \lambda_2) \langle \psi_3 J(\psi_1, \psi_2) \rangle \} + \text{p.t.o.}
\end{aligned}$$

$$+ \sum_{i=1}^{3} \gamma a_i b_i \langle \psi_i \partial\phi_i/\partial x \rangle \ . \qquad (2.4.13)$$

So E_A is conserved by interactions within any triad of components when $\gamma = 0$ and E_A is conserved amongst any triad of pairs (ψ_i, ϕ_i) when $\gamma \neq 0$. Similarly

$$dF_A/dt = d/dt \ \tfrac{1}{2} \sum_{i=1}^{3} \lambda_i^2 a_i^2 \langle \psi_i \psi_i \rangle$$

$$= \sum_{i=1}^{3} \lambda_i \gamma a_i b_i \langle \psi_i \partial\phi_i/\partial x \rangle \ . \qquad (2.4.14)$$

So F_A is conserved by interactions within any triad when $\gamma = 0$ or within any triad of pairs when $\gamma \neq 0$.

When the complex representation exp irkπx is used E_A and F_A are expressed by

$$E_A' = \tfrac{1}{2} \sum_{i=1}^{3} \lambda_i |c_i|^2 \langle \psi_i^* \psi_i \rangle \quad ; \quad F_A' = \tfrac{1}{2} \sum_{i=1}^{3} \lambda_i^2 |c_i|^2 \langle \psi_i^* \psi_i \rangle \qquad (2.4.15)$$

in which c_i is the complex amplitude of the complex basis function ψ_i. E_A' and F_A' are conserved within any triads.

Let $e_i = \tfrac{1}{2}\lambda_i |C_i|^2$ and consider changes Δe_i which conserve E_A' and F_A'

$$\Delta E_A' = \sum_{i=1}^{3} \Delta e_i = 0 \quad ; \quad \Delta F_A' = \sum_{i=1}^{3} \lambda_i \Delta e_i = 0 \ . \qquad (2.4.16)$$

For definiteness let $\lambda_1 < \lambda_2 < \lambda_3$. Assume first that

$$\Delta e_1 \geq 0, \ \Delta e_2 \geq 0 \text{ and } \Delta e_1 + \Delta e_2 > 0.$$

Then $\Delta e_3 < 0$ and writing $\alpha_{32} = \lambda_2 + \tfrac{1}{2}(\lambda_3 - \lambda_2)$, $\Sigma\ (\lambda_i - \alpha_{32})\ \Delta e_i < 0$ which contradicts the conservation of F_A'. A similar contradiction results if one assumes that

$$\Delta e_1 \leq 0, \ \Delta e_2 \leq 0 \text{ and } \Delta e_1 + \Delta e_2 < 0.$$

So Δe_1 and Δe_2 must have opposite signs or both be zero. Similarly Δe_3 and Δe_2 must have opposite signs or both be zero. Consequently Δe_1 and Δe_3 are of the same sign or zero. Energy can only cascade from mode ψ_2 to both ψ_1 and ψ_3 or to ψ_2 from both ψ_1 and ψ_3; it cannot cascade only to modes with lower total wavenumbers or only to modes with higher total wavenumbers. (11) only depends on the

amplitudes a_i and not on their rates of change so the instantaneous energy transfer is simply the linear superposition of each of the triad interactions. So if energy flows to higher wavenumbers it must also flow to lower wavenumbers. This anti-cascade argument was suggested by that for barotropic flows due to Fjørtoft (1953). A similar result for 3D quasi-geostrophic flow was derived by Charney (1971 & 1973).

F_A is the truncated system's analogue of the total enstrophy F of the flow

$$F = \tfrac{1}{2} \int_V (Q - f_o - \gamma y)^2 \, dV. \tag{2.4.17}$$

On truncation of (17) one obtains

$$\begin{aligned} F &= \tfrac{1}{2} \int_V \sum_{i=1}^{N} \lambda_i a_i \psi_i \sum_{j=1}^{N} \lambda_j a_j \psi_j \, dV \\ &= \tfrac{1}{2} \sum_{i=1}^{N} \lambda_i^2 a_i^2 \langle \psi_i \psi_i \rangle . \end{aligned} \tag{2.4.18}$$

$dF_A/dt = 0$ is the analogue of the full conservation law. E_A is less strictly related to E, the total kinetic and available potential energy of the flow;

$$\begin{aligned} E &= \tfrac{1}{2} \int_V \left((\partial\Psi/\partial x)^2 + (\partial\Psi/\partial y)^2 + \tfrac{1}{B} (\partial\Psi/\partial z)^2 \right) dV \\ &= \tfrac{1}{2} \iint \left[\partial\Psi/\partial x \, \Psi \right]_{x=0}^{L_x/L} dy\,dz + \tfrac{1}{2} \iint \left[\partial\Psi/\partial y \, \Psi \right]_{y=0}^{1} dx\,dz \\ &+ \tfrac{1}{2} \iint \left[\partial\Psi/\partial z . \Psi/B \right]_{z=0}^{1} dx\,dy - \tfrac{1}{2} \int_V \Psi (\partial^2\Psi/\partial x^2 + \partial^2\Psi/\partial y^2 + \partial/\partial z (\tfrac{1}{B} \partial\Psi/\partial z)) dV. \end{aligned} \tag{2.4.19}$$

On truncation E reduces to

$$E = \tfrac{1}{2} \sum_{i=1}^{N} \lambda_i a_i^2 \langle \psi_i \psi_i \rangle + \tfrac{1}{2} \iint \left[\bar{\Psi} \, \partial\bar{\Psi}/\partial y \right]_{y=0}^{1} dx\,dz. \tag{2.4.20}$$

So dE/dt is related to dE_A/dt by

$$dE/dt = dE_A/dt - \iint \sum_{i=1}^{N} \sum_{j=1}^{N} a_i \, da_j/dt \left[\bar{\psi}_i \bar{\psi}_j / \eta \right]_{y=0}^{1} dx\,dz. \tag{2.4.21}$$

When η has the same value at y = 0 and y = 1 the axisymmetric modes are either symmetric or anti-symmetric about mid-channel. Only axisymmetric modes with the same vertical structure and opposite

lateral symmetries give rise to differences between dE/dt and dE_A/dt.

Linear stability and the anti-cascade argument

The anti-cascade argument implies that a zonal flow containing only modes with total wavenumbers smaller than that of the gravest wave mode is not subject to exponentially or algebraically growing instabilities. The absence of normal mode instabilities may be proved by considering the exchange of energy between a small amplitude normal mode wave and the zonal flow to second order in the wave's amplitude, a. Corrections of $O(a^2)$ to the amplitude of the zonal flow due to the wave are included in the energetics but not of course in the linear stability analysis; their neglect in the linear stability analysis results in neglected contributions to the energy and enstrophy budget of $O(a^4)$. Corrections to the shape of the zonal flow are neglected because they also give energy contributions of $O(a^4)$.

Let $\{ w_i, 1 \leq i \leq N\}$ be the complex amplitudes of the components of a normal mode when it is of unit amplitude and W be its amplitude. Then

$$E_w = W^2 \sum_{i=1}^{N} \lambda_i |w_i|^2 = W^2 e_w \; ; \; F_w = W^2 \sum_{i=1}^{N} \lambda_i^2 |w_i|^2 = W^2 f_w \qquad (2.4.22)$$

Similarly let Z be the amplitude of the zonal flow and e_z, f_z be the energy and enstrophy of the zonal flow when of unit amplitude. Changes to the shape of the zonal flow are neglected (see above). Denoting amplitudes at t=0 by W_0, Z_0 and at t=1 by W_1, Z_1, energy and enstrophy conservation imply

$$\begin{aligned} W_0^2 e_w + Z_0^2 e_z &= W_1^2 e_w + Z_1^2 e_z \\ W_0^2 f_w + Z_0^2 f_z &= W_1^2 f_w + Z_1^2 f_z . \end{aligned} \qquad (2.4.23)$$

So

$$(W_1^2 - W_0^2) = e_z/e_w (Z_0^2 - Z_1^2) = e_z/e_w \cdot f_w/f_z \cdot (W_1^2 - W_0^2). \qquad (2.4.24)$$

But $e_w/f_w < e_z/f_z$ when the zonal modes all have smaller total wavenumber than any of the wave modes. So $W_0 = W_1$; any normal mode disturbance on the zonal flow must be neutrally stable.

The linear stability problem may be written as a real valued linear dynamical system

$$da_i/dt = X_{ij}\, a_j \tag{2.4.25}$$

in which a_i are the real valued amplitudes of all the wave modes included in the truncation and X_{ij} are the elements of a real square matrix $\underline{\underline{X}}$. All the purely real eigenvalues of $\underline{\underline{X}}$ have at least one eigenvector and are zero by the above argument. The real parts of the complex eigenvalues are also zero since the argument just given can be applied to them using $t=0$ and $t=2\pi/\lambda_i$ with λ_i the imaginary part of the eigenvalue. Any generalised eigenvectors would nevertheless be algebraically unstable. We show next that they would violate the energy and enstrophy conservation laws.

Let $\hat{\underline{x}}$ be any unit eigenvector of $\underline{\underline{X}}$ with $\lambda=0$ and $\hat{\underline{y}}$ the first unit generalised eigenvector related to $\hat{\underline{x}}$ by

$$(\underline{\underline{X}} - \lambda \underline{\underline{I}})\,\hat{\underline{y}} = p\,\hat{\underline{x}} \quad ; \quad (\underline{\underline{X}} - \lambda \underline{\underline{I}})\,\hat{\underline{x}} = \underline{0} \,. \tag{2.4.26}$$

Consider the evolution of a wave $\underline{w} = x\hat{\underline{x}} + y\hat{\underline{y}}$

$$d\underline{w}/dt = d/dt\,(x\hat{\underline{x}} + y\hat{\underline{y}}) = py\,\hat{\underline{x}} \,. \tag{2.4.27}$$

So $y(t) = y(t=0) = y_0$ and $x(t) = x_0 + p\,y_0\,t$. It is convenient to let $x_0 = 0$, to denote the components of $\hat{\underline{x}}$ by x_i and of $\hat{\underline{y}}$ by y_i and to set

$$e_x = \sum \lambda_i x_i^2 \quad ; \quad e_{xy} = \sum \lambda_i x_i y_i \quad ; \quad e_y = \sum \lambda_i y_i^2$$

and similarly for f_x, f_{xy} and f_y. Then

$$\begin{aligned} E_w + E_z &= x^2 e_x + 2xy\, e_{xy} + y^2 e_y + z^2 e_z \\ &= y_0^2 (p^2 t^2 e_x + 2pt\, e_{xy} + e_y) + z^2 e_z \,, \end{aligned}$$

$$F_w + F_z = y_0^2 (p^2 t^2 f_x + 2pt f_{xy} + f_y) + z^2 f_z .$$

From energy and enstrophy conservation

$$d/dt (z^2) = -2y_0^2 p (pt\, e_x/e_z + e_{xy}/e_z)$$

$$= -2y_0^2 p (pt\, f_x/f_z + f_{xy}/f_z)$$

which implies that

$$y_0^2 p \{ p(e_x/e_z - f_x/f_z)t + e_{xy}/e_z - f_{xy}/f_z \} = 0. \qquad (2.4.28)$$

$e_x/e_z \neq f_x/f_z$ as before so (28) cannot hold for more than one value of t unless $y_0^2 p = 0$. Hence $p = 0$ and we conclude that the energy and enstrophy conservation laws imply that there are no generalised eigenvectors with $\lambda=0$. A somewhat similar treatment of the generalised eigenvectors with pure imaginary eigenvalues (see Hirsch & Smale 1974) yields the same conclusion.

Non-linear stability

The arguments of the previous sub-section do not imply that energy cannot be extracted at all by small amplitude waves. The modes with pure imaginary eigenvalues need not be orthogonal so the total energy of a small amplitude wave containing two or more eigenvectors may fluctuate in a quasi-periodic manner with time. Fluctuations in the energy and enstrophy of large amplitude disturbances are also not covered by the preceding arguments. An upper bound on the fractional increase in enstrophy of large amplitude disturbances to an unperturbed flow containing amplitude in only the gravest mode of the system may, however, be deduced from the anti-cascade and conservation properties by considering the maximum amount of energy which could be extracted from the gravest mode. For any initial configuration the maximum amount is achieved by cascades which transfer all the perturbation's energy to the second gravest mode; any other state could still extract energy from the gravest mode. Let E and F be the initial energy and enstrophy of the disturbance , E_2 and $\lambda_2 E_2$ the final energy and enstrophy of the second gravest mode and ΔE_1 and $\lambda_1 \Delta E_1$ the change in energy and enstrophy of the gravest mode. Then

$$E_2 - E = \Delta E_1 \quad ; \quad \lambda_2 E_2 - F = \lambda_1 \Delta E_1 .$$

Elimination of E_2 reveals that

$$\Delta E_1 = (F - \lambda_2 E)/(\lambda_2 - \lambda_1) .$$

This places a clear bound on the largest possible fractional increase in the enstrophy of the disturbance;

$$|\Delta F(\text{disturbance})| < \lambda_1 F/(\lambda_2 - \lambda_1) . \qquad (2.4.29)$$

(McIntyre & Shepherd (1987) have given a very different proof of a somewhat similar result). Note that the bound on the fractional increase becomes very large when $\lambda_2 - \lambda_1$ becomes small. For sinusoidal zonal flows (see next sub-section) $\lambda_2 = \lambda_1$ at the SWC to unstable normal modes.

The above argument cannot be used when the unperturbed flow contains modes other than gravest modes; this is true even when the basic flow is zonal because wave-wave interactions could excite zonal perturbations with wavenumbers smaller than some components of the basic flow.

An interpretation of the SWC of sinusoidal flows

The sinusoidal flows defined by (2.2.20) have axisymmetric streamfunctions which are eigenfunctions of P (see (5)) with total wavenumber $\lambda = \alpha^2$. Any sinusoidal internal jet may be reproduced as a single axisymmetric mode in the spectral model by appropriate choice of η in (6). In a Cartesian channel the gravest lateral structure of a wave subject to the no-normal flow condition at the sides is $\sin \pi y$. When $B=B_0$ the internal jet boundary condition (1) allows the wave to have no vertical variation with depth. The argument given in the last sub-section hence shows that there are no unstable normal mode waves of azimuthal wavenumber k_s on a sinusoidal flow when

$$k_s^2 \geq \alpha^2 - \pi^2 . \qquad (2.4.30)$$

The analysis of section 2.2 (see (2.2.14) - (2.2.24)) shows that

this value of k_s marks the SWC to normal mode disturbances, providing $\partial q/\partial y$ changes sign within the flow domain. The analysis given here makes the importance of the no-normal flow condition for the Burger number of the SWC particularly clear. It also clarifies the sensitivity of the cut-off to the lateral curvature of the zonal flow for domains with $k_s^2 << \pi^2$. For writing $\alpha^2 = \alpha_V^2 + \alpha_H^2$ with α_V the contribution to α from the vertical curvature and α_H that from lateral curvature (30) implies that

$$\alpha_V^2 > k_s^2 + \pi^2 - \alpha_H^2 \quad \text{for instability.} \tag{2.4.31}$$

When the zonal flow has no lateral shear $\alpha_H = 0$ and the requirement for instability is $\alpha_V^2 > k_s^2 + \pi^2$, whilst when the zonal flow has half sine lateral curvature $\alpha_H = \pi$ and instability requires $\alpha_V > k_s$.

SWCs for more complex flows

The anti-cascade stability argument applied to

$$u = -1/2 \cos \pi z \, (\sin \pi y + a_3 \sin 3\pi y) \tag{2.4.32}$$

shows that the flow has no unstable normal modes with

$$K_s^2 > 8\pi^2 + \pi^2/8 \tag{2.4.33}$$

This flow has a SWC azimuthal wavenumber at all values of γ (the non-dimensional beta parameter) despite the fact that for some values of γ the $\partial q/\partial y=0$ contour lies within the domain and cuts across isotachs; for these values of γ (2.1.2) is not satisfied by any choice of $\bar{u}_s$. From (2.2.29) any Cartesian model of a cylindrical annulus will have $k_1 < 2$ so (33) is not stringent enough to determine a SWC Burger number corresponding to an upper axisymmetric transition.

The spectral model may be used to represent any zonal flow to any degree of accuracy if (9) is used as the side-boundary condition. The addition of extra modes with higher wavenumber pushes the upper bound on the SWC azimuthal wavenumber out to increasingly high

values where it is of little interest. The upper bound is definitely a gross over-estimate in many cases; for example the discussion centred on (2.2.40) shows that (32) has $k_s^2 = \pi^2/B$ (which is to be contrasted with (33)) when $\gamma=0$ and $-1/27 < a_s < 1/9$.

Some insight into this may be gleaned from the spectral model by considering perturbations to axisymmetric modes

$$\bar{\Psi}_1 = \cos\pi z \cos\pi y \quad ; \quad \bar{\Psi}_2 = \cos\pi z \cos 3\pi y, \qquad (2.4.34)$$

with amplitudes $\bar{a}_1$, $\bar{a}_2$ and total wavenumbers Λ_1, Λ_2 respectively, consisting of a mode ψ_3 with total wavenumber $\lambda_3 < \Lambda_2$ and one (or more) mode(s) ψ_4 with $\lambda_4 > \Lambda_2$. The key point is that if $I_2 \equiv \langle \psi_3 J(\bar{\Psi}_2, \psi_4) \rangle$ is non zero then $I_1 \equiv \langle \psi_3 J(\bar{\Psi}_1, \psi_4) \rangle$ will also be non zero. So if energy can cascade from $\bar{\Psi}_2$ into ψ_3 and ψ_4 it can also cascade from ψ_3 into $\bar{\Psi}_1$ and ψ_4, or in the opposite direction from $\bar{\Psi}_1$ and ψ_4 into ψ_3. When $|\bar{a}_1| >> |\bar{a}_2|$ the cascade from a_3 or a_4 into $\bar{\Psi}_1$ will shut down the instability. Let us assume that it is a_3 rather than a_4 which decays in the presence of $\bar{\Psi}_1$. Then from (11) the cascade into ψ_3 from $\bar{\Psi}_2$ and ψ_4 gives

$$da_3/dt \ \langle \psi_3 \psi_3 \rangle = (\Lambda_2 - \lambda_4) I_2 \bar{a}_2 a_4 \qquad (2.4.35)$$

whilst that from ψ_3 into $\bar{\Psi}_1$ gives

$$da_3/dt \ \langle \psi_3 \psi_3 \rangle = (\Lambda_1 - \lambda_4) I_1 \bar{a}_1 a_4 . \qquad (2.4.36)$$

$|\Lambda_1 - \lambda_4| > |\Lambda_2 - \lambda_4|$ and I_1 and I_2 are of the same order of magnitude so the decay of a_3 would be greater than its growth irrespective of the amplitude of a_4. The growth of a_4 which is proportional to a_3 will cease as a_3 is decreased to zero.

The anti-cascade property can be applied to show that some large amplitude wavenumber one free modes (White 1986) are stable to all small amplitude wave disturbances. The free modes consist of a barotropic wave, Ψ_1, with the gravest horizontal variation commensurate with the channel boundary conditions

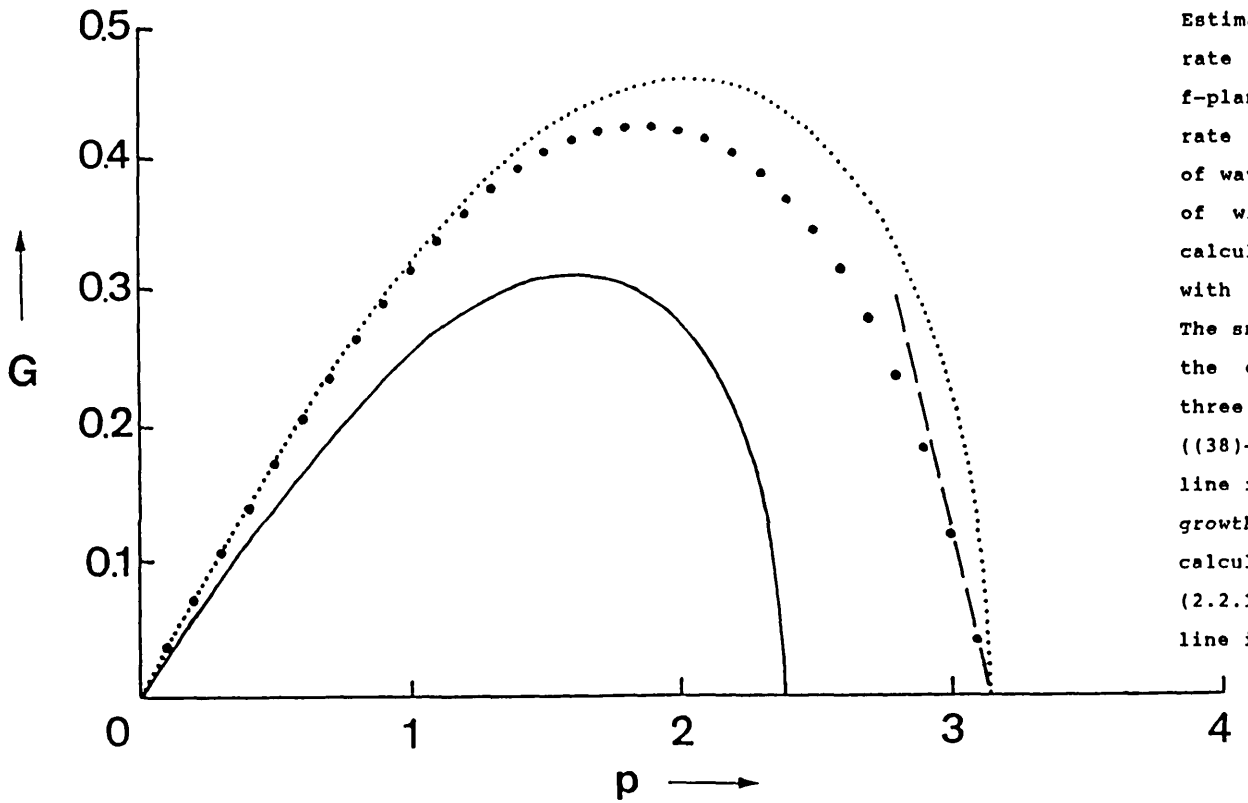


Figure 2.12

Estimates of the normal mode growth rate curve for $u=-\frac{1}{4}\cos\pi z$ on an f-plane. The non-dimensional growth rate $G=pc_i$ is plotted as a function of wavenumber p (2.3.3). The curve of widely spaced heavy dots was calculated using the shooting method with a 100 level discretization. The small closely spaced dots mark the growth rates according to a three mode spectral truncation ((38)-(40)). The straight dashed line indicates the gradient of the *growth rate curve at the SWC* ($p=\pi$) calculated analytically from (2.2.16)-(2.2.18). The continuous line is the Eady growth rate curve.

$$\Psi_1 = \cos r\pi x \sin \pi y \qquad (2.4.37)$$

and a single axisymmetric flow component $\bar{\Psi}_0$ with the same total wavenumber as Ψ_1. Investigations of the stability of barotropic waves and some free modes with higher azimuthal wavenumbers using a numerical code which implements the spectral model are presented in chapter 4. Some interesting patterns in the truncation dependence and reliability of coarse truncations described in chapter 4 are also apparent in the stability of zonal flows to which we now turn.

Linear growth rates of heavily truncated models

The simplest representation of the stability of the zonal flow

$$u = -1/2 \cos \pi z \quad ; \quad \bar{a}_0 \bar{\Psi}_0 = y/2 \cos \pi z \quad ; \quad \gamma = 0 \qquad (2.4.38)$$

uses a three mode truncation involving $\bar{\Psi}_0$ and

$$\psi_1 = \cos rk\pi x \sin \pi y \quad ; \quad \psi_2 = \sin rk\pi x \cos \pi z \sin \pi y . \qquad (2.4.39)$$

The linear growth rate of ψ_1 and ψ_2 in this model is found to be given by

$$(pc)^2 = \frac{p^2(p^2 - \pi^2)}{8(p^2 + \pi^2)} \quad ; \quad p^2 = \pi^2 \beta (r^2 k^2 + 1) . \qquad (2.4.40)$$

Figure 2.12 displays $p.c_i$ as a function of p according to this formula (small closely spaced dots) and as calculated using a numerical method with fine resolution described in section 2.6 (widely spaced heavy dots). The agreement of the two growth rate curves is remarkable. The SWCs agree exactly and the curves are almost coincident at small values of p.

The agreement at small p is readily explained. The three mode truncation only neglects the amplitudes of modes with higher vertical wavenumbers; only one aziumuthal wavenumber enters the linear calculation and since the zonal flow has no lateral shear different lateral wavenumbers are decoupled. Some amplitude will cascade down into ψ_3 = cosrkπx sinπy cos2πz and ψ_4 =

sinrkπx sinπy cos3πz etc. To assess the importance of ψ_3 we neglect the amplitude in ψ_4 and compare the growth rates of ψ_1 and ψ_2 due to the decay of the zonal flow mode with the decay of ψ_2 into $\bar{\Psi}_0$ and ψ_3. To do this note that $B\Lambda_0 = \pi^2$, $B\lambda_1 = p^2$, $B\lambda_2 = \pi^2 + p^2$, $B\lambda_3 = 4.\pi^2 + p^2$ and , writing $I \equiv \langle \bar{\Psi}_0 J(\psi_3, \psi_2) \rangle$, that $\langle \bar{\Psi}_0 J(\psi_1, \psi_2) \rangle = 2I$. For the first triad

$$\begin{aligned} \langle \psi_1 \psi_1 \rangle \, p^2 \, da_1/dt &= 2p^2 I \bar{a}_0 a_2 \\ \langle \psi_2 \psi_2 \rangle (\pi^2 + p^2) \, da_2/dt &= 2(\pi^2 - p^2) I \bar{a}_0 a_1 \end{aligned} \qquad (2.4.41)$$

whilst for the $(\bar{\Psi}_0, \psi_2, \psi_3)$ triad

$$\begin{aligned} \langle \psi_2 \psi_2 \rangle (\pi^2 + p^2) \, da_2/dt &= -(3\pi^2 + p^2) I \bar{a}_0 a_3 \\ \langle \psi_3 \psi_3 \rangle (4\pi^2 + p^2) \, da_3/dt &= p^2 I \bar{a}_0 a_2 . \end{aligned} \qquad (2.4.42)$$

From (41b) and (42a) the ratio of the decay of a_2 to its growth is about $3/2\ a_3\ /a_1$. Looking for normal mode solutions we let $da_3/dt\ /\ da_1/dt = a_3\ /\ a_1$ and find from (41a) and (42b) that $a_3/a_1 = p^2\ /\ 4\pi^2$. Hence the neglect of ψ_3 gives a relative error in a_2 of order p^2/π^2 and the three mode truncation gives growth rates with a fractional error of order $(p/\pi)^2$. The effect of ψ_4 is to drain some amplitude from ψ_3 so this estimate is likely to be an upper bound.

The omission of modes ψ_3, ψ_4 is an inaccurate approximation for $p \simeq \pi$. (2.2.19) shows that the gradient of the growth rate curve on the long wave side of the cut off is finite whereas the three mode truncation, (40), gives an unbounded gradient. Truncations at cosMπz with even values of M actually tend to misplace the SWC; with M = 2 the cut off occurs at $p_s/\pi \simeq 0.717$. This can be ascribed to the fact that these even truncations overestimate the amplitudes in the cos mπz modes with even m; the energy dammed up at the last mode in the truncation weakens the mode next to it resulting in a slight strengthening of the cos(M-2)πz mode. It is possible to show that truncations with M odd all locate the SWC correctly at $p = \pi$. But when $\gamma \neq 0$ the equivalent of the three term truncation, which uses two pairs of wave modes, also mislocates the SWC (Bell & White 1988a).

The linear stability calculation for the three mode truncation with

$$\bar{u} = -1/2 \cos\pi z \sin\pi y \quad ; \quad \bar{a}_0\bar{\psi}_0 = \frac{1}{2\pi} \cos\pi z \cos\pi y \quad ; \quad Y = 0 \qquad (2.4.43)$$

and wave modes as in (39) gives

$$(sbc)^2 = \frac{8}{9\pi^2} s^2 (sb)^2 \frac{[(sb)^2 - \pi^2]}{b^2 + \pi^2} \quad ; \quad s^2 = \frac{r^2k^2}{1 + r^2k^2} \quad . \qquad (2.4.44)$$

This formula is a useful complement to the results of more accurate estimates of the growth rate curve; see section 2.7 and Bell & White (1988a).

Section 2.5 A SUMMARY AND DISCUSSION OF SWC RESULTS

Many authors have discussed short wavelength disturbances to baroclinic zonal flows, but unfortunately a clear resulting consensus of opinion has not been firmly established. To make the facts of the matter as clear as possible the types of zonal flows which are i) known and ii) expected to possess SWCs are summarised at the start of this section. A critical discussion of the most important arguments presented by previous authors follows.

Baroclinic zonal flows store available potential energy (APE) which, as far as parcel exchange arguments are concerned, may be extracted by symmetric disturbances or wave-like disturbances of any wavelength. Interpretations of SWC arguments naturally focus on the constraints on short waves which prevent them from extracting the APE. Of course, the waves are constrained in more than one way and it need not be possible to separate the influences of the various constraints. Different arguments and constraints may be regarded as illuminating the same process from different angles and should be thought of as complementary rather than competing.

Summary of SWC results

In the interest of clarity only zonal flows with uniformly positive thermal wind shear ($\partial u/\partial z \geq 0$) will be addressed (see section 2.2 paragraph 1 for $\partial u/\partial z \leq 0$). Consider first the problem for (quasi-geostrophic) inviscid normal modes; i.e. solutions of (2.1.15) - (2.1.18). Flows with

$$\text{isotachs coincident with each } \partial q/\partial y=0 \text{ level} \qquad (2.5.1)$$

and

$$\partial q/\partial y<0 \text{ adjacent to the lower boundary,}\quad \partial q/\partial y>0 \text{ adjacent to the upper boundary} \qquad (2.5.2)$$

have SWCs. That is they have no unstable normal modes of

sufficiently large azimuthal wavenumber k; but they may nevertheless support instabilities at arbitrarily large Burger numbers (see section 2.2). Any baroclinic internal jet which can be described by a finite number of eigenfunctions of the potential vorticity operator, P, (see (2.4.5)) also has a SWC.

It is probable that all flows with boundary thermal gradients ($\partial u/\partial z>0$ at both upper and lower boundaries) and $\partial q/\partial y>0$ adjacent to the lower boundary or $\partial q/\partial y<0$ adjacent to the upper do not have SWCs. The SWC result for baroclinic internal jets and the tendency for short waves to be shallow suggest that any flow which is dominated by a finite number of eigenfunctions of P will possess a SWC if it is a baroclinic internal jet or if it satisfies (2) and its $\partial q/\partial y=0$ levels are not close to the horizontal boundaries.

Baroclinic internal jets which are the gravest modes consistent with the boundary conditions also enjoy Lyapunov stability to large amplitude disturbances.

Critical assessment of arguments by other authors

Bretherton's (1966a) discussion of short waves is important in that it focussed attention on the constraints imposed on a wave by the requirement that it make no net meridional transport of generalised potential vorticity. This constraint explains why only one of the two short wave Eady modes forms an unstable mode in Green's problem with small β (Bretherton 1966a p333). The argument led McIntyre (1970) to prove that flows close to Eady's but containing small interior potential vorticity gradients support unstable short waves if $\partial q/\partial y>0$ adjacent to the lower boundary or $\partial q/\partial y<0$ by the upper (his equation (5.5)).

Bretherton also argued (p332) from the no net flux constraint that neutral normal modes with interior critical levels must be very rare. It is true (and important) that such solutions only occur on laterally uniform flows at isolated wavenumbers, but less clear that this is true for laterally sheared flows; the condition that $\partial q/\partial y$ should not assume different signs for different values of y on the

critical level does not seem to be as restrictive as he asserts. Bretherton concluded that stability must be very rare. The absence of neutral normal modes does not, however, imply the existence of unstable normal modes or any other instabilities and Bretherton's conclusion is wrong. Unfortunately this error, which has stood uncorrected for 20 years, has had considerable influence and caused needless misinterpretation of Bretherton's (and also McIntyre's) work.

A very general approach to the stability of steady flows stems from the work of Arnol'd (1965, 1966). The basic idea is that the local conservation of potential vorticity can be used to construct a rather general class of integrals which are conserved by the motion. From the general form a specific integral(s) is constructed which is stationary w.r.t. any perturbation of the streamfunction (satisfying the boundary conditions) about the steady flow. Basic flows for which the integral is a maximum or minimum must clearly be stable in a strong sense. Blumen (1968) gives a clear account of the method for small amplitude disturbances and Holm et. al. (1985) and McIntyre & Shepherd (1987) consider large amplitude disturbances. McIntyre & Shepherd, for instance, establish that an internal jet with $0<a<T_s<A$ (for a, A positive constants) has Lyapunov stability if it is the gravest mode of the problem (see section 4.4). A very similar result is deduced using a truncated spectral model in section 2.4.

A qualitative discussion of short waves is given by Fjørtoft (1951) (see also Hide (1969) and Bell & White (1988a)). He calculates the horizontal divergence of a wave according to reasonable scaling arguments and infers the size of the vertical motions thus induced assuming the disturbance to have a depth comparable with that of the fluid. The inferred slope of the motions in the meridional plane for sufficiently short waves is found to exceed the slope of the isotherms associated with the baroclinic shear through the thermal wind relation. Deep short wave disturbances are hence unable to release the flow's APE and are stable. In combination with the constraint on the net meridional flux of generalised potential vorticity this argument provides a

straightforward interpretation of the conditions under which SWCs occur (see (1) and (2) above).

Bretherton(1966b) proposed a second qualitative description of baroclinic instability in two layer models; Hoskins et al. (1985) (p921) discuss the application of the description to continuously stratified fluids. The baroclinic wave is viewed as being comprised of two Rossby waves, one above the $\partial q/\partial y=0$ level and the other below it. The upper Rossby wave lies in a $\partial q/\partial y>0$ background and must advect itself upstream against the flow ($u - u_s > 0$). The lower wave lies in a background $\partial q/\partial y<0$ and must advect itself against the upstream flow ($u - u_s < 0$). When the two waves have an appropriate phase shift (the wave as a whole leans back with height) the lower wave can amplify the upper and vice versa. According to the Rossby wave formula,

$$c_r = u - u_s - \partial q/\partial y \,/\, (k^2 + l^2 + m^2/B), \qquad (2.5.3)$$

short wave disturbances ($k>>1$) are unable to advect themselves against the zonal flow sufficiently strongly (when $T_s=\partial q/\partial y/(u-u_s)$ is finite); the vertical shear of the zonal flow tears the two halves apart and short wave normal modes are not possible.

Section 2.6 A NUMERICAL METHOD FOR THE CALCULATION OF GROWTH RATES

This section describes the methods employed in the calculations of normal mode growth rate curves presented in the next section. All the calculations use a finite difference formulation of vertical derivatives. Trial combinations of phase velocity and streamfunction structure at the lower boundary are integrated to the upper boundary to ascertain whether they satisfy the boundary condition there. Further trials establish the dependence of the error at the upper boundary on alterations to the phase velocity or streamfunction structure and enable interpolation towards accurate solutions. Three codes have been used to calculate growth rate curves. The first applies only to laterally uniform flows. The second treats laterally sheared flows which are symmetric about mid-channel. It could reproduce all the results obtained from the first. The third code applies only to flows on an f plane which are anti-symmetric about mid-level and symmetric about mid-channel. It was developed to combat difficulties encountered in the calculations for example 1 of section 2.7. Its formulation is described at the end of the section. The method for laterally uniform flows is similar to that of Green (1960). Its extension to laterally sheared flows has not (to the best of my knowledge) been used before.

Vertical Integration

The methods use (2.1.15) re-arranged as

$$\partial/\partial z(1/B\,\partial\phi/\partial z) = K^2\phi - \partial^2\phi/\partial y^2 - q_y\phi/(u-c)\,. \qquad (2.6.1)$$

ϕ is represented only at discrete levels on a uniform grid as in fig. 2.13. The grid is arranged to straddle the boundaries and has N points in the interior spaced at intervals of length 1/N. The codes have taken $B(z) = B_0$ and represent $\partial^2\phi/\partial z^2$ by the standard second order accurate scheme

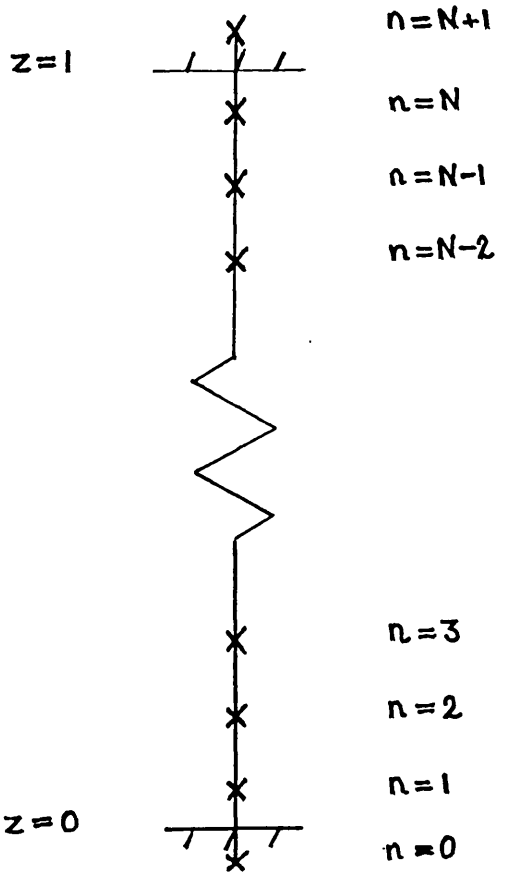


Figure 2.13

The vertical grid used in the shooting method. Grid points are marked by crosses and the boundaries at z=0 and z=1 by horizontal lines. The grid spacing is uniform ($\Delta z=1/N$) and the grid straddles the boundary.

$$\partial^2\phi/\partial z^2 \leftrightarrow N^2\{\phi_{n+1} - 2\phi_n + \phi_{n-1}\}. \qquad (2.6.2)$$

Hence

$$\phi_{n+1} = 2\phi_n - \phi_{n-1} + B/N^2\{K^2\phi_n - \partial^2\phi_n/\partial y^2 - \frac{q_y\phi_n}{u-c}\}. \qquad (2.6.3)$$

For laterally uniform flows $\phi = X(z)\sin\ell\pi y$, where ℓ is a positive integer. The horizontal wavenumbers k and ℓ only appear in (3) in the combination $k^2 + \ell^2\pi^2$. For notational convenience we set $\ell=1$ so that $B(k^2 + \ell^2\pi^2) = p^2$. From (2.1.16), $\partial q/\partial y = \gamma - 1/B\ \partial^2 u/\partial z^2$, so

$$\phi_{n+1} = 2\phi_n - \phi_{n-1} + 1/N^2\{p^2 - \frac{(B\gamma - \partial^2 u_n/\partial z^2)}{u_n - c}\}\phi_n \qquad (2.6.4)$$

and ϕ can be integrated from a pair of initial values ϕ_0 and ϕ_1 to give ϕ_N and ϕ_{N+1}.

With laterally sheared flows the lateral structure of ϕ varies with height. We represent ϕ_n by a half range sine expansion, which satisfies the side boundary conditions, truncated at wavenumber M;

$$\phi_n = \sum_{m=1}^{M} \Phi_n^m \sin m\pi y. \qquad (2.6.5)$$

Clearly the main task is to determine the coefficients of the half range expansion of $\partial q/\partial y\ \phi_n\ /\ (u - c)$ from the coefficients Φ_n^m. The method actually used assumes that u(y,z) is even about mid-channel

$$u(y-1/2, z) = u(1/2 - y, z) \qquad (2.6.6)$$

and that u(y,z) is a known function of y and z. Details of its formulation are given in appendix B. A less complicated approach would be to take a half range sine transform of Φ_n^m to find ϕ_n on a lateral grid, calculate $\partial q/\partial y\ \phi_n\ /\ (u - c)$ on the grid and transform back. This approach does not require any symmetries in u(y,z) and does not appear to have any drawbacks; the present author overlooked it when devising the code.

Upper and lower boundary conditions

The simplest horizontal b.c. is that for an internal jet;

$$\partial\phi/\partial z = 0 \qquad \text{on} \quad z = 0, 1 . \tag{2.6.7}$$

The grid is arranged so that

$$\phi_1 = \phi_0 \quad \text{for laterally uniform flows} \tag{2.6.8}$$

$$\Phi_1^m = \Phi_0^m \quad \text{for laterally sheared flows}$$

are good representations of this condition at z=0. The error in the solution at the upper boundary may reasonably represented by

$$E = \phi_{N+1} - \phi_N \qquad \text{or} \qquad E^m = \Phi_{N+1}^m - \Phi_N^m . \tag{2.6.9}$$

A trial solution is viewed as acceptable if |E| or the r.m.s. of the E^m is smaller than a given "tolerance" value.

The most general form of the boundary conditions which has been used for laterally uniform flows is

$$(u-c)\,\partial\phi/\partial z - \partial u/\partial z\,\phi + i\epsilon\beta^2 X\phi = 0 \quad \text{at } z = 0,1 \tag{2.6.10}$$

$$\epsilon = +1 \quad \text{at } z = 0 \quad ; \quad \epsilon = -1 \quad \text{at } z = 1 ,$$

where X is defined by (2.3.47). When $\partial u/\partial z \neq 0$ it has been applied at the lower boundary with

$$u(z=0) \leftrightarrow (u_0 + u_1)/2 \quad ; \quad \partial u/\partial z\,(z=0) \leftrightarrow N(u_1 - u_0)$$

$$\phi(z=0) \leftrightarrow (\phi_0 + \phi_1)/2 \quad ; \quad \partial\phi/\partial z\,(z=0) \leftrightarrow N(\phi_1 - \phi_0). \tag{2.6.11}$$

Substitution of (11) into (10) and rearrangement yields

$$\phi_1\left(u_0 - c + i\beta^2 X/(2N)\right) = \phi_0\left(u_1 - c - i\beta^2 X/(2N)\right) \tag{2.6.12}$$

A reasonable expression for the error at the upper boundary is then

$$E = \left(u_N - c - \frac{i\beta^2 X}{2N}\right)\phi_{N+1} - \left(u_{N+1} - c + \frac{i\beta^2 X}{2N}\right)\phi_N . \tag{2.6.13}$$

Inviscid solutions for laterally sheared flows with $\partial u/\partial z \neq 0$ have been obtained by applying formulae (12) and (13), with $X = 0$, at each point of a P point grid with points at $y = j/(P+1)$, $1 \leq j \leq P$. The trial vector Φ_0^m is specified in spectral form, so a half range sine transform of Φ_0^m is used to represent ϕ_0 on the P point grid. ϕ_1 calculated using (12) is transformed back into spectral component form. A similar process involving two transforms is used at the upper boundary to produce a spectral representation of E.

The values, ϕ^p, of ϕ on the P point grid just mentioned may be found, when $P \geq M$, by setting $\Phi^m = 0$ for $M < m < P$, extending Φ^m by setting

$$\Phi^{m+P} = -\Phi^{P-m} \quad \text{for} \quad 1 \leq m \leq P-1 \qquad (2.6.14)$$

and calculating by FFT

$$\phi^p = \frac{1}{\sqrt{2P}} \sum_{m=0}^{2P-1} \Phi^m \sin \frac{2p\pi m}{2P} \, . \qquad (2.6.15)$$

ϕ^p is transformed back to spectral form by by setting $\phi^{p+P} = -\phi^{p-p}$, $1 \leq p \leq P-1$ and calculating

$$\Phi^m = \frac{1}{\sqrt{2P}} \sum_{p=0}^{2P-1} \phi^p \sin \frac{2m\pi p}{2P} \, . \qquad (2.6.16)$$

Iteration Procedure

For laterally uniform flows one may set $\phi_0 = 1$. At a given wavenumber p the only undetermined variable is the complex phase velocity. Starting from a given value of c one may calculate the complex error $E = E_r + i\, E_i$ given by (9). One can also try $c = c + \Delta c_r$ to find the change $\Delta R = \Delta R_r + i\, \Delta R_i$ in the error and try $c = c + i\, \Delta c_i$ to find the error change $\Delta I = \Delta I_r + i\, \Delta I_i$. If the error function is roughly linear in arbitrary small variations δc_r and δc_i

$$\delta E \simeq \partial E/\partial c_r \, \delta c_r + \partial E/\partial c_i \, \delta c_i \, , \qquad (2.6.17)$$

so a better estimate of c is given by $c + \delta c_r + i\,\delta c_i$ with

$$-\begin{pmatrix} E_r \\ E_i \end{pmatrix} = \begin{pmatrix} \Delta R_r/\Delta c_r & \Delta I_r/\Delta c_i \\ \Delta R_i/\Delta c_r & \Delta I_i/\Delta c_i \end{pmatrix}\begin{pmatrix} \delta c_r \\ \delta c_i \end{pmatrix} . \qquad (2.6.18)$$

Clearly a good initial estimate of c helps the iteration to converge rapidly. The codes increment (or decrement) p in steps of chosen size and use the previous solution as the first guess.

For laterally sheared flows it is permissible to set $\Phi_0^s = 1$ if one is looking for solutions with some sin $s\pi y$ structure. The amplitudes of the other M-1 spectral components Φ_0^m in (5) must be determined. Let E^m, a complex vector of length M, be the error of a first guess solution, with $\Phi_0^p = F_0^p$ and $c = C$; E_p^m be the error at the upper boundary when $\phi_0 = \sin p\pi y$; and ΔE_c^m be the change in the error from the first guess when a small increment Δc is added to the trial phase speed c. Then a better estimate of the solution, if linear interpolation is a guide, is $\Phi_0^p = F_0^p + \delta\Phi^p$ with $c = C + \delta c$ where $\delta\Phi^p$ and δc are the solutions of

$$-E^m = \sum_{p=1;\, p\neq s}^{M} \Delta E_p^m\, \delta\Phi^p + \frac{\Delta E^m}{\Delta c}\,\delta c \qquad (2.6.19)$$

and $\delta\Phi_0^s = 0$. The inversion of this set of M complex linear simultaneous equations is performed by calling standard NAG packages.

Checks on coding

Two types of check were made on the code for laterally sheared flows. Firstly solutions for flows with no lateral shear were compared with results from well established code. Secondly neutral solutions for

$$u = \cos\pi z\,(a\sin\pi y + \sin 3\pi y)$$

were computed using very poor vertical resolution (N=3). The neutral solutions have no vertical dependence and the same lateral dependence as the flow itself (see section 2.2). The second check is complementary to the first and the combination is quite a strict test of the entire code.

Frailties of the method

The method is unreliable for values of c_i smaller than the typical difference between zonal flow velocities at adjacent grid levels covering a steering level ($u=c_r$) with $q_y \neq 0$. For then the simulated meridional transports of potential vorticity are sensitive to the distance between the steering level and the nearest gridpoint. The transport is overestimated, for a given c_i, when the steering level is close to a gridpoint and underestimated when it is midway between gridpoints. In Green's problem the steering level of the unstable solution moves down towards the lower boundary as p is increased. This causes spurious maxima and minima in the growth rate curve as the steering level moves past successive gridpoints on its way towards the boundary (Bell & White 1988b).

A second problem concerns the growth of modes with large lateral wavenumber. For these modes $\partial^2\phi/\partial y^2$ dominates $\partial q/\partial y\ \phi\ /\ (u - c)$ in (1). To an adequate approximation a mode dominated by sin $M\pi y$ lateral structure has

$$\phi \simeq (a\exp gz + b\exp -gz)\sin M\pi y \ ;\ g^2 = BM^2\pi^2. \quad (2.6.20)$$

For large values of B and M in (5) this creates a serious obstacle to numerical methods. Very small errors in the posited structure of ϕ_0 will excite a small error in the exp(gz) mode which will "blow up". When B = 1 and M = 9, for example, g = 9π and exp(g) ≃ 2. 10^{12}. Hence double or quadruple precision calculations become necessary for modest truncations well within the range of Burger numbers of interest. It appears that this problem will affect any numerical method which attempts to solve the stability problem (2.1.15) - (2.1.18) directly.

To combat this difficulty and produce the results of example 1 in section 2.7 a third code was written for inviscid internal jets on an f plane which are symmetric about mid-channel and anti-symmetric about mid-level;

$$u(y-\tfrac{1}{2},z) = u(\tfrac{1}{2}-y,z) \ ;\ u(y,z-\tfrac{1}{2}) = -u(y,\tfrac{1}{2}-z). \quad (2.6.21)$$

For these flows $\partial q/\partial y$ is also anti-symmetric about mid-level. If $(\phi(y,z-\frac{1}{2}),c)$ is a solution of the stability problem then so are $(\phi(y,\frac{1}{2}-z),-c)$ and $(\phi^*(y,\frac{1}{2}-z),-c^*)$. Howard's theorem shows that no more than one unstable solution breaks off from each neutral solution so the unstable solutions must be linearly dependent

$$\phi^*(y,1/2-z) = A\,\phi(y,z-1/2) \;;\; |A|=1,\; c_r=0. \qquad (2.6.22)$$

Solutions with $A = \exp(i\alpha)$ when multiplied by $\exp(-i\alpha/2)$ produce solutions with $A = 1$. So A is essentially arbitrary and we choose $A = 1$. Then the solutions must consist of linear combinations of real valued functions which are even about mid-level and pure imaginary solutions which are odd about mid-level. It is easiest to use a grid similar to that drawn in figure 2.13 but with $z = \frac{1}{2}$, rather than $z=0$, lying between levels 0 and 1. For truncation at wavenumber M, as in (5), ϕ_n is then represented by

$$\phi_n = \sum_{m=1}^{M} a^m \Phi_n^m \sin m\pi y + i\,b^m \Psi_n^m \sin m\pi y \qquad (2.6.23)$$

with

$$\Phi_1^m = \Phi_0^m \;;\quad \Psi_1^m = -\Psi_0^m. \qquad (2.6.24)$$

Expressing the complex error vector as a real vector of length 2.M, linear interpolation may be accomplished by solving a set of 2.M real valued simultaneous equations.

Section 2.7 SOME GROWTH RATE CURVES

The non dimensional growth rate most relevant to annulus experiments may be inferred from the thermal wind relation for fluids with $\rho = \rho_0(1 - \varepsilon(T - T_0))$,

$$f \Delta u / H \simeq g/\rho \, \partial\rho/\partial y = - g \varepsilon \partial T/\partial y ,$$

and the definition of the Brunt Vaisala frequency; $N^2 = -g/\rho \, \partial\rho/\partial z = g\varepsilon\partial T/\partial z$. Denoting the slope of the isotherms by θ ($\theta = -\partial T/\partial y \; / \; \partial T/\partial z$) one finds that

$$rpc_i = \frac{NH}{fL} Kc_i \simeq K_i^* c_i^*/(N\theta) . \qquad (2.7.1)$$

Many annulus experiments take the imposed temperature difference to be a constant and vary only the rotation rate. For these experiments rpc_i is a good measure of the growth rate, since $N\theta$ will suffer only small indirect variations with rotation rate, and p is proportional to the square root of the imposed thermal Rossby number.

1. Flows with lateral curvature

The main motivation for the development of the codes for the calculation of the growth rates of disturbances to laterally sheared flows was to explore the adequacy of the three term truncation result (2.4.44). This result suggests that, in geometry representative of laboratory experiments, the laterally sheared flow

$$u = -1/2 \cos \pi z \sin \pi y \; ; \; \gamma = 0 \qquad (2.7.2)$$

could support rapid growth rates at considerably greater wavenumbers (and hence thermal Rossby numbers) than a laterally uniform internal baroclinic jet.

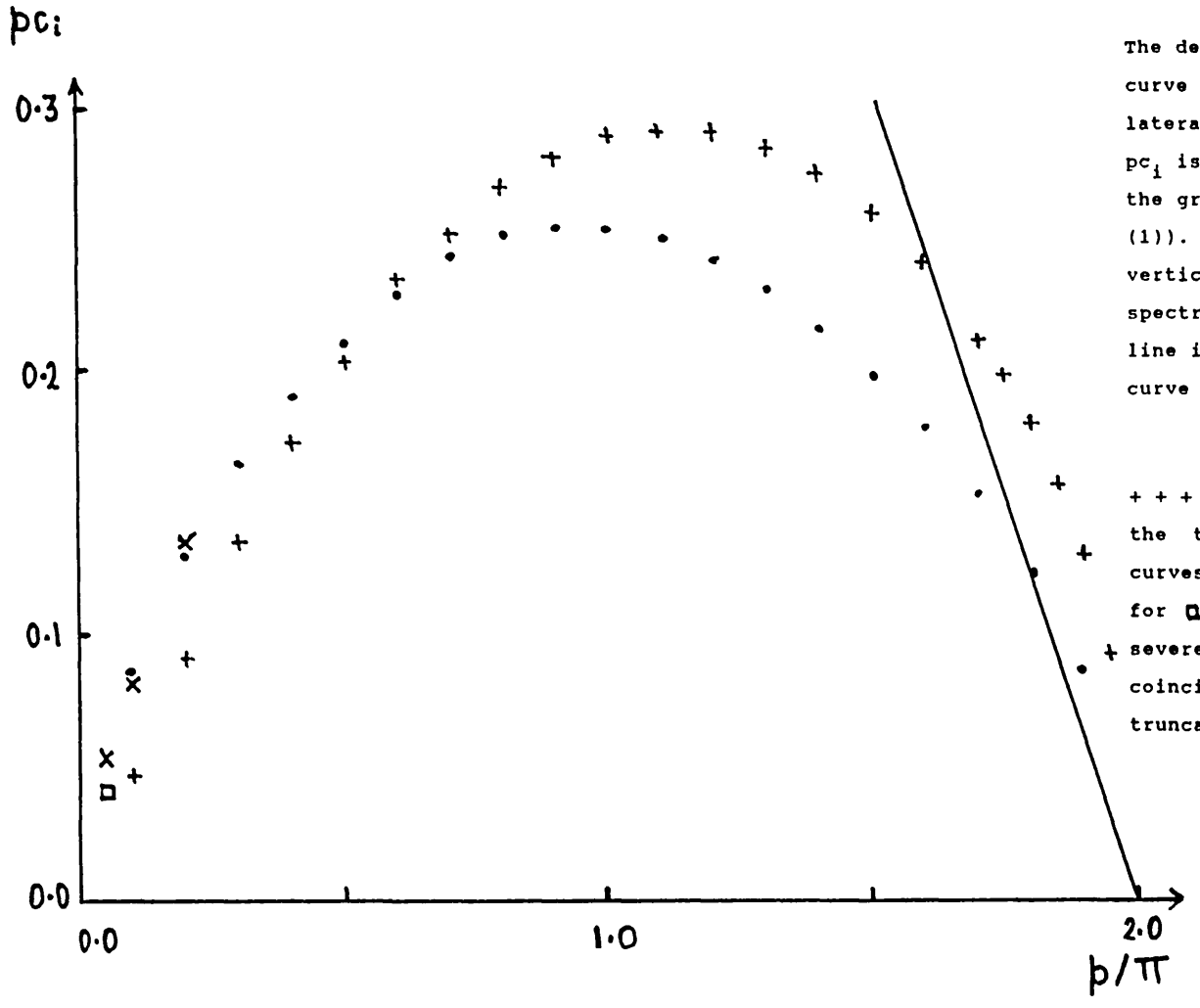


Figure 2.14

The dependence of estimates of the growth rate curve for $u=-\frac{1}{4}\cos\pi z \sin\pi y$ and $r^2=0.25$ on (a) lateral truncation and (b) vertical truncation. pc_i is an appropriate non-dimensional measure of the growth rates for annulus experiments (see (1)). N denotes the number of levels in the vertical grid (fig. 2.13) and M the lateral spectral truncation (2.6.5). The continuous line indicates the gradient of the growth rate curve at the SWC according to (2.2.16)-(2.2.18).

Figure 2.14a

+ + + indicates the growth rates (2.4.44) for the three mode truncation. N=10 for all other curves: M=3 for • • • ; M=5 for x x x ; M=7 for □ □ □ ; M=9 for Δ Δ Δ. Values for less severe truncations are not marked if they are coincident with more severe ones; all truncations listed agree for $p \geq \pi/2$.

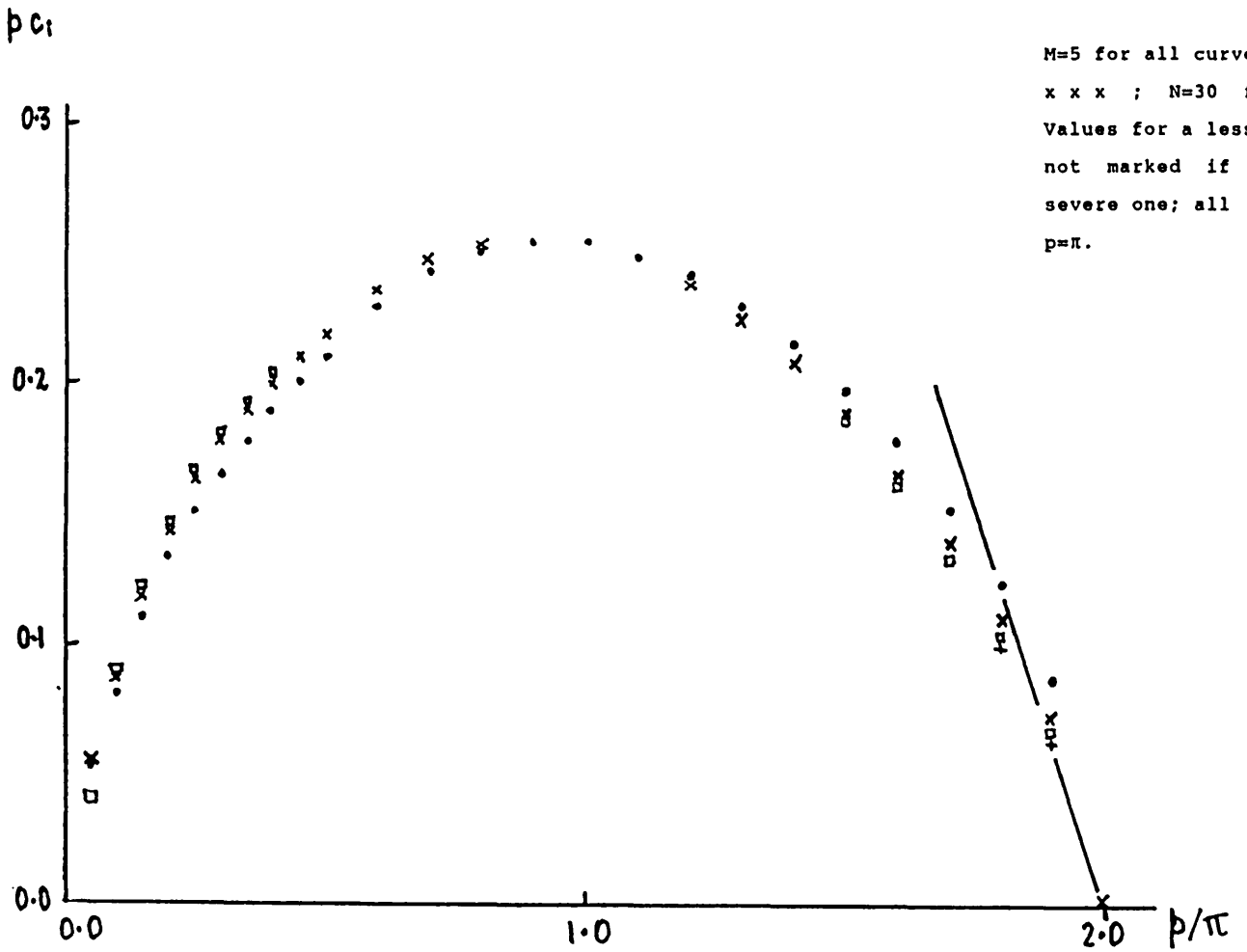


Figure 2.14b

M=5 for all curves: N=10 for • • • ; N=20 for x x x ; N=30 for □ □ □ ; N=40 for + + + . Values for a less severe truncation are again not marked if coincident with the next less severe one; all truncations listed agree for p=π.

The cut-off wavenumber p_s for (2) is given by

$$p_s = \pi/r \quad ; \quad r^2 = k^2/(k^2 + \pi^2) \tag{2.7.3}$$

(see (2.2.25) and (2.2.26)). According to (2.2.29) k = 1 is representative of annulus experiments. So ideally solutions with r $\simeq$ 1/3 and $p_s \simeq 3\pi$ would be investigated. The code BUR4, which used double precision but did not take advantage of the solution's symmetries was unable to deliver reliable solutions with more than 10 points in the vertical (N=10) and 7 lateral wavenumbers (M=7) for $p \geq 1.2\ \pi$. The code SYM4 which does use the symmetries noted in (2.6.22) was able to produce solutions up to $p \simeq 2\pi$ for resolutions with M=9 and N=10 or M=5 and N=40; N being the implied number of points between z=0 and z=1.

The dependence of solutions with $r^2 = 0.25$ on lateral resolution (M) is presented in fig. 2.14a and the dependence on vertical resolution (N) in fig. 2.14b. The three term truncation result (2.4.44) is also plotted on fig. 2.14a for comparison. For $p \geq \pi$ the growth rates are insensitive to the lateral truncation and the amplitudes of the streamfunction components decrease rapidly with lateral wavenumber. The growth rates are much more sensitive to the vertical resolution. The gradient of the growth rate curve at the SWC is plotted on both figs. 2.14; it is clear that inadequate vertical resolution (with an even number of gridpoints) results in an over-estimation of the growth rates near the cut-off.

Near $p=\pi$ the growth rate reaches a maximum which depends little on the variations in truncation covered by figs. 2.14. For smaller values of p, in particular $p \simeq \pi/5$, the growth rates are sensitive to both vertical and lateral truncation and the lateral structure is rich.

The results of fig. 2.14a suggest that higher lateral modes may play little part in solutions with high wavenumber p and that severe lateral truncations for $p^2 >> 1$ may be accurate. Following this hypothesis fig. 2.15 displays growth rates for flow (2) with $r^2 = 0.1$ obtained using SYM4 with poor lateral resolution and

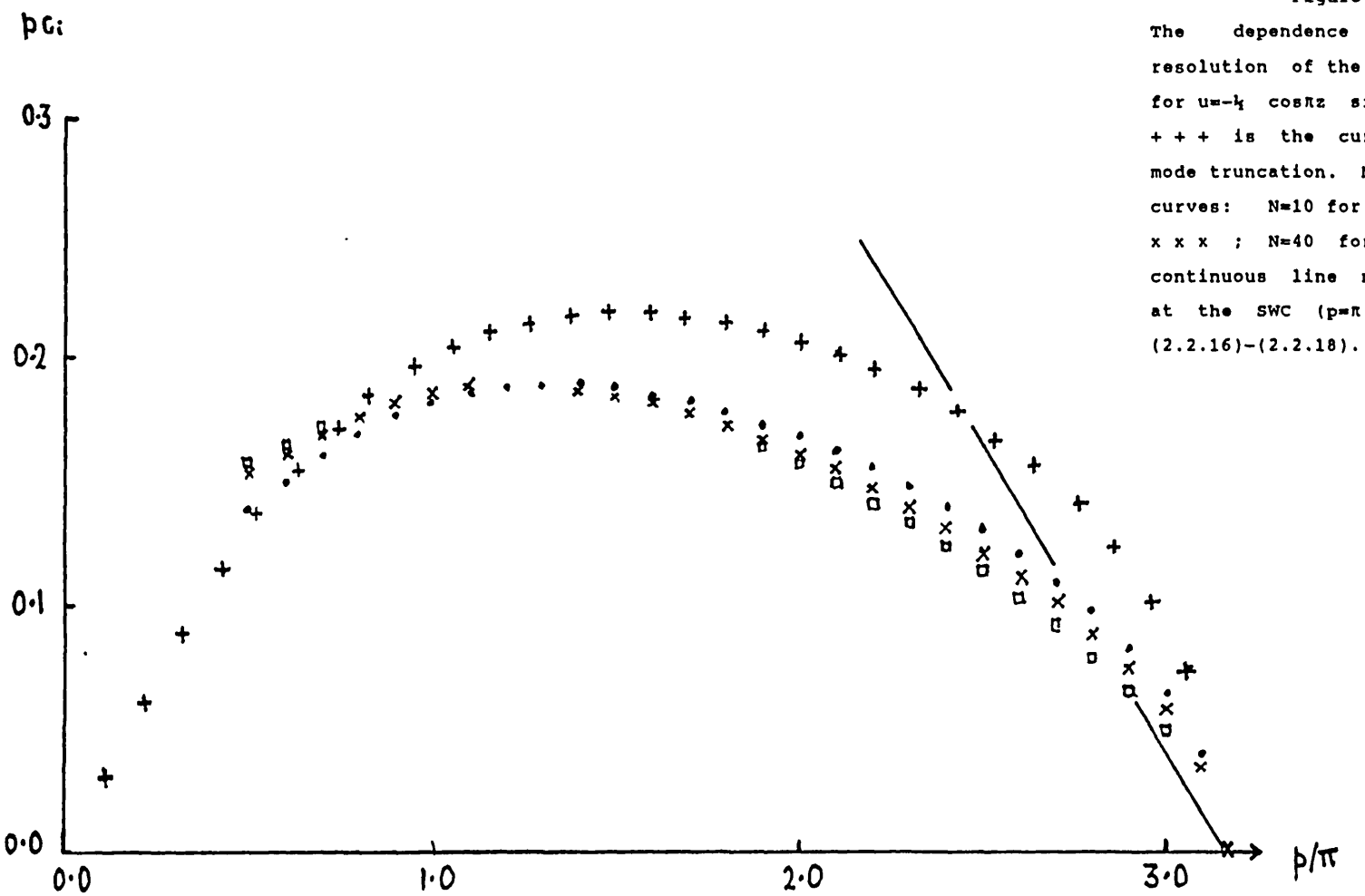


Figure 2.15

The dependence on vertical resolution of the growth rate curve for $u=-\frac{1}{4}\cos\pi z \sin\pi y$ and $r^2=0.1$. + + + is the curve for the three mode truncation. M=3 for all other curves: N=10 for • • • ; N=20 for x x x ; N=40 for □ □ □ . The continuous line marks the gradient at the SWC ($p=\pi$) according to (2.2.16)-(2.2.18).

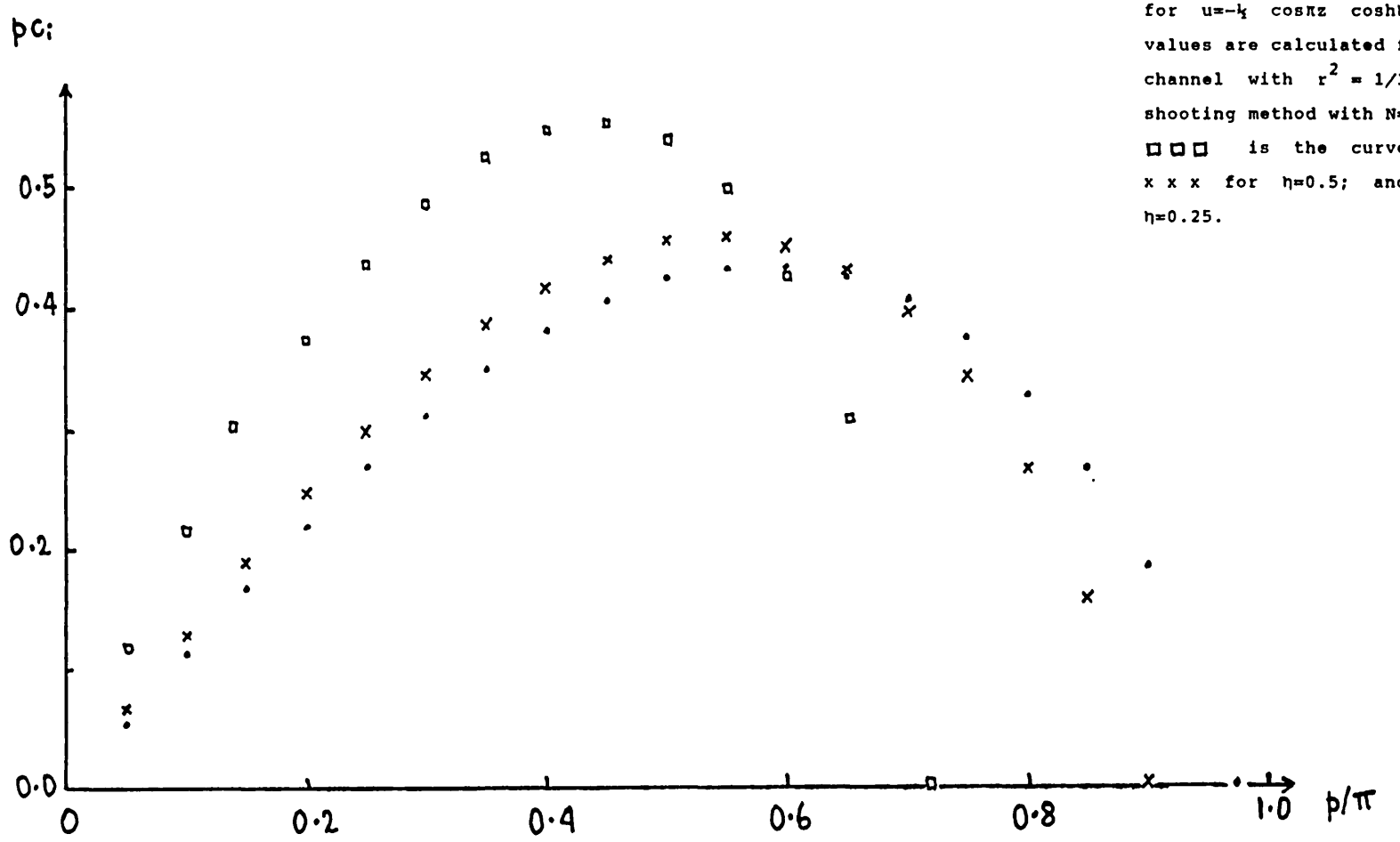


Figure 2.16

Estimates of the growth rate curves for $u=-\frac{1}{4}\cos\pi z\cosh\eta\pi(y-\frac{1}{4})$. All values are calculated for an f-plane channel with $r^2 = 1/17$, using the shooting method with N=40 and M=5; □ □ □ is the curve for $\eta=1.0$; x x x for $\eta=0.5$; and • • • for $\eta=0.25$.

relatively high vertical resolution. The three term truncation result is again displayed for intercomparison. The features noted for fig. 2.14b apply also to fig. 2.15.

The codes BUR4 and SYM4 are particularly effective at small values of p where they can be used at relatively high resolution without difficulty. Some results for the stability of the hyperbolically sheared flow (2.2.25) on an f-plane with $r^2 = 1/17$ are displayed in fig. 2.16; the calculations used SYM4 with N=40 and M=5. The movement of the SWC to lower wavenumbers, p, as the ratio R of the flow at the sides to the middle is increased is clearly evident as is the similar movement of the growth rate maximum.

The codes which have been developed could be used to investigate flows with boundary thermal gradients on beta planes, provided the flow were symmetric about mid-channel. A simplification of the code outlined after (2.6.6) would enable flows without the mid-channel symmetry to be studied. The stabilising effect of a barotropic lateral shear anti-symmetric about mid-channel (James 1987) is particularly worthy of investigation.

2. Decaying branches in the presence of Ekman pumping

Not all the decaying normal modes on the flow

$$u = -1/2 \cos \pi z \quad ; \quad \gamma = 0$$

are connected to the non-singular normal modes which develop from the neutral normal modes of the inviscid problem as the Ekman number is increased. Numerical investigation of one set of normal modes showed them to decay more rapidly as the wavenumber p increased. The arguments presented in section 2.3, concerning the continuity of solution branches with both p and X, raised questions about the neutral modes from which these solutions originate. The structure of the rapidly decaying modes at large wavenumbers also called for elucidation.

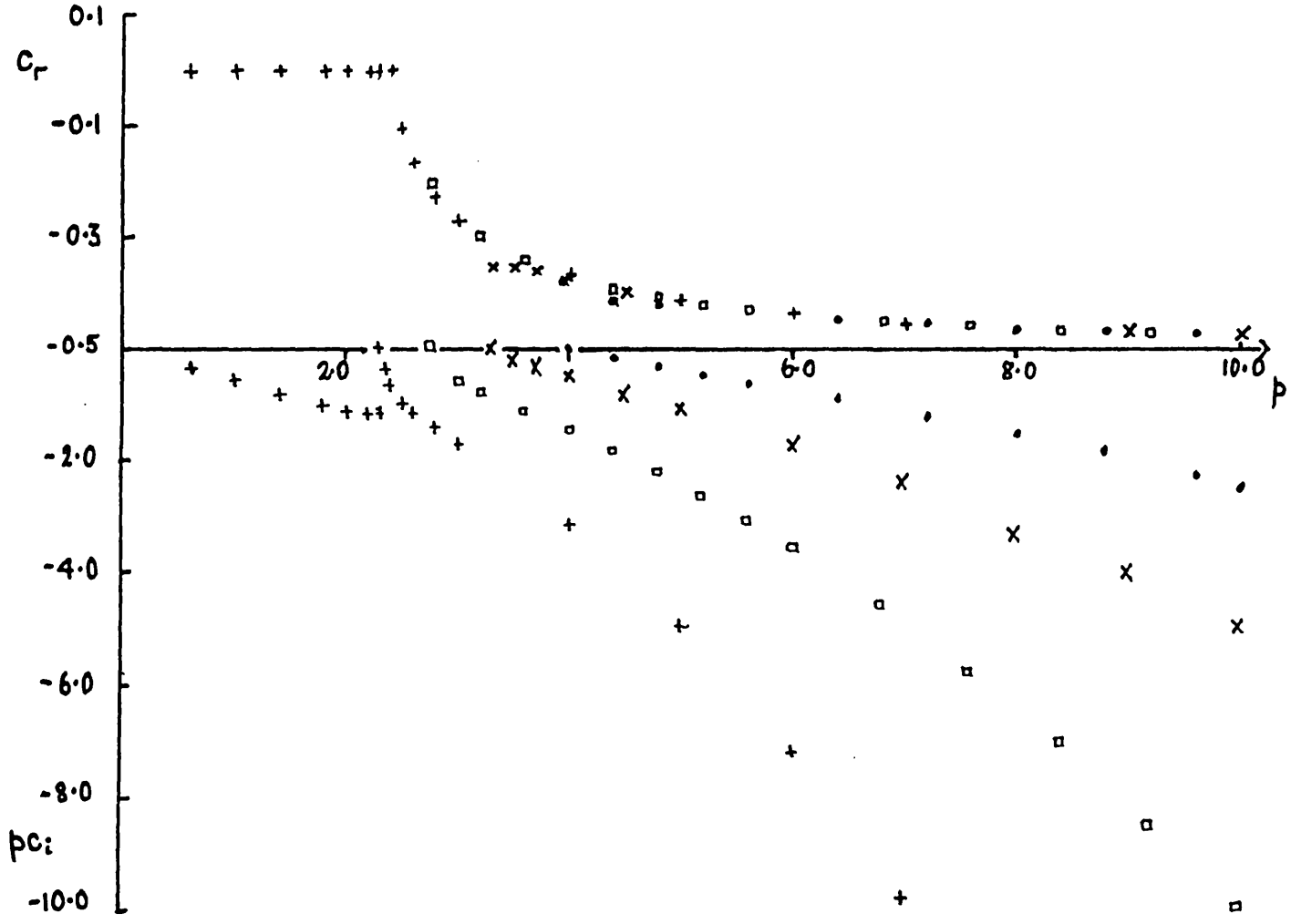


Figure 2.17

The phase speeds and decay rates of some modes subject to Ekman pumping. The phase speeds ($-0.5<c_r\leq 0$) and non-dimensional decay rates ($pc_i<0$) presented are for modes on the f-plane flow $u=-\frac{1}{2}\cos\pi z$. $X\equiv E^{\frac{1}{4}}/(kR_o)$ is the Ekman parameter; see (2.3.47). X=0.025 for the curves • • • ; X=0.05 for x x x ; X=0.1 for □ □ □ ; and X=0.2 for + + + . All values were calculated using the shooting method with N=100.

Starting from a decaying solution with X=0.08 and p=3.3 connected solutions were found by small variations of X or p. Fig. 2.17 shows some of the connected solutions obtained at four values of X, namely 0.025, 0.05, 0.1 and 0.2, using a 100 level resolution. The figure displays both the decay rate pc_i and the real phase speed c_r of these solutions as functions of p. For all solutions the decay rate increases as p increases and the phase velocity tends to the zonal flow velocity at the (lower) boundary. In these respects the solutions resemble those found by Barcilon (1964) and illustrated in fig. 14 of Hide & Mason (1975). The structure of the solutions clearly depends crucially on the Ekman pumping at the lower boundary and its contribution to the generalised potential vorticity flux since the solutions are not present in the absence of Ekman pumping.

The solution branches terminate in singular neutral modes. The neutral solution with $c_i \uparrow 0$ for X=0.1 has been investigated using NEUSING (see section 2.3). With 50 level resolution the solution has p=2.71 and c_r=-0.22; with 100 levels p=2.747 and c_r=-0.203; and with 140 levels p=2.755 and c_r=-0.201. No solutions were found for X=0.05, but for X=0.049 solutions with 100 level resolution (p=3.32, c_r=-0.355) and 50 levels (p=3.26, c_r=-0.338) were detected. Tracing branches of neutral solutions by varying X proved to be unprofitable. It may nevertheless be confidently asserted, on the basis of the continuity arguments, that the neutral solutions form a single branch in the (X, p) plane with $p \uparrow \infty$ as $X \downarrow 0$.

Some light was shed on the difficulties in finding singular solutions by searching for solutions with fixed values of N_1 and N_2 (the number of levels in the grid below and above the critical level) rather than calculating N_1 and N_2 as a function of c_r. Convergence was greatly improved but the solutions found to be very sensitive to N_1/N_2. Future work would need to make the grid separation exactly equal on each side of the critical level so that the treatment of the logarithmic singularity in $\partial\psi/\partial z$ is identical on each side.

Fig. 2.17 shows that the X=0.2 decaying solution branch has a bifurcation point at $p \simeq p_B \simeq 2.5$, $c_i \simeq -0.35$. The branch of

solutions ($\phi(z)$, c) with $p > p_B$ and $c_r < 0$ has a mirror branch ($\phi^*(-z)$, $-c^*$). These two branches meet at p_B. Below p_B there are two sets of solutions with $c_r=0$ which decay at different rates. The structure of the neutral solution branches in the range $0.1< X < 0.2$ has not been fully explored; the branch of singular neutral solutions with $X < 0.1$ might simply intersect the branch of non-singular solutions (with $c_r=0$) between X=0.1 and X=0.2, but another branch of singular neutral solutions, which appears to extend from X=0.1 (p =2.747, $c_r=-0.2035$ for N=100) to X=0.2 (p=1.995, $c_r=-0.3105$ for N=100), might complicate matters.

3. Green's problem

Green (1960) examined the stability of a flow with uniform baroclinic shear on a beta plane. He found that the flow had no SWC (see section 2.5). The unstable short wavelength normal modes he found are trapped near the lower boundary. The arguments of sections 2.3 and 2.5 suggest that such short wave boundary instabilities should not occur in annulus experiments because of the absence of vertical shear at the boundaries and/or the particular distribution of the potential vorticity of the zonal flow. Jonas (1980), however, has presented some convincing experimental evidence that at least some of the weak waves obtained at high Taylor numbers are trapped near the lower boundary. It is not clear how to reconcile these theoretical and experimental results but it may be that the distributions of the shears of the zonal flow at high Taylor numbers just above the UAT are significantly different from those presented in fig. 2.1.

This discrepancy breathes new life into the question of whether short wave instabilities could be virulent enough to prevent a sharp upper axisymmetric transition on zonal flows with shears markedly different from those of fig. 2.1. Suppose, for the sake of argument, that the zonal flow was of the form

$$u = \sin \pi/2 (z-1/2) \quad \cosh \eta (y-1/2). \qquad (2.7.4)$$

At large values of B, the lateral shear of the flow would dominate

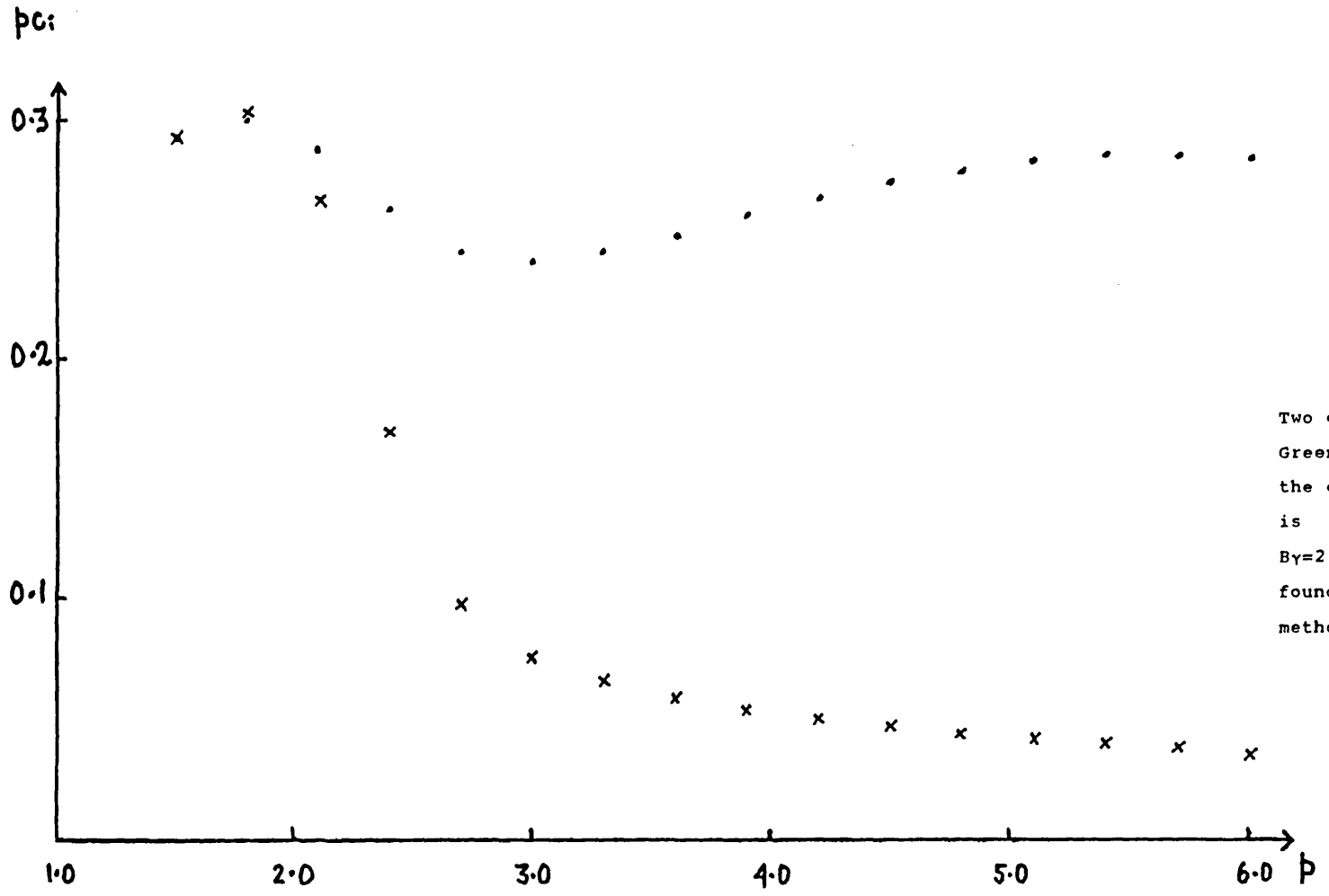


Figure 2.18

Two growth rate curves for Green's problem. x x x is the curve when $B\gamma=\frac{1}{4}$; . . . is the curve when $B\gamma=2(p/3)^2$. All values were found using the shooting method with N=100.

contributions to $\partial q/\partial y$ in (2.1.15). We can model the trapped unstable short waves of such flows near one of the boundaries by taking $u=z$ and using an effective beta parameter representative of the flow near that boundary; $\gamma = \beta L^2 / \Delta u = \eta^2$ for η defined by (4). The stability problem is then

$$d^2\phi/dz^2 - p^2\phi = -B\gamma\phi/(u-c) \quad 0 \leq z \leq 1$$
$$(u-c)\,d\phi/dz - du/dz\,\phi = 0 \quad \text{at} \quad z=0,1. \tag{2.7.5}$$

In most annulus experiments $p^2 = B(k^2 + \pi^2) \simeq B\pi^2$ so $B\gamma \simeq (p/\pi)^2\gamma$.

Most investigations of Green's problem have taken $B\gamma$ to be a constant. The growth rate curve marked by crosses on fig. 2.17 is an example with $B\gamma = 1/2$. For this curve it is plausible to suppose that Ekman pumping would stabilize the weakly unstable short wave tail and leave a sharp transition between stability and instability. The curve marked by small circles on fig. 2.18 has $B\gamma = 2\,(p/3)^2$. The short waves are clearly very unstable in this case. Ekman pumping would probably stabilize the very short wave disturbances but such a cut-off would be rather sensitive to the Taylor number.

4. Flows trapped near a boundary: Polar lows

Short wavelength boundary instabilities require strong zonal flow shears at a horizontal boundary and strong (interior) potential vorticity gradients next to it. Flows which decay exponentially with height

$$u = \Delta u \exp(-sz) \tag{2.7.6}$$

satisfy both requirements. The growth rate curves for s=2 (crosses), s=4 (circles) and s=6 (triangles) are displayed on fig. 2.19 (the calculations used a 100 level resolution). For each curve Δu was chosen so that the APE of all the flows was the same by setting

$$\int_0^1 (du/dz)^2\,dz = 1\,. \tag{2.7.7}$$

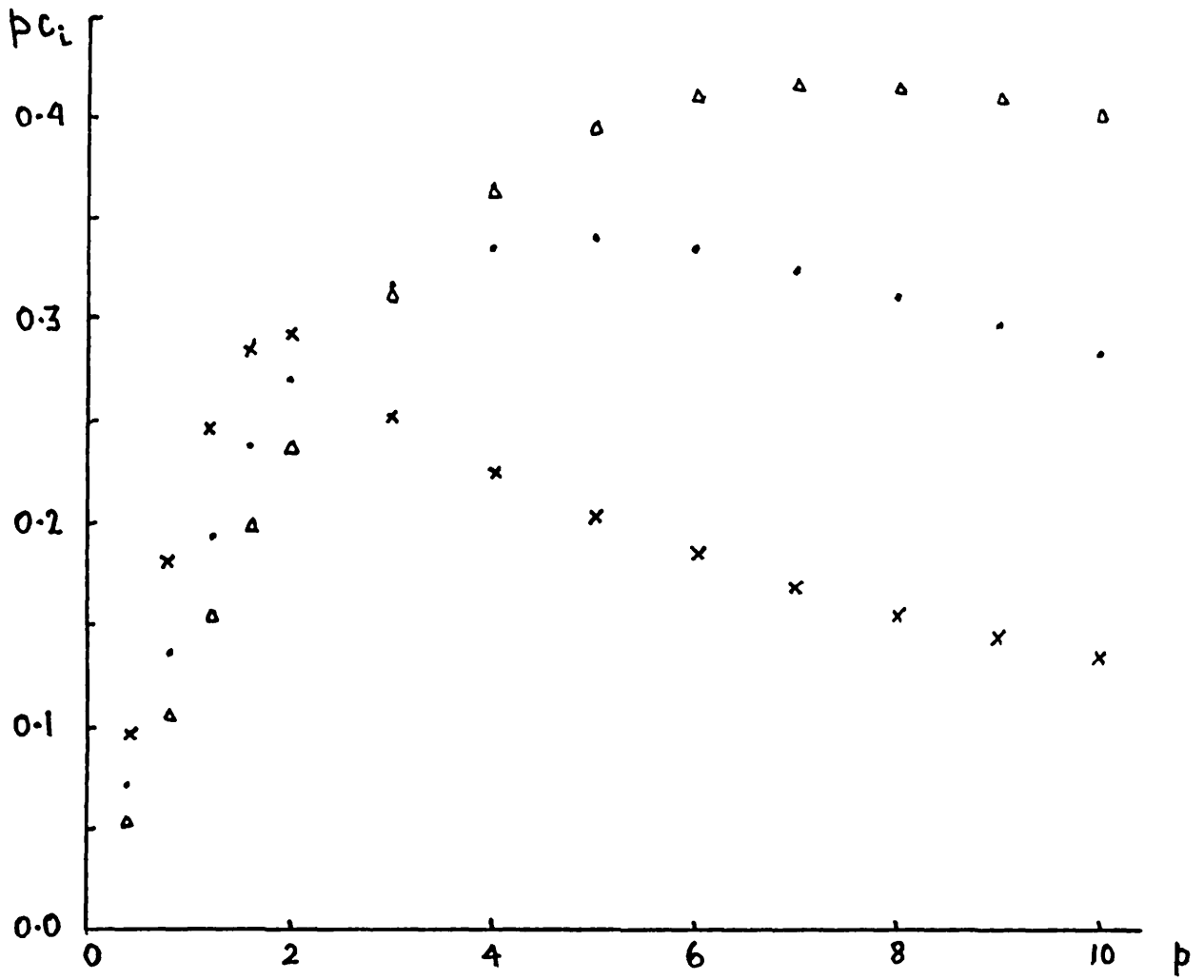


Figure 2.19

Growth rate curves for flows $u=\Delta u \exp -sz$ on an f-plane. Δu is chosen so that all flows have the same total APE; see (7). s=2 for x x x ; s=4 for . . . ; s=6 for Δ Δ Δ. Values calculated by the shooting method with N=100.

The wavenumber of maximum instability clearly increases as the flow becomes more trapped, as does the growth rate maximum.

Polar lows are small shallow disturbances which develop in shallow but intense baroclinic regions created by polar airstreams flowing southward over a relatively warm sea (Reed (1987) gives a review). Harrold & Browning (1969) and Mansfield (1974) give accounts of these disturbances by modelling the low level inversion which often caps the baroclinic layer as a rigid lid and applying Eady's theory. Growth rates in the Eady model depend only on $p = N H K / f$, K being the disturbance's total horizontal wavenumber. The small height H of the capping inversion compared with the tropopause and/or low values of the static stability account for the short wavelength of the polar lows.

Polar lows are thought to be strongly modified by various processes such as latent heat release. An ingredient which has been neglected but the above results suggest may be important is the potential vorticity gradients associated with the trapped vertical shear of the zonal flow. These gradients may be coarsely modelled by a shallow layer of baroclinic shear overlaid by a deep barotropic layer (cf the semi-internal jet discussed by Bell & White (1988a)). Such a coarse model might well be quite successful in modelling rapidly growing disturbances.

CHAPTER THREE

THE EVOLUTION OF A SMALL AMPLITUDE WAVE ON A LATERALLY SHEARED BAROCLINIC JET

Section 3.1 INTRODUCTION

The stability of baroclinic zonal flows against wave perturbations of infinitesimal amplitude and the thermal Rossby number of the transition between stability and instability have been discussed in some detail in the previous chapter. Those investigations attempt to account for the stability of the axisymmetric state at high thermal Rossby number (Θ) and the location of the upper axisymmetric transition found in laboratory experiments.

The interaction and equilibration of waves in the wave-regime is a more difficult and less well developed theoretical subject. One profitable approach to it, pioneered for plane parallel shear flows by Stuart (1960) and Watson (1960), is to restrict attention to small amplitude waves close to the stability transition in which the advection by the axisymmetric flow remains dominant but the non-linear advection is sufficient to counteract the weak linear growth. The aim of these analyses is to provide insight into the mechanisms which govern the wave evolution and equilibration. The hope is that the mechanisms uncovered will be important for waves of small to moderate amplitude.

The growth, saturation and decay (i.e. life cycle) of synoptic scale baroclinic eddies in the atmosphere is clearly a crucial process for the forecasting and understanding of the weather. The atmosphere itself is a hostile environment in which to test ideas of wave evolution since it is subject to complicated boundary conditions and forcing processes. Numerical models which can be

operated with chosen boundary conditions, selected physical processes and under controlled conditions are more promising test beds (Simmons & Hoskins 1980). The most thoroughly explored testing grounds, however, are the laboratory experiments with rotating annuli of stably stratified fluid.

The experiments driven by thermal forcing in the annulus depicted in fig. 1.1 (a=2.5cm, b=8.0cm, H=14.0cm) display transitions, at moderate and high Taylor numbers, from an axisymmetric flow at high Θ, through amplitude vacillating flows, to steady waves at low Θ. At lower Taylor numbers the transition is directly from an axisymmetric flow to a steady wave. Many details of the structure of amplitude vacillations have been investigated. Hignett (1985) reports that the waves grow more rapidly than they decay and that both the period of the vacillation and the ratio of the maximum to the minimum amplitude increase with the rotation period. More recently, Johnson (private communication) has found that at the small amplitude (growth) stage the mean zonal flow has its maximum at mid-channel whilst in the decay phase its minimum is at mid-channel. Finally Jonas (1981) found the areas of the regime diagram occupied by amplitude vacillation to be sensitive to the fluid's Prandtl number.

According to the results of example 1 in section 2.2 the zonal flow during the decay phase is more stable to linear disturbances than that during the growth phase. So the amplitude vacillation could be interpreted as involving a linear growth stage by the end of which the wave-mean flow interaction has reshaped the zonal flow enough to make it sub-critical. Wave-mean flow interactions are the principal non-linear interactions in Drazin's (1970) inviscid WNL analysis (and also in Pedlosky's (1970) two layer analysis), but Drazin's analysis neglects interior potential vorticity gradients, so detailed comparisons would be inappropriate.

Experiments with two-layer systems driven by a rotating lid give a <u>different</u> sequence of transitions from those found in the thermal annuli (White 1988). As the Burger number (inverse Froude number) decreases the two layer system passes from axisymmetric flow through

steady waves to amplitude modulated waves (Hart 1979 fig 5). Clearly it is important to account for this difference. One possibility is that the amplitude vacillation found in the thermal experiments is related to the relaxation response of the axisymmetric flow to thermal forcing.

We explore an alternative possibility here by considering the transports which appear at lowest order in a weakly non-linear expansion of the quasi - geostrophic equations about a marginally stable mode. Consider first the continuous equations relevant to the thermal annuli. The neutral mode ($\psi_0 \equiv \phi_0 . \cos kx$ with phase speed c_0 and potential vorticity q_0) satisfies

$$(\bar{u} - c_0) q_0 + \bar{q}_y \psi_0 = 0 \tag{3.1.1}$$

$$(\bar{u} - c_0) \partial\phi_0/\partial z - \partial\bar{u}/\partial z \, \phi_0 = -i \epsilon p^2 X \phi_0 \tag{3.1.2}$$

$$X = E^{1/2}/(kR_0) \quad ; \quad \epsilon = 1 \text{ on } z=0, \quad \epsilon = -1 \text{ on } z=1 .$$

The lowest order non-linear interactions possible are provided by the advection of the neutral mode's potential vorticity field by its own streamfunction in the interior and the advection of the neutral mode's thermal field by its own streamfunction at the boundaries. Using (1) the first of these may be expressed as

$$J(\psi_0, q_0) = \psi_0 \partial\psi_0/\partial x \, \partial/\partial y \{ -q_y/(\bar{u} - c_0) \} . \tag{3.1.3}$$

This self-advection is zero for laterally uniform flows and non-zero for most laterally sheared flows. The advection of the thermal field at the boundary may be expressed, after using (2), as

$$J(\psi_0, \partial\psi_0/\partial z) = \psi_0 \partial\psi_0/\partial x \, \partial/\partial y \{ \partial\bar{u}/\partial z/(\bar{u} - c_0) \}$$
$$- \frac{\epsilon p^2 X}{k} \left\{ \left| \frac{\partial\psi_0}{\partial x} \right|^2 \frac{\partial}{\partial y} \left(\frac{1}{\bar{u} - c_0} \right) + \frac{k^2 \phi_0 \partial\phi_0/\partial y}{\bar{u} - c_0} \right\} . \tag{3.1.4}$$

If the flow is inviscid the thermal advection is zero both for laterally uniform flows and internal baroclinic jets. In the presence of Ekman pumping the thermal advection can induce both a zonal mean and a first (azimuthal) harmonic thermal correction.

The above analysis shows that inviscid neutral waves on most laterally sheared flows will generate their own harmonics and that the lateral curvature of a zonal flow should not be neglected in weakly non-linear analyses; WNL analyses for purely baroclinic inviscid flows will apply only to flows for which $q_y/(u-c)$ is very nearly independent of y. This point might give some insight into the conclusion of Simmons & Hoskins (1980) that alterations to the lateral shear of a zonal flow can significantly alter the maximum amplitude achieved by normal mode disturbances of similar linear growth rates.

The neutral mode marking the axisymmetric transition, being a gravest mode, has relatively little vertical variation (see sections 2.2 - 2.4). For example, the Jacobian $J(\psi_0, q_0)$ for $u = f(y).\cos \pi z$ is independent of height. Near the wave's steering level (i.e. where $u=c_0$) the advection of q_0 by the zonal flow does not dominate that of the wave itself. In this critical layer all azimuthal harmonics of the wave will be excited at the order of the square of the amplitude of the neutral mode.

The development of a small amplitude inviscid wave on the zonal flow

$$\bar{u} = (1 - a_s + a_s \sin \pi y) \sin \pi z \; ; \; -1/2 \leq z \leq 1/2, \; 0 \leq y \leq 1 \qquad (3.1.5)$$

is analysed in section 3.2. The mathematical problem which results is similar to that of a Rossby wave impinging on a critical layer, as studied by Warn & Warn (1978), Stewartson (1978) and Killworth & McIntyre (1985). Maslowe (1986) presents a useful review of non-linear critical layers. Numerical solutions for the present problem are reported in section 3.3 and various deficiencies of the solution indicated and discussed in section 3.4.

The effect of lateral shear on inviscid WNL evolution in two-layer analyses is easier to describe. The advection of the normal mode's potential vorticity by itself again induces a harmonic of the neutral mode. The zonal flow advection dominates the wave advection in both layers (they are not critical layers usually,

though see Pedlosky (1982b)) so complications in a critical layer do not arise. The neutral mode's advection of the harmonic results in a modification of the streamfunction in the principal wavenumber. The amplitude equation is hence of Landau's type,

$$\partial A/\partial t = c\Delta B^{1/2} A - \mu A^3. \qquad (3.1.6)$$

Here A denotes the wave amplitude, t time, $\Delta B = B_0 - B$ the supercritical value of the Burger number and c and μ constant real valued coefficients[1]. The solutions evolve monotonically towards steady waves of amplitude $(c\Delta B^{1/2}/\mu)^{1/2}$, provided $c/\mu > 0$.

The lateral structure of the experimental two-layer flows has been investigated by Hart (1972). He argues that each layer's depth independent axisymmetric zonal flow is arrested by a Stewartson type $E^{\frac{1}{4}}$ layer (E being the Ekman number as defined in (2.3.47)) whilst the zonal motions of wave perturbations are halted by a viscous Stokes layer. The latter is thinner than the Stewartson layer in most experiments so that much of the axisymmetric flow's lateral shear in its side boundary layer should probably be viewed as being within the inviscid domain of the perturbation field. The two-layer zonal flow's lateral shear is neglected principally for analytical convenience. WNL theories making this assumption do so without due regard for its effect on the dynamics.

In summary, a laterally sheared two layer flow, with $\partial\{q_y/(u-c_0)\}/\partial y \neq 0$ may be expected to give a transition (at high Taylor number) from axisymmetric flow to steady waves governed by Landau's equation. The evolution of inviscid waves on a continuous zonal flow with $\partial\{q_y/(u-c_0)\}/\partial y \neq 0$ involves a non-linear critical layer and is discussed in the next three sections.

[1] The linear growth rate is assumed to be proportional to $\Delta B^{1/2}$ (rather than ΔB).

Section 3.2 AN ANALYSIS OF A NON-LINEAR BAROCLINIC CRITICAL LAYER

Problem Definition

The quasi-geostrophic equations for an f plane

$$\partial/\partial t\,(Bq) + J(\psi, Bq) = 0 \qquad -1/2 \le z \le 1/2,\ 0 \le y \le 1 \tag{3.2.1}$$

$$Bq = B\,(f_o + \nabla_H^2 \psi) + \partial^2\psi/\partial z^2 \tag{3.2.2}$$

$$\partial^2\psi/\partial z\partial t + J(\psi, \partial\psi/\partial z) = 0 \qquad \text{on } z = \pm 1/2 \tag{3.2.3}$$

$$\partial\psi/\partial x = 0 \quad \text{and} \quad \partial^2\bar{\psi}/\partial y\partial t = 0 \qquad \text{on } y = 0, 1 \tag{3.2.4}$$

will be assumed to govern the motion: (1) represents conservation of the quasi-geostrophic potential vorticity (2) following the geostrophic motion; (3) represents conservation of potential temperature at the rigid (no normal flow) horizontal boundaries and (4) no normal flow and energy conserving conditions at the side boundaries. The stratification is assumed to be independent of height (see (2)) and the channel is taken to be periodic in x with repeat length L_x/L_y. The zonal flow is assumed to be be of the form

$$\bar{u} = 1/2 \sin \pi z \,(1 - a_s + a_s \sin \pi y). \tag{3.2.5}$$

This internal baroclinic jet has a non-singular critical level at z=0 and $\partial/\partial y\{\bar{q}_y/\bar{u}\} \neq 0$ (see (3.1.3)). The flow is studied at a Burger number slightly smaller than that of the SWC. The evolution of the only unstable weakly growing normal mode is traced as it grows from a very small initial amplitude.

The form of the neutral normal mode and the appropriate scalings are summarised first. The outer solutions (i.e. those outside the critical layer) are then developed and the jump conditions across the critical layer established. Presentations of the inner solution and the matching of the inner solution with the jump in the outer

thermal field across the critical layer follow. Formulae for the unstable normal mode which serves as the initial condition conclude the section. A summary of the key formulae is given at the start of the next section.

Neutral mode solution

T_s is well defined for (5), and independent of height;

$$BT_s = B\bar{q}_y/\bar{u} = \pi^2 + \frac{B\pi^2 a_s \sin\pi y}{1 - a_s + a_s \sin\pi y} . \qquad (3.2.6)$$

The results of section 2.2 show that the SWC for (5) is marked by a non-singular neutral mode

$$\psi_0 = \rho_0(y) \cos r\pi x \quad ; \quad r = 2L_y/L_x \qquad (3.2.7)$$

which is independent of height. $\rho_0(y)$ in (7) is the gravest mode eigenfunction of

$$d^2\rho/dy^2 + \left\{ \frac{\pi^2 a_s \sin\pi y}{1 - a_s + a_s \sin\pi y} \right\} \rho + \lambda\rho = 0 \qquad 0 \le y \le 1 \qquad (3.2.8)$$

$$\rho = 0 \qquad \text{on} \quad y = 0, 1$$

and B_0, the Burger number at the SWC, is related to the gravest eigenvalue, λ_0, by

$$B_0 = \frac{\pi^2}{\lambda_0 + r^2\pi^2} . \qquad (3.2.9)$$

Scaling considerations

The Burger number is assumed to be just less than its SWC value

$$B = B_0 - \Delta B . \qquad (3.2.10)$$

The normal mode growth rate is then of order ΔB (see section 2.2 or (42) & (46) below) so we introduce an appropriate slow time τ

$$\tau = \Delta\beta\, t. \qquad (3.2.11)$$

The generation of harmonics and non-linear corrections to the dominant wave mode are argued in section 3.1 to be of the order of the square of the wave amplitude, A, (when $\partial/\partial y\{\partial q/\partial y/u\} \neq 0$) so the non-linear terms can arrest the linear growth if

$$A = \Delta B\, a(\tau) \qquad (3.2.12)$$

- $a(\tau)$ being of order 1 when linear growth and non-linear advection become comparable. Within the critical layer the zonal velocity $\bar{u}$ of (5) is of the order $\pi z/2$ providing a_s (see (5)) is of $O(\Delta B^0)$ rather than $O(\Delta B^1)$. The critical layer hence has a depth of $O(\Delta B)$; for $z >> \Delta B$ zonal flow advection dominates. So the vertical co-ordinate ζ,

$$z = \Delta B\, \zeta \qquad (3.2.13)$$

will be used within the critical layer. The inner solution has domain $-\infty<\zeta<+\infty$. The inner solution for $\zeta \to \pm\infty$ is matched to the outer solutions with $z \to 0^{\pm}$.

Outer solutions

Two outer solutions are required; one above and one below the critical layer. The equations governing the two are very similar and for the most part they will not be painstakingly distinguished. We pose the perturbation expansion for the streamfunction suggested by (12)

$$\psi = \bar{\psi}_0 + \Delta B\, \psi_1 + \Delta B^2 \psi_2 + \ldots. \qquad (3.2.14)$$

in which $\bar{\psi}_0$ is the streamfunction of the zonal flow (5). Substituting into (1) and (3) and using (10) one obtains at order ΔB

$$L_0 \partial\psi_1/\partial x \equiv (\beta_0 \nabla_H^2 + \partial^2/\partial z^2)\partial\psi_1/\partial x - 1/\bar{u}\,(\beta_0 \partial^2\bar{u}/\partial y^2 + \partial^2\bar{u}/\partial z^2)\partial\psi_1/\partial x$$
$$= 0 \qquad (3.2.15)$$

$$\partial^2\psi_1/\partial z\partial x = 0 \quad \text{on} \quad z = \pm 1/2 .$$

So the outer solution to first order is proportional to the gravest mode solution

$$\psi_1 = \{ a(\tau) \cos r\pi x + b(\tau) \sin r\pi x \} \rho_0(y). \qquad (3.2.16)$$

Strictly speaking the most general outer solutions should be matched across the critical layer to reach this conclusion; but the depth independent solution (16) does satisfy the inner equations at lowest order as will become apparent. An undetermined axisymmetric perturbation should also be retained at this point but examination of the problem at the next order ((17) below) shows that it is not excited at this order.

At order ΔB^2 (1) divided by $\bar{u}$ and (2) yield

$$L_0 \partial\psi_2/\partial x = -B_0/\bar{u}\, \nabla_H^2 \partial\psi_1/\partial\tau - 1/\bar{u}\, J(\psi_1, B_0 \nabla_H^2 \psi_1) + 1/(B_0 \bar{u})\, \partial^2\bar{u}/\partial z^2\, \partial\psi_1/\partial x \qquad (3.2.17)$$

$$\partial^2\psi_2/\partial z \partial x = 0 \quad \text{on} \quad z = \pm 1/2 .$$

The l.h.s of (17a) involves an elliptic operator (defined in (15)) on $\partial\psi_2/\partial x$ and the r.h.s. of (17a) terms $-B_0 \partial q_1/\partial\tau$ / $\bar{u}$ and $-J(\psi_1, B_0 q_1)$ / $\bar{u}$ which are singular at z=0 on the edge of the two domains. The forcing terms on the r.h.s. of (17a) do not induce any axisymmetric motions as anticipated in (16). (17) becomes a well defined problem if jump conditions on ψ_2 and $\partial\psi_2/\partial z$ across z=0 are specified

$$[\psi_2]_{0^-}^{0^+} = H(x,y) \quad ; \quad [\partial\psi_2/\partial z]_{0^-}^{0^+} = I(x,y). \qquad (3.2.18)$$

H(x,y) and I(x,y) are understood to depend on the inner solution. One way to solve (17), (18) for the non-axisymmetric part of ψ_2 is to expand the inhomogeneous (forcing) terms on the r.h.s of (17a) and (18) in terms of the eigenfunctions of L_1;

$$L_1 \equiv L_0 - \partial^2/\partial z^2 . \qquad (3.2.19)$$

(17), (18) then reduce to equations of the form

$$
\begin{aligned}
&d^2\Phi_2^{K,\ell}/dz^2 + \alpha^{K,\ell}\Phi_2^{K,\ell} = f^{K,\ell}(z)\\
&[\Phi_2^{K,\ell}]_{0^-}^{0^+} = \hat{H}^{K,\ell}\ , \quad [d\Phi_2^{K,\ell}/dz]_{0^-}^{0^+} = \hat{I}^{K,\ell}\\
&d\Phi_2^{K,\ell}/dz = 0 \qquad \text{on } z = \pm 1/2\\
&\psi_2 = \sum_{K,\ell} \Phi_2^{K,\ell}(z)\,\rho^{K,\ell}(y)\,\exp ir K\pi x\ .
\end{aligned}
\tag{3.2.20}
$$

Note that $\alpha^{K,1} > 0$, except for the gravest solution - which is the neutral normal mode ρ_0. When $\alpha^{K,1} > 0$ the complementary solutions of (20a) enable solutions to be found whatever $f^{K,1}(z)$, $\hat{H}^{K,1}$ and $\hat{I}^{K,1}$. So it is adequate to consider only the projections of the r.h.s. of (17), (18) into $\rho_0(y)$ to find the condition for (17),(18) to have a solution. This projection results in the following problem

$$
\begin{aligned}
&d^2\chi/dz^2 = E/\sin\pi z + F\\
&d\chi/dz = 0 \qquad \text{at } z = \pm 1/2\\
&[\chi]_{0^-}^{0^+} = P\ , \quad [d\chi/dz]_{0^-}^{0^+} = R
\end{aligned}
\tag{3.2.21}
$$

in which P and R are to be determined from the interior solution and χ, E and F are given by

$$
\begin{aligned}
&\chi = \iint \rho_0(y)\,\exp ir\pi x\ \partial\psi_2/\partial x\ dx\,dy\\
&E = -2B_0 \iint \frac{\rho_0(y)\,\exp ir\pi x\ \nabla_H^2\,\partial\psi_1/\partial\tau}{1 - a_s + a_s\sin\pi y}\ dx\,dy\\
&F = -\pi^2/B_0 \iint \rho_0(y)\,\exp ir\pi x\ \partial\psi_1/\partial x\ dx\,dy\ .
\end{aligned}
\tag{3.2.22}
$$

Integrating (21) once yields

$$
\begin{aligned}
d\chi/dz &= E/\pi\{\ln|\tan\pi z/2| - \ln|\tan\pi/4|\} + F(z - 1/2), \quad z > 0\\
&= E/\pi\{\ln|\tan\pi z/2| - \ln|\tan\pi/4|\} + F(z + 1/2), \quad z < 0
\end{aligned}
$$

and hence

$$
\begin{aligned}
[d\chi/dz]_{-z}^{+z} &= F(2z - 1)\\
&\to -F \qquad \text{as } z \downarrow 0\ .
\end{aligned}
\tag{3.2.23}
$$

This condition on the jump in potential temperature across the critical layer provides a relationship between the inner and outer solutions which must be satisfied if the solution sought is to be valid. It is the only consistency condition at this order because the complementary solution χ = const. can accommodate any jump R in the streamfunction. For reference we note the time derivative of (23) (obtained by differentiating w.r.t. τ and using (22))

$$\left[\iint \rho_0(y)\exp ir\pi x\ \partial^2\psi_2/\partial z\partial\tau\ dx dy\right]_{-z}^{+z} \to \frac{\pi^2}{B_0}\iint \rho_0 \exp ir\pi x\ \partial\psi_1/\partial\tau\ dx dy \qquad (3.2.24)$$

as $z \downarrow 0$.

Inner Solution

From (2) and (13) the potential vorticity in the critical layer, Q, is related to the streamfunction, Ψ, by

$$BQ = B\nabla_H^2\Psi + 1/\Delta B^2\ \partial^2\Psi/\partial\zeta^2 + Bf_0 . \qquad (3.2.25)$$

The conservation of potential vorticity, (1), is expressed by

$$\Delta B\ B\ \partial Q/\partial\tau + J(\Psi, BQ) = 0 . \qquad (3.2.26)$$

We pose a series expansion for Ψ in powers of ΔB;

$$\Psi = \Delta B\,\Psi_1 + \Delta B^2\Psi_2 + \Delta B^3\Psi_3 + \ldots . \qquad (3.2.27)$$

For $\zeta \to \pm\infty$, Ψ must match to the outer solution with $z \to 0^{\pm}$ which is given to $O(\Delta B)$ by

$$\psi_{out} = \Delta B\{\pi\zeta/2 - \Delta B^2/2\ (\pi\zeta)^3/6 + O(\Delta B^4)\}\{-(1-a_s)y + a_s/\pi\ \cos\pi y\} \qquad (3.2.28)$$
$$+ \Delta B\,(a(\tau)\cos r\pi x + b(\tau)\sin r\pi x)\,\rho_0(y) + O(\Delta B^2).$$

The terms of $O(\Delta B^3)$ in (28) describe the initial zonal mean flow to $O(\Delta B^3)$. They are included for reference in the derivation of (34) below.

At $O(\Delta B^0)$, (26) is

$$\partial^3 \Psi_1 / \partial \xi^2 \partial \tau \quad + \quad J(\Psi_1 , \partial^2 \Psi_1 / \partial \xi^2) \quad = \quad 0 . \qquad (3.2.29)$$

A (barotropic wave) solution which matches to the outer is

$$\Psi_1 = -\frac{\pi}{2} \xi \{ (1-a_s) y - \frac{a_s}{\pi} \cos \pi y \} + \beta_0 (a(\tau) \cos r\pi x + b(\tau) \sin r\pi x) . \qquad (3.2.30)$$

The mean flow in the outer is time independent so $\partial \bar{\Psi}_1 / \partial \tau = 0$.

At $O(\Delta B^1)$, (26) is

$$\partial^3 \Psi_2 / \partial \xi^2 \partial \tau \quad + \quad J(\Psi_1 , \partial^2 \Psi_2 / \partial \xi^2) \quad = \quad 0 .$$

The partial solution

$$\Psi_2 = \bar{\Psi}_2 (y, \xi) + \sum_{k,l} (c_{k,l} \cos rk\pi x + d_{k,l} \sin rk\pi x) \rho^{k,l}(y) \qquad (3.2.31)$$

is in fact adequate. Solutions with $\partial^2 \Psi_2 / \partial \zeta^2$ non-zero would require potential vorticity gradients of order 1 which cannot develop for ζ of order 1 since Q is conserved by the horizontal geostrophic motion. (Such solutions are probably important in the development of flows with singular critical levels.) Wave solutions directly proportional to ζ are not required because the $O(\Delta B)$ outer streamfunction (16) is barotropic (i.e. $\partial \psi_1 / \partial z = 0$).

At $O(\Delta B^2)$, the highest order considered here, (26) is

$$\partial W / \partial \tau \quad + \quad J(\Psi_1 , W) \quad = \quad 0 \qquad (3.2.32)$$

for W defined by

$$W \equiv (BQ)_1 \quad = \quad B_0 \nabla_H^2 \Psi_1 + \partial^2 \Psi_3 / \partial \xi^2 . \qquad (3.2.33)$$

(32) and (33) assert that the $O(\Delta B)$ potential vorticity, W, is conserved following the $O(\Delta B)$ geostrophic motion. Using (28) one finds that the zonal flow by itself initially has

$$\bar{W} = \frac{\pi^3}{2}\delta^3 \left\{ (1-a_s)y - (1+B_0)\frac{a_s}{\pi}\cos\pi y \right\} . \qquad (3.2.34)$$

The inviscid boundary conditions (4) imply that $\Psi_3' = 0$ on $y = 0,1$ and (30) with (8) implies that $\nabla_H^2 \Psi_1' = 0$ on $y = 0,1$ so

$$W' = 0 \qquad \text{on} \quad y = 0,1 . \qquad (3.2.35)$$

Furthermore $\partial^2\bar{\Psi}_1/\partial y\partial\tau = 0$ and $\partial^2\bar{\Psi}_3/\partial y\partial\tau = 0$ on $y = 0,1$. For a pure wave perturbation $\bar{\Psi}_3 = 0$ at time $\tau = 0$ so using (34)

$$\partial\bar{W}/\partial y = \frac{\pi^3}{2}\delta^3 (1 - a_s) \qquad \text{on} \quad y = 0,1 . \qquad (3.2.36)$$

Matching conditions

The streamfunctions of the inner and outer solutions may be matched together using the asymptotic matching principle (van Dyke 1975):

expansion to $O(\Delta B^m)$ in ζ of the outer solution in z to $O(\Delta B^n)$
= expansion to $O(\Delta B^n)$ in z of the inner solution in ζ to $O(\Delta B^m)$
(3.2.37)

It is adequate here to match only the jump in $\partial\chi/\partial z$ given in the outer solution by (23). Because of the lack of vertical derivatives in Ψ_1', Ψ_2' and ψ_1', the first condition imposed by (37) is obtained with m=3 and n=2;

$$\left[\partial\Psi_{in}'/\partial z\right]_{-z_0}^{+z_0} = \left[\tfrac{1}{\Delta B}\partial\Psi_{in}'/\partial\xi\right]_{-z_0}^{+z_0} = \left[\Delta B^2\partial\Psi_3'/\partial\xi\right]_{\xi=-z_0/\Delta B}^{\xi=+z_0/\Delta B}$$
$$= \Delta B^2 \int_{-z_0/\Delta B}^{+z_0/\Delta B} \partial^2\Psi_3'/\partial\xi^2\, d\xi . \qquad (3.2.38)$$

So from (24) and (38) the matching condition (37) implies the following solvability condition.

$$\int_{-1/\delta}^{+1/\delta}\iint \rho_0 \exp i\pi x \frac{\partial^3\Psi_3}{\partial\xi^2\partial\tau}\, dx\,dy\,dz \to \frac{\pi^2}{B_0}\iint \rho_0 \exp i\pi x \frac{\partial\psi_1}{\partial\tau}\, dx\,dy \qquad (3.2.39)$$
$$\text{as } \delta \downarrow 0.$$

Initial conditions

An appropriate expression for the linearly unstable normal mode initial condition may be found by linearising (1) about the zonal flow (5),

$$\partial q'/\partial \tau + \bar{u}\,\partial q'/\partial x + \bar{q}_y\,\partial \psi'/\partial x = 0 \qquad (3.2.40)$$

and letting $B = B_0 - \Delta B$ and

$$\psi' = \mathrm{Re}\{\phi(y,z)\exp ir\pi(x-ct)\}. \qquad (3.2.41)$$

Posing an expansion for ϕ in powers of ΔB based on the neutral normal mode ϕ_0 marking the SWC,

$$\phi = \Delta B\,\phi_0(y) + \Delta B^2\phi_1 + \ldots \quad ; \quad c = \Delta B c_1 + \ldots \qquad (3.2.42)$$

$$(d^2/dy^2 - r^2\pi^2)\phi_0 + \bar{q}_y/\bar{u}\,\phi_0 = 0 \quad , \quad \phi_0 = 0 \text{ at } y = 0,1 \ , \qquad (3.2.43)$$

at order ΔB^2 one finds that

$$(B_0\nabla_H^2 + \partial^2/\partial z^2 + B_0\bar{q}_y/\bar{u})\phi_1 = -\frac{\pi^2}{B_0}\phi_0 - B_0 c_1 \frac{\bar{q}_y/\bar{u}\,\phi_0}{\bar{u}-c} \qquad (3.2.44)$$

$$\partial\phi_1/\partial z = 0 \quad \text{on} \quad z = \pm 1/2 \quad ; \quad \phi_1 = 0 \quad \text{on} \quad y = 0,1.$$

The Fredholm condition for (44) to have a solution (appendix A) is that the free (complementary) solution ρ_0 be orthogonal to the forcing of ϕ_1. This condition determines the dependence of c on ΔB. Standard manipulations (described in more detail in section 2.2) show that c_1 is pure imaginary and

$$\iint \rho_0^2\,\pi^2/B_0\,dy\,dz = -\iint c_1 \frac{B_0(\bar{q}_y/\bar{u})\rho_0^2}{\bar{u}-c}\,dy\,dz = \pi|c_1|\int_y \frac{B_0(\bar{q}_y/\bar{u})\rho_0^2}{|\partial\bar{u}/\partial z|_{z=0}}\,dy \ . \qquad (3.2.45)$$

So

$$c_1 = \frac{i\mathcal{R}}{2B_0^2} \quad ; \quad \mathcal{R} = \frac{\int \pi^2\rho_0^2\,dy}{\int \rho_0^2(\bar{q}_y/\bar{u})(1-a_s+a_s\sin\pi y)^{-1}\,dy} \qquad (3.2.46)$$

The ratio $R = B_0$ when $a_s = 0$ and is not far from 1 for $a_s \simeq 0.5$.

Manipulation of (32) using (33), (30), (6) and (8) shows that $\partial^2 \Psi_3'/\partial\zeta^2$ for this normal mode satisfies

$$(\bar{u} - c)\partial^3 \Psi_3' / \partial \xi^2 \partial x \;=\; B_0 (\bar{q}_y / \bar{u})\, \partial \Psi_1' / \partial \tau \,. \qquad (3.2.47)$$

So the normal mode with an initial outer streamfunction of small amplitude ε

$$\Psi_1' \;=\; \varepsilon \cos r\pi x \;\rho_0(y) \qquad (3.2.48)$$

has accompanying vertical variation

$$\partial^2 \Psi_3' / \partial \xi^2 \;=\; \frac{\varepsilon B_0 (\bar{q}_y/\bar{u}) \rho_0 \{ |c|^2 \cos r\pi x + \tilde{u} |c| \sin r\pi x \}}{|\tilde{u} - c|^2} \qquad (3.2.49)$$

in which

$$c = \Delta B\, c_1 \quad ; \quad \tilde{u} \;=\; \Delta B \frac{\pi}{2} \xi \,(\, 1 - a_s + a_s \sin \pi y \,). \qquad (3.2.50)$$

Section 3.3 A NUMERICAL STUDY OF THE NON-LINEAR EVOLUTION

Summary of analysis

The previous section presents an analysis of the development of a wave on a baroclinic zonal flow, (3.2.5), which has $\bar{q}_y = \bar{u} = 0$ on $z=0$, starting from a pure unstable normal mode disturbance of small amplitude. The analysis is set at a Burger number, $B_0-\Delta B$, slightly below the SWC value B_0, where there is only one linearly unstable normal mode and its evolution is slow. The wave amplitude is argued to become of order ΔB and to be governed by the evolution in a critical layer with a depth of order ΔB. The wave's streamfunction to order ΔB is barotropic and the same inside as outside the critical layer (see (3.2.30) and (3.2.16)). It has the same spatial form as the neutral normal mode (3.2.7). The $O(\Delta B)$ potential vorticity within the critical layer, W defined by (3.2.33), is conserved following the $O(\Delta B)$ geostrophic motion (3.2.32); the boundary conditions on W are provided by (3.2.35) and (3.2.36). The amplitudes $a(\tau)$ and $b(\tau)$ of the $O(\Delta B)$ wave streamfunction are determined by the constraint (3.2.39) which ensures that the rate of change of the jump in the thermal field across the critical layer (caused by potential vorticity advection within it) does not resonantly force the neutral normal mode outside the critical layer. Expressing $\partial^3\Psi_3/\partial\tau\partial\zeta^2$ in (3.2.39) in terms of $J(\Psi_1, W)$ and Ψ_1 using (3.2.32) and (3.2.33) and then using (3.2.30), yields

$$\begin{aligned}
&\int_\zeta \iint \rho_0 \exp(ir\pi x)\,.\,-J(\Psi_1, W)\,dx\,dy\,d\zeta \\
&= \iint \frac{\pi^2}{B_0}\rho_0^2 \exp(ir\pi x)\,(\partial a/\partial\tau \cos r\pi x + \partial b/\partial\tau \sin r\pi x)\,dx\,dy \\
&\quad + \int_\zeta \iint -B_0\,(\bar{q}_y/\bar{u})\,\rho_0^2 \exp(ir\pi x)(\partial a/\partial\tau \cos r\pi x + \partial b/\partial\tau \sin r\pi x)\,dx\,dy\,d\zeta .
\end{aligned} \tag{3.3.1}$$

Thus the evolution is determined by (3.2.32) with Ψ_1 given by (3.2.30) which must be solved subject to (3.2.35), (3.2.36) and (1). These equations are solved numerically by representing W at a given time step on a discrete uniform 3D grid and using standard finite

difference expressions for $J(\Psi_1, W)$ and the boundary conditions. The time derivatives of W, $a(\tau)$ and $b(\tau)$ calculated from (3.2.32) and (1) are used in a leapfrog scheme to infer W, a and b at the next time step. Integrations commence with a normal mode wave perturbation of amplitude $a(\tau=0) = \varepsilon$ ($b(0) = 0$) for which W is given by (3.2.33), (3.2.34) and (3.2.48) - (3.2.50).

Details of numerical method

The grid on which W is stored has IXPTS along the channel with the ith x ordinate, $x_i = L_x/L_y$ i/IXPTS; IYPTS across the channel with y_j = (j-0.5)/IYPTS; and IZPTS in the vertical arranged symmetrically about $\zeta=0$ with a uniform level spacing $\Delta\zeta$. The boundary conditions (3.2.35) and (3.2.36) are applied by setting

$$W(y_0) = -W(y_1) + 2\bar{W}(y_1) - \pi^3 \beta (1-a_s)/(2.\,IYPTS)$$

$$W(y_{IYPTS+1}) = -W(y_{IYPTS}) + 2\bar{W}(y_{IYPTS}) + \pi^3 \beta (1-a_s)/(2.\,IYPTS) \qquad (3.3.2)$$

which is accurate to second order in Δy. $J(\Psi_1, W)$ is calculated using Arakawa's energy and enstrophy conserving form (Haltiner & Williams 1980 pp175-6). The l.h.s. of (1) is calculated by simple summation over all the grid points. The integrals multiplying $\partial a/\partial\tau$ and $\partial b/\partial\tau$ in (1) are depth independent and found by simple summation across the channel using the values of ρ_0 and $(\bar{q}_y/\bar{u})$ on the grid; ρ_0 is calculated prior to the integration by a second order finite difference approximation to (8) on the model's lateral grid and normalised by setting its maximum value to 1. Excitation of the computational mode during the course of the integrations is avoided by using Miyakoda's method for leapfrogging from a small initial forward time step to the standard time step (Hoskins & Simmons 1975) and employing a weak Robert filter (Haltiner & Williams (1980) p147).

The time step must be chosen so as to satisfy the CFL criterion ($U\,\Delta\tau < \Delta x$) at the top and bottom levels and the depth of the layer ($\Delta\zeta$. IZPTS) needs to be large enough and the vertical resolution ($\Delta\zeta$) fine enough to give a fair representation of the critical layer integral

$$I(h) \equiv \int_{-h}^{+h} c_1/(\zeta^2 + c_1^2)\, d\zeta = 2\tan^{-1}(h/c_1). \qquad (3.3.3)$$

The initial amplitude of the wave should be chosen to ensure that W has no closed contours within the critical layer initially. These form first (as ε is increased) near the sidewall. Comparison of the initial fields of $\partial\bar{W}/\partial y$ and $\partial W'/\partial y$ shows that $|\partial\bar{W}/\partial y| > |\partial W'/\partial y|$ and hence there are no closed contours when

$$|\partial\rho_0/\partial y|\, \varepsilon < c_1 \qquad \text{at } y = 0, 1. \qquad (3.3.4)$$

Finally the grid must be be symmetric about $\zeta=0$ to avoid misrepresentation of the integral on the l.h.s. of (1), which has an antisymmetric logarithmic dependence on ζ (see the discussion of matching in section 3.4).

Numerical results

The presentation here is limited to results for the flow (3.2.5) with $a_s=0.5$ in a channel with $L_x = 2\pi L_y$. This value of a_s is midway between the special values $a_s=0$ and $a_s=1$ for which $J(\Psi_1, W)=0$ (see (3.1.3) and (3.2.6)). The slow growth rate for this value, $c_1 \simeq 0.2$ (see (3.2.46)), implies that $F(h)\equiv I(h)/I(\infty)$, the fraction of the critical layer integral (3) contributed within depth h of the critical layer, is 0.75 for h=0.5 and 0.87 for h=1.0. Three integrations (A, B and C) with $\Delta\zeta=0.1$ have been investigated. The grids used were

run A IXPTS = IYPTS = 20 ; IZPTS = 21
run B IXPTS = IYPTS = 20 ; IZPTS = 11
run C IXPTS = IYPTS = 40 ; IZPTS = 11

The time step $\Delta\tau = 0.015$ is sufficiently short to avoid numerical instability and was used in all the integrations as was the initial amplitude $\varepsilon = 0.02$ which avoids closed contours (see (4)).

Fig. 3.1 presents the evolution of the amplitude $a(\tau)$ over the 30 slow time units of each integration (A full line, B dashed line,

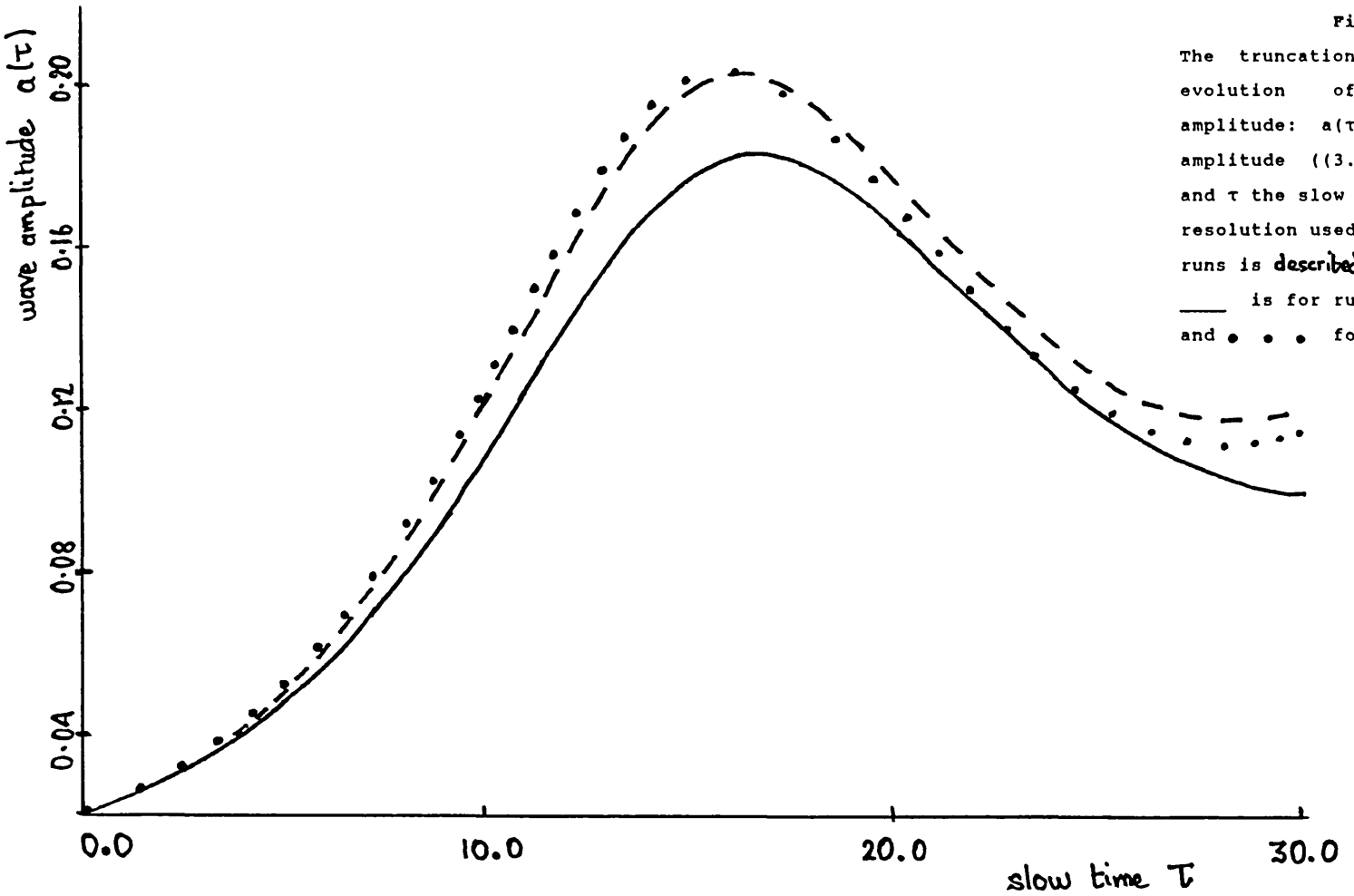


Figure 3.1

The truncation dependence of the evolution of the WNL wave's amplitude: a(τ) is the small wave amplitude ((3.2.12) and (3.2.16)) and τ the slow time (τ=ΔB t). The resolution used in each of the three runs is described on page 130. Curve ____ is for run A; - - - for run B; and • • • for run C.

C circles). The evolution is qualitatively similar in all three cases. Over the first 10 slow time units, $a(\tau)$ grows approximately exponentially, doubling its amplitude every 4 time units. It reaches its peak amplitude between τ = 15 and 20. The maximum amplitude and growth rate of run A are smaller than those of runs B and C probably because run A covers a greater depth of the critical layer than runs B and C; the amplitudes in fig. 3.1 are probably overestimates of the true values.

The amplitude $b(\tau)$ was found to remain very close to zero throughout all integrations. This feature may be understood by consideration of solutions of the form

$$W = \sum_{n=1}^{\infty} e_n(\zeta)\{ f_n(y) \sin 2n\pi x + g_n(y) \cos(2n-1)\pi x \} + \sum_{n=0}^{\infty} o_n(\zeta)\{ h_n(y) \sin(2n+1)\pi x + j_n(y) \cos 2n\pi x \} \qquad (3.3.5)$$

in which o_n is an odd function of ζ and e_n an even function and f_n, g_n, h_n and j_n are arbitrary functions of y. Investigation of (3.2.32) using (3.2.30) reveals that when $b(\tau)$ = 0, a solution W which has the form (5) initially maintains it. When W is given by (5) it contains no amplitude in an even function of ζ with sin $r\pi x$ dependence so from (1) $\partial b/\partial\tau = 0$ for solutions of this form. The initial condition (3.2.48) - (3.2.50) is also of this form so $b(\tau)$ = 0 and W maintains the symmetries of (5).

The potential vorticity of the zonal flow is zero on the critical level. Since W is conservatively advected by horizontal motions it remains zero on ζ=0 for all τ; that the initial state does have W=0 on ζ=0 may be verified from (3.2.48) - (3.2.50). The initial field $W(\tau=0)$ and the streamfunction Ψ_1 (with $b(\tau)$=0) at all times τ also have the following symmetry

$$W(-x + L_x/2L_y, y, -\zeta, \tau=0) = -W(x, y, \zeta, \tau=0)$$
$$\Psi_1(-x + L_x/2L_y, y, -\zeta, \tau) = -\Psi(x, y, \zeta, \tau).$$

Substitution of these expressions in (3.2.32) shows that

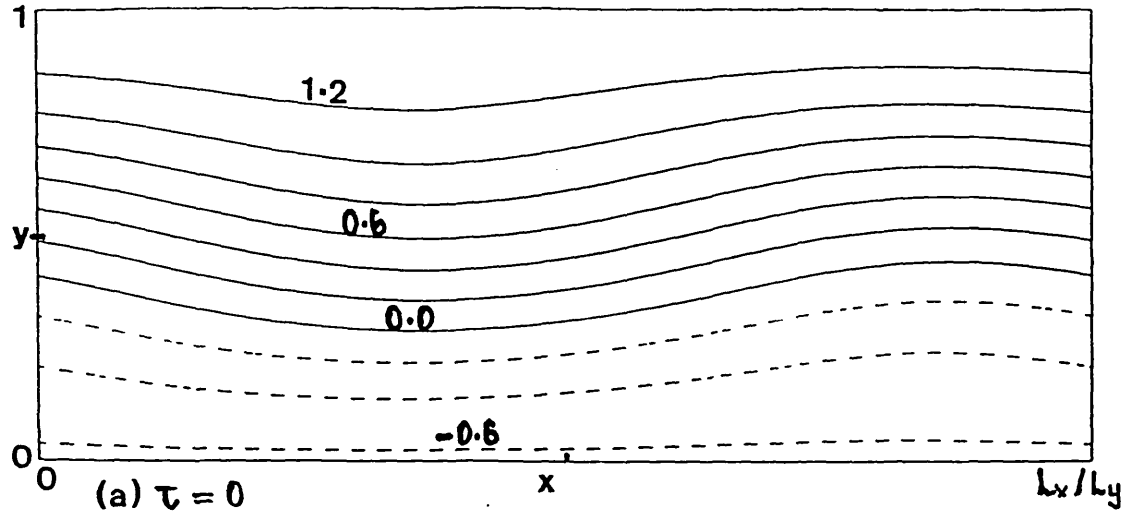
$$\partial W/\partial\tau(-x + L_x/2L_y, y, -\zeta, \tau=0) = -\partial W/\partial\tau(x, y, \zeta, \tau=0).$$

Consequently

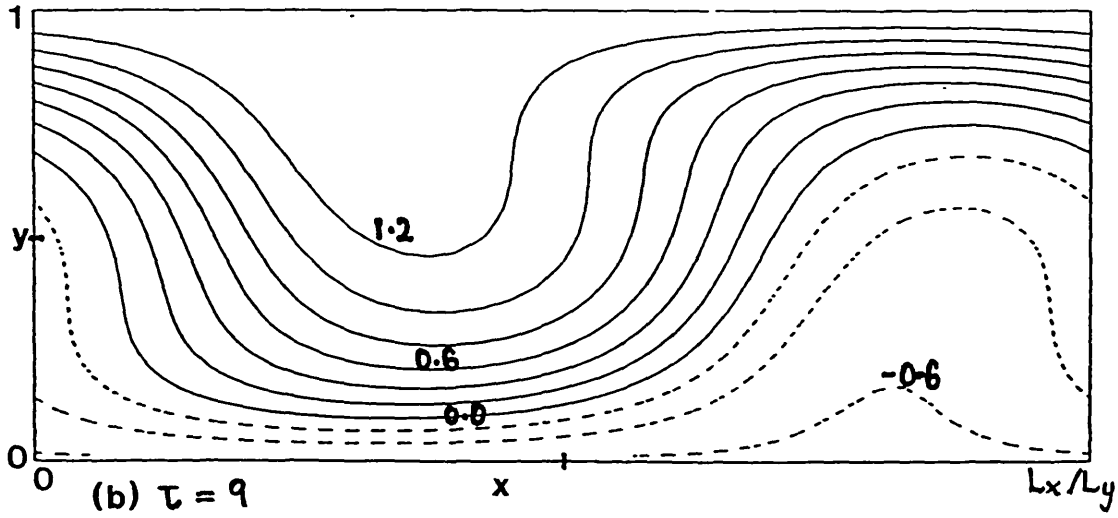
$$W(-x + L_x/2L_y\ ,\ y\ ,\ -\zeta\ ,\ \tau) = -W(x, y, \zeta, \tau) \qquad (3.3.6)$$

- the potential vorticity field, W, above the critical level may be used to infer W below the critical level.

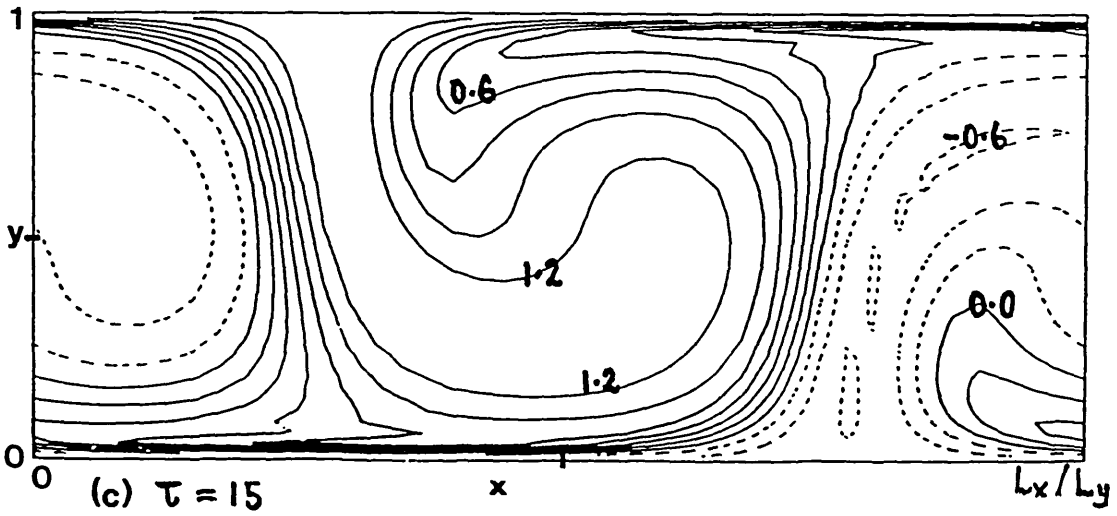
Fields of W from run C on the horizontal section $\zeta = 0.1$ are presented in fig. 3.2 for times a) $\tau = 0$, b) $\tau = 9$, c) $\tau = 15$, d) $\tau = 18$ and e) $\tau = 21$. Horizontal sections of W from the same run at a) $\zeta = -0.1$, b) $\zeta = 0.2$ and c) $\zeta = 0.4$ for $\tau = 15$ are presented in fig. 3.3. From fig. 3.2.a it is clear that the initial potential vorticity distribution contains a small amplitude wave perturbation of the zonal flow field. The wave has reached a large amplitude at $|\zeta| = 0.1$ by $\tau = 9$ (fig. 3.2.b) and by $\tau = 15$ (fig. 3.2.c) the effects of the wave's self advection are unmistakable; the main lobe has been wrapped round on itself and strongly sheared regions are becoming apparent. Around $x \simeq 0.25\ L_x/L_y$, $y \simeq 0.1$ and $x \simeq 0.75\ L_x/L_y$, $y \simeq 0.9$ the advecting flow has the characteristics of a deformation field. The sharp corners of W which develop in these regions by $\tau = 15$ appear to generate the grid scale wave disturbances which contaminate the solutions at times $\tau=18$ (fig. 3.2d) and $\tau=21$ (fig. 3.2e). It is argued in section 3.4 that these are not numerical representations of real small scale instabilities. The formation between $\tau=9$ and $\tau=15$ and narrowing between $\tau=15$ and $\tau=18$ of a tail of high potential vorticity fluid ($W \geq 1.2$) around $x \simeq 0.25\ L_x/L_y$, $y \simeq 0.5$ appears to be well represented though the $W=1.2$ contour is lost from the tail by $\tau=21$. The emergence of $W=1.4$ contours by $\tau=18$ and $\tau=21$ is further evidence of non-conservation of W within the integration. Figure 3.3 shows that the rolling up of the potential vorticity field has also progressed considerably at $\zeta=0.2$ (fig 3.3b) and begun at $\zeta=0.4$ (fig. 3.3c) by $\tau=15$. The wrapping clearly proceeds most rapidly in the layers closest to the critical level. The symmetry of the solution about the critical level, expressed by (6), is evident at $\tau=15$ from fig 3.3a for $\zeta=-0.1$ and fig. 3.2c for $\zeta=0.1$.



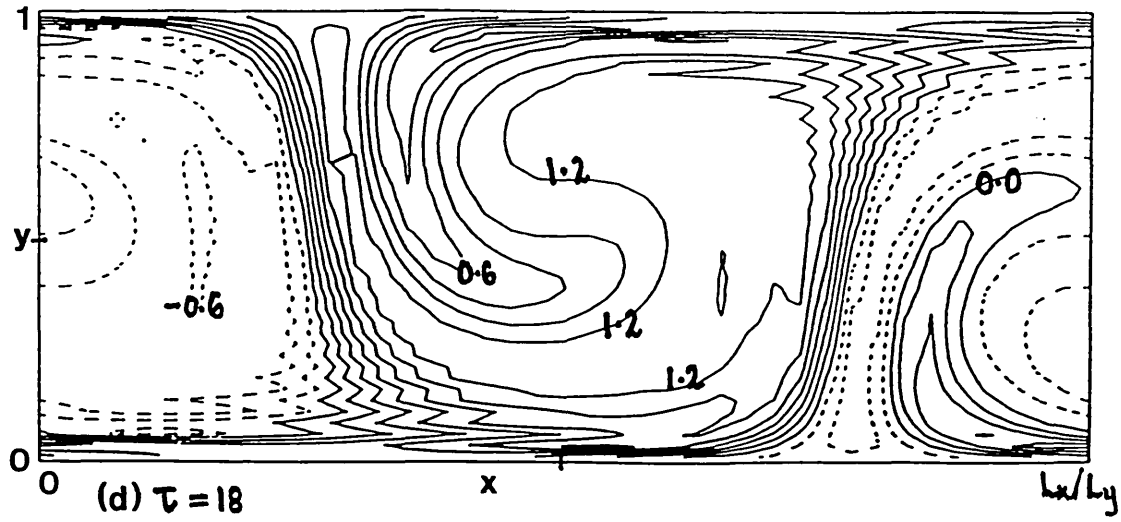
(a) $\tau = 0$



(b) $\tau = 9$



(c) $\tau = 15$



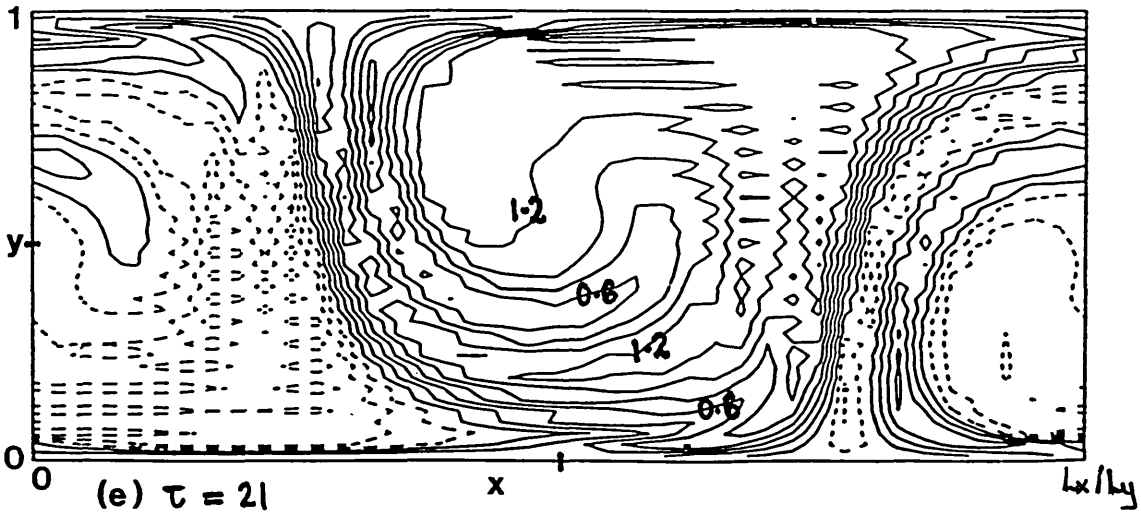


Figure 3.2

The evolution of the potential vorticity distribution just above the steering level. The horizontal cross-sections present W, the O(ΔB) potential vorticity of the inner solution, (3.2.33), obtained in run C at ζ=0.1 at times (a) τ=0.0, (b) τ=9.0, (c) τ=15.0, (d) τ=18.0, (e) τ=21.0.

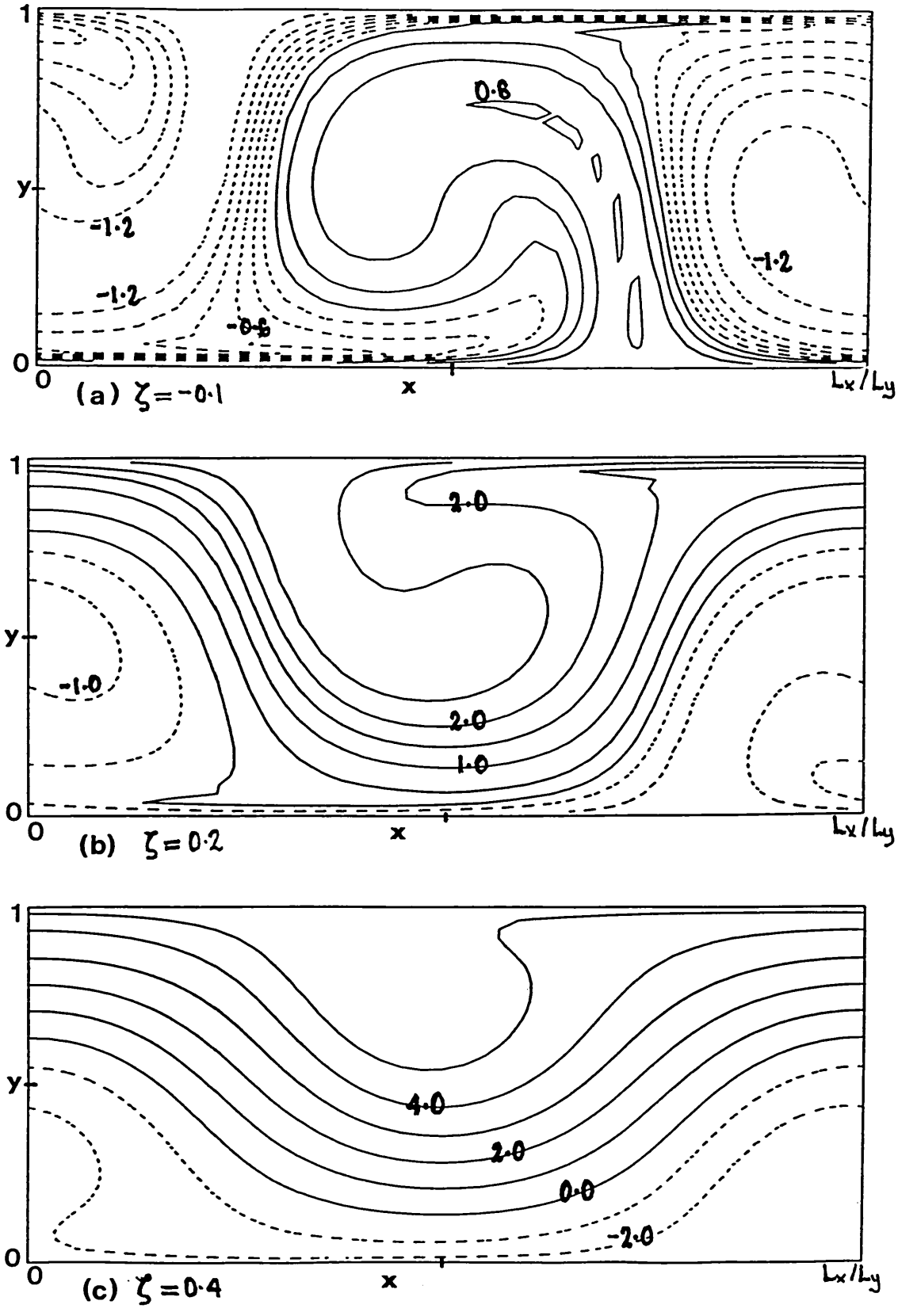


Figure 3.3

The distribution of the potential vorticity within the critical layer at time $\tau=15$. The horizontal cross-sections of W from run C are presented at heights (a) $\zeta=-0.1$, (b) $\zeta=0.2$ and (c) $\zeta=0.4$ (fig. 3.2(c) presents the cross-section for $\zeta=0.1$).

Section 3.4 DISCUSSION

The simulation described above exhibits several features of the irreversible deformation of the potential vorticity field within a critical layer suggested to be of importance in stratospheric dynamics by McIntyre & Palmer (1983, 1984, 1985). But since it produces a flow with monotonically increasing small scale structure it appears to be unable to give a proper description of the perfectly periodic phenomenon of amplitude vacillation found in rotating annulus experiments. Since the aim of this work is to illuminate amplitude vacillation near the UAT it is important to discuss aspects of the model's formulation which may be unsound or unrealistic and responsible for this qualitative difference from the laboratory flows.

Validity of Matching

The initial condition (3.2.49) for $\partial^2 \Psi_3'/\partial\zeta^2$ involves a term which is essentially proportional to $1 / \zeta \sin r\pi x$. Hence

$$\int_1^z W \exp ir\pi x \, d\zeta \sim \ln z .$$

It is crucial for the validity of the solutions presented that these logarithmic terms are antisymmetric in ζ so that

$$\int_{-\infty}^{+\infty} W \exp ir\pi x \, d\zeta - \int_{-z}^{+z} W \exp ir\pi x \, d\zeta \sim 1/z .$$

Furthermore the solution is not developed far enough in section 3.2 to provide confirmation that matching is proceeding satisfactorily. Some confidence concerning these points may be gained by a comparison of the equations governing the outer solution for small z with that of the inner for large $|\zeta|$. One can only hope to achieve matching if there is an intermediate range of values of z ($\Delta B \ll z \ll 1$) for which the inner and outer equations are both valid.

Writing

$$\Omega = \partial^2 \Psi_3 / \partial \zeta^2 + \pi^2 \bar{\Psi}_1 \qquad (3.4.1)$$

it is possible to reduce (3.2.32) to

$$\partial \Omega / \partial \tau - B_0 \bar{q}_y / \bar{u} \, \partial \Psi_1' / \partial \tau + J(\Psi_1, \Omega) + J(\Psi_1', -B_0 \bar{q}_y / \bar{u} \, \Psi_1') = 0 \qquad (3.4.2)$$

by using relations (3.2.6) and (3.2.7). For large ζ, $J(\bar{\Psi}_1, \Omega)$ dominates (2). Given suitable starting conditions Ω' will decay with $1/\zeta$ so that the principal source terms for Ω' in (2) are

$$\partial \Omega / \partial x \simeq \frac{\Delta B}{\bar{u}} B_0 \, \bar{q}_y / \bar{u} \, \partial \Psi_1' / \partial \tau + \frac{\Delta B}{\bar{u}} J(\Psi_1', B_0 \bar{q}_y / \bar{u} \, \Psi_1') , \qquad (3.4.3)$$

$(\partial \Omega / \partial \tau)/\bar{u}$ and $\{J(\Psi_1', \Omega)\}/\bar{u}$ being comparatively negligible. The principal source terms in (3) are hence the same as those for small z in (3.2.17). The presence of those singular terms suggests that $L_0 \partial \psi_2 / \partial x$ will be dominated by vertical derivatives for small z in which case (3.2.17) is, for small z, approximated by

$$\partial / \partial x \, \partial^2 \psi_2 / \partial z^2 \simeq -\frac{B_0}{\bar{u}} \nabla_H^2 \partial \psi_1' / \partial \tau - \frac{1}{\bar{u}} J(\psi_1', B_0 \nabla_H^2 \psi_1') . \qquad (3.4.4)$$

So the equation governing the outer solution for small z is the same as that governing the inner for large ζ. (3) shows that at large ζ, $\Omega = O(1/\zeta)$ is strictly antisymmetric in ζ; Ω will involve symmetric terms only of order $1/\zeta^2$. So the convergence of integrals of Ω calculated symmetrically in ζ will not cause serious error. This point is consistent with the cancellation of the logarithmic terms in the derivation of (3.2.23).

Stability

Fig. 3.2 demonstrates that strong horizontal gradients of potential vorticity develop within the critical layer as a result of advection by the streamfunction Ψ_1. There are regions where the gradients are approximately perpendicular to the streamfunction contours and of opposite sign above and below the critical layer (i.e. at ζ = +0.1 and -0.1). Killworth & McIntyre (1985) and Haynes (1985) have demonstrated that Stewartson's analytical solution for the development of a forced barotropic Rossby wave

impinging on a critical layer, which develops similar reversals in Q gradients, is unstable. So it is important to establish whether the solution presented here is subject to similar instabilities. The first point worth noting is that any solution of (3.2.32) subject to (3.2.39) excludes such instabilities because the advecting streamfunction is forced to be of the form (3.2.30). Linear perturbations to waves (other than the gravest mode) are conservatively advected by (3.2.32). Instabilities of azimuthal wavenumber k > 1 require a perturbation streamfunction $\delta\Psi_1'$ of wavenumber k within the critical layer. Now $\delta\psi_1'$ will have hyperbolic dependence on z outside the critical layer

$$\begin{aligned} \delta\psi_1' &\simeq a \cosh \alpha^{k,l}(z-1/2)\,\rho^{k,l}(y) \qquad z>0 \\ &\simeq a \cosh \alpha^{k,l}(z+1/2)\,\rho^{k,l}(y) \qquad z<0. \end{aligned} \tag{3.4.5}$$

Hence $\partial^2/\partial\zeta^2(\delta\Psi_1')$ within the critical layer would need to be non-zero to match to the outer solution. As noted after (3.2.31), this quantity and even $\partial^2/\partial\zeta^2(\delta\Psi_2')$ are zero within a non-singular critical layer. So it appears that the first order streamfunction in the critical layer is forced to be proportional to that of the neutral normal mode and that instabilities involving other azimuthal wavenumbers are only possible in higher order solutions.

Singular critical layers

Figs. 2.1b and 2.1c suggest that the $\partial\bar{q}/\partial y = 0$ contour of the zonal flow slants strongly across isotach contours. These figures may be misleading in this regard since $\partial\bar{q}/\partial y$ is likely to be sensitive to the choice of vertical ordinate and calculations using (x,y,T) coordinates (T being temperature) could well produce contours of $\partial\bar{q}/\partial y$ much more closely parallel to isotachs. It would be remarkable, however, if the axisymmetric flow had a $\partial\bar{q}/\partial y = 0$ level strictly coincident with an isotach. Unfortunately an analysis of the evolution of a singular normal mode in such a critical layer may well be quite different from that just presented and considerably more difficult. A preliminary obstacle to such an analysis is the lack of explicit solutions for singular neutral

modes on laterally sheared flows. It is quite conceivable that instabilities associated with neutral modes with singular critical levels could be present at higher Burger numbers than those resulting from non-singular normal modes. So it may be helpful to think of non-singular neutral modes as the markers of the main UAT and instabilities arising from singular neutral modes (with $\partial\bar{q}/\partial y$ small where $u = c_r$) as responsible for the weak waves found above the main transition. The analysis of section 3.2 is offered as a model of amplitude vacillation just below the UAT on the assumption that departures of isotachs from the $\partial\bar{q}/\partial y = 0$ contour whilst important in very small amplitude waves can be neglected in waves of moderate amplitude.

Vertical advection within the critical layer

The restriction of the model to waves of moderate amplitude also emerges from an examination of the conditions under which the solution satisfies the quasi-geostrophic equations. The neglect of the ageostrophic horizontal velocities requires $Ro \ll \Delta B$. It is plausible that no qualitative errors would arise from their neglect but as we now show the neglect of vertical velocity advection also requires $Ro \ll B_0 \Delta B$. Let w

$$w = (L_y w^*) / (U H) \tag{3.4.6}$$

be the non-dimensional form of the vertical velocity w* and other non-dimensional quantities be defined as in (2.1.12). Then according to quasi-geostrophic theory (see (2.1.9))

$$w \simeq - R_o / B \; . \; D_g/Dt \; \partial\psi/\partial z \tag{3.4.7}$$

In the critical layer

$$D_g/Dt \; \partial\psi/\partial z \simeq \Delta B \; \partial\Psi'/\partial x \; \partial^2\bar{\Psi}_1/\partial y \partial z$$

so

$$\frac{w \, \partial/\partial z}{u \, \partial/\partial x} \sim \frac{R_o}{B \, \Delta B} \tag{3.4.8}$$

and vertical velocity advection can only be neglected if $Ro \ll B_0 \Delta B$.

Influences of diffusive transports

Diffusive transports have been deliberately neglected to this point in an attempt to establish whether inviscid processes can describe amplitude vacillation successfully. The largest diffusive transports will be confined to the Ekman and side boundary layers. The lateral shear of the zonal flow at the boundaries will, however, cause Ekman pumping of axisymmetric vertical motion which will affect the whole of the flow. The stability of such a zonal flow with its meridional circulation is a major problem in its own right.

Diffusion can, of course, play a dominant role in the critical layer (Drazin & Reid 1981 p421). Conditions under which advection dominates diffusion may be found by inserting the scaling used in section 3.2 into the diffusive version of the potential vorticity equation, which in dimensional form is

$$D_g/Dt^* \, q_v^* = \left(\nu^* \nabla_H^{*2} + \kappa^* \frac{f^2}{N^2} \frac{\partial^2}{\partial z^{*2}} \right) \nabla_z^{*2} \psi^* . \qquad (3.4.9)$$

At order (ΔB^2) in the critical layer one obtains

$$\Delta B^2 \, D_g W/Dt = \frac{\kappa L^2}{\Delta B H^2} \frac{\partial^4 \Psi_z}{\partial S^4} + \Delta B \left(\kappa + \nu B \frac{L^2}{H^2} \right) \nabla_H^2 \frac{\partial^2 \Psi_z}{\partial S^2} + \Delta B \, \nu B \, \nabla_H^4 \Psi_1 \qquad (3.4.10)$$

where $\nu = \nu^*/(UL)$, $\kappa = \kappa^*/(UL)$.

In typical laboratory experiments $L/H \simeq 1/3$, $\upsilon^*/\kappa^* \simeq 10$ (for water glycerol solutions) and, near the UAT, $B \simeq 1$. In these conditions the inviscid WNL critical layer theory requires

$$\kappa L^2/(\Delta B^3 H^2) \ll 1 \quad \text{and} \quad \nu B/\Delta B \ll 1 . \qquad (3.4.11)$$

Taking $\kappa^* = 1.3.10^{-3}$ $cm^2 s^{-1}$, L=5cm, H=15cm, U=0.5cm/s and $\upsilon = 10\kappa$, the conditions in (11) reduce to

$$\Delta B^3 \gg 10^{-4} \quad , \quad \Delta B \gg 5.10^{-3}$$

which are satisfied moderately well when $\Delta B \geq 0.2$.

The generation of sharp gradients of W within the critical layer makes dissipative processes important even when ΔB is large enough for them to be formally small according to the above scaling arguments. Diffusive transports will control the total enstrophy within the critical layer and dissipate the strongest gradients of W. Vertical diffusion may act to re-establish the potential vorticity gradient of the zonal flow within the critical layer, if it becomes depleted by horizontal diffusion. Horizontal diffusion can also control the roughness which develops in the inviscid integrations as the following, somewhat ad hoc, integration shows.

Horizontal diffusion is represented qualitatively by replacing (3.2.32) with

$$\partial W/\partial \tau = -J(\Psi, W) + \tilde{\nu}\{(L_x/2L_y)^2 \partial^2/\partial x^2 + \partial^2/\partial y^2\} W \qquad (3.4.12)$$

and $-J(\Psi,W)$ in the l.h.s. of (3.3.1) with the r.h.s. of (12). In the numerical code the r.h.s. of (12) is calculated at time step n using values of W at step n in $J(\psi,W)$ and values of W at step n-1 in the diffusive term. This method is stable when

$$4\tilde{\nu}\ (L_x/L_y)^2\ \Delta t\ /\ \Delta x^2 < 1$$

(Roache 1976 p61). Horizontal sections of W at ζ=0.1 for a) τ=9, b) τ=18 and c) τ=27 from an integration using the same grid and initial conditions as in figs. 3.2 and 3.3 and $\tilde{\nu} = 2 \cdot 10^{-4}$ are presented in fig. 3.4. It is clear that even such a small horizontal diffusion coefficient has a significant impact on the integration, controlling grid scale roughness and opposing the tightening of W contours.

Values of $\tilde{\nu}$ representative of Hignett's experiments can be found by comparing the second term of the r.h.s. of (10) with the l.h.s. of (10);

$$\tilde{\nu} \simeq (\kappa + \nu B L^2/H^2)/\Delta B \simeq 10^{-3}/\Delta B.$$

Clearly values of $\tilde{\nu} \geq 10^{-3}$ are appropriate. Unfortunately the initial evolution of the amplitude of the wave modelled by (12) is

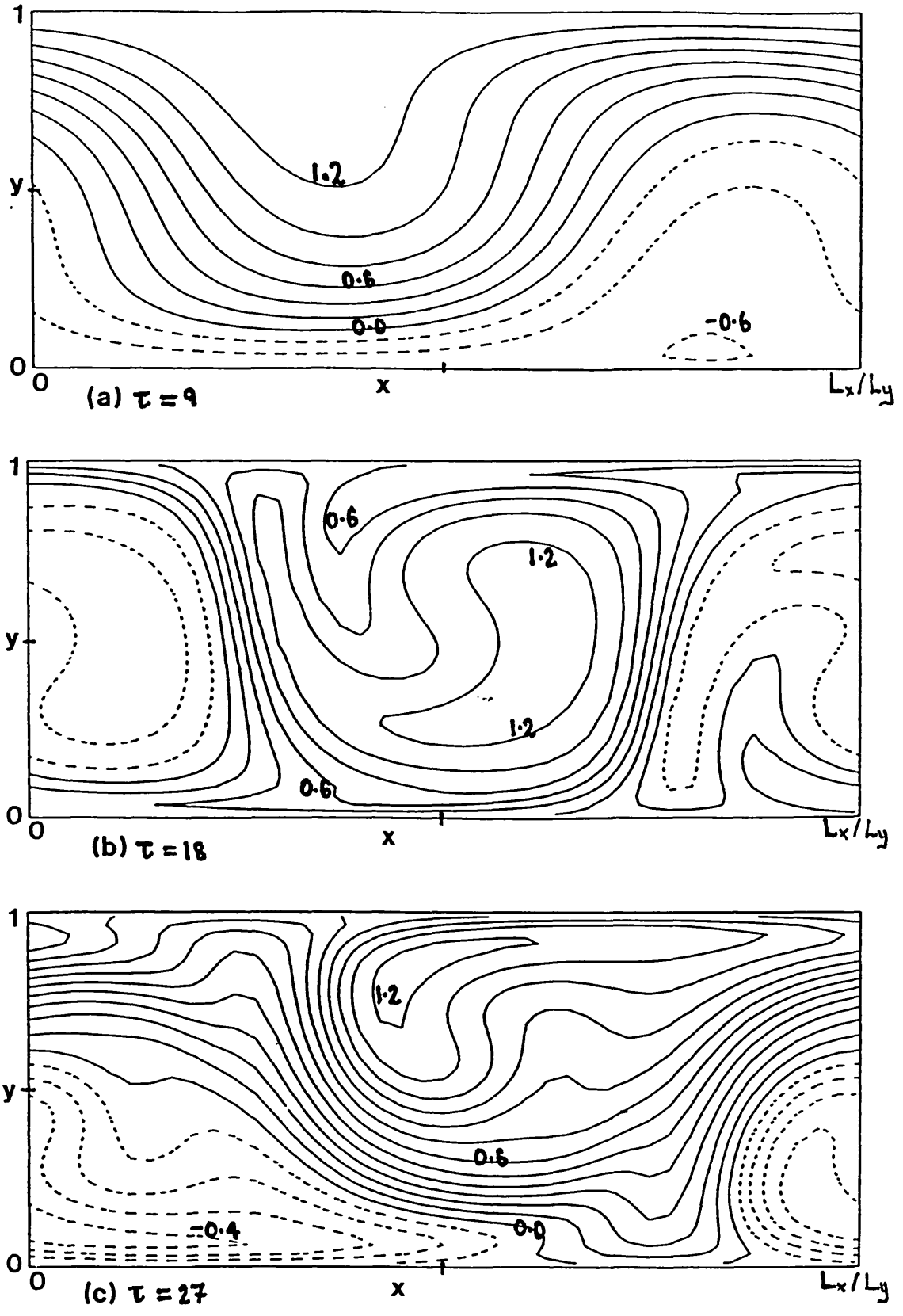


Figure 3.4

The evolution of the potential vorticity when subject to dissipation. The cross-sections of W are taken at ζ=0.1 at times (a) τ=9, (b) τ=18, (c) τ=27. The fields were obtained using the same resolution and initial conditions as run C but with a horizontal diffusion coefficient $\tilde{\nu}=2.10^{-4}$ (see (3.4.12)).

severely retarded when $\tilde{\nu} \geq 10^{-3}$; the inviscid normal mode analysis is inapplicable for such values of $\tilde{\nu}$. The results of figs. 3.2 & 3.4 by themselves, however, make it clear that the level of dissipation present in Hignett's experiments would strongly influence the critical layer's evolution and suggest that amplitude vacillations may become less regular as υ and κ are decreased.

CHAPTER FOUR

THE STABILITY OF ROSSBY WAVES AND OTHER FREE MODES IN A BOUNDED f-PLANE CHANNEL

Section 4.1 INTRODUCTION

The 3D spectral model introduced in section 2.4 is able to simulate some exact finite amplitude steady state solutions of the inviscid quasi-geostrophic equations on an f-plane. The simplest solutions have streamfunctions which contain only one eigenfunction of the potential vorticity operator. These Rossby waves are accurately represented in the spectral model by amplitude vectors with a single non-zero component. They are solutions irrespective of the Burger number or the aspect ratio of channel width to length. Combinations of Rossby waves with the same total wavenumbers (2.4.5) are also steady state solutions; the ratios of the amplitudes of the components of these solutions are arbitrary, but for fixed boundary conditions and channel aspect ratios they only occur at fixed (isolated) Burger numbers.

Streamfunctions of the form

$$\Psi = A_1 \sinh \alpha\pi(y-1/2) \cos\pi z + A_2 \exp ir k\pi x \sin\pi y \qquad (4.1.1)$$

are inviscid steady state solutions provided

$$1/B = \alpha^2 + r^2k^2 + 1 \ . \qquad (4.1.2)$$

White (1986) suggests that such solutions may give significant insight into the large amplitude steady waves which feature in rotating annulus experiments.

Combinations of Rossby waves with total wavenumber λ_0, say, are

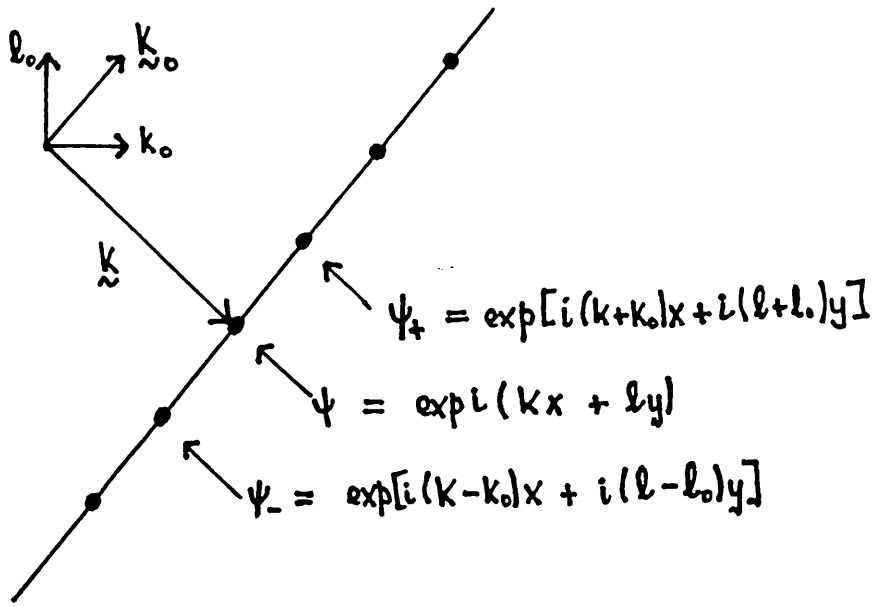


Figure 4.1

The "line" of perturbation modes involved in the linear stability of a plane wave $\psi_0=\exp i(k_0x+l_0y)$. The modes are represented by points on the horizontal wavenumber plane (k,l). Mode ψ_0 has wavenumber $\underset{\sim}{k}_0$. Its linear stability to a perturbation based on a mode $\psi=\exp i(kx+ly)$ with wavenumber $\underset{\sim}{k}$ is at issue. The advection of the potential vorticity of ψ_0 by ψ and that of ψ by ψ_0 only induces modes ψ_+ and ψ_-. The interaction of these and subsequent modes with ψ_0 induces only the modes lying on the line of wavenumbers illustrated.

also constant amplitude uniformly drifting solutions of the q-g equations on a β-plane. Their drift rates follow from Rossby's celebrated formula;

$$c_r = -\gamma / \lambda_o \ . \qquad (4.1.3)$$

Uniformly drifting (including steady state) waves or wave combinations which are solutions irrespective of their amplitudes will be referred to as free modes.

The stability of Rossby waves is of considerable interest. Lorenz (1972) investigated the stability of plane barotropic waves

$$\psi_o \equiv \exp i(K_o x + l_o y) \equiv \exp i\theta_o \qquad (4.1.4)$$

on an infinite β-plane and found that, for the types of normal mode disturbance he permitted, a wave of velocity amplitude U and horizontal wavenumber K was stabilised when the ratio $M \equiv (U\ K^2)/\beta$ lay below a certain value. Gill (1974) extended Lorenz's work, considering the limits of small and large M particularly carefully. He showed that at small M only wave perturbations which satisfy a resonance condition (defined below) are unstable and that coarsely truncated representations of the stability are reliable. At large M a wider range of waves is unstable at the coarsest truncation but the truncation is less reliable; it is not valid asymptotically (for $1/M \to 0$).

The Jacobian advection of a wave

$$\psi \equiv \exp i(kx + ly) \equiv \exp i\theta \qquad (4.1.5)$$

by wave ψ_0 of (4) induces waves with streamfunctions $\exp i(\theta \pm \theta_0)$, so the linear stability of wave ψ_0 to a perturbation based on ψ involves only a "line" of waves (see fig. 4.1) with streamfunctions $\exp i(\theta + n\theta_0)$. A wave perturbation is said to satisfy the (second order) resonance condition if the linear drift rates, according to (3), of ψ and $\psi_+ \equiv \exp i(\theta + \theta_0)$ or $\psi_- \equiv \exp i(\theta - \theta_0)$ are the same as those of the main wave ψ_0. When $M \ll 1$ wave perturbations which

do not satisfy the resonance condition will tend to drift relative to each other; unable to maintain a constant shape they cannot be normal mode perturbations. As the main wave amplitude and hence M is increased it may advect the perturbation so as to combat the Rossby drift; the unstable normal mode solutions found by Gill at small M which almost satisfy the resonance condition are of this type.

The stability of Rossby waves in a finite domain differs from that in an infinite domain in at least two important respects. Firstly the disturbances to the main wave must satisfy the imposed boundary conditions; these greatly restrict permissible disturbances. Secondly the main wave, whose stability is in question, cannot be represented by a single plane wave; it is rather a combination of two plane waves

$$\begin{aligned} \psi &= \exp ir K\pi x \sin \pi y \\ &= i/2 \exp i(rK\pi x - \pi y) - i/2 \exp i(rK\pi x + \pi y). \end{aligned} \tag{4.1.6}$$

As a result the perturbations are not confined to a line of eigenmodes as in fig. 4.1 but explore a net of eigenmodes such as that of fig. 4.2 (p158).

Hoskins (1973) and Baines (1976) have investigated the stability of barotropic Rossby waves on a sphere

$$\psi_n^l = P_n^l(\phi) \exp il\theta \; ; \; q_n^l = \nabla_H^2 \psi_n^l = -n(n+1)\psi_n^l \tag{4.1.7}$$

using truncated spectral representations of the barotropic vorticity equation. Hoskins considered the waves' stability to perturbations including only zonal flows and azimuthal harmonics (i.e. azimuthal wavenumbers 0, 1, 21,...). Baines also considered the stability to waves including other azimuthal wavenumbers (such as sideband instabilities). Baines noted that wave modes ψ_1^1, ψ_1^0, ψ_1^{-1} contain the three components of the fluid's angular momentum. Their amplitudes are hence constants of the inviscid motion. The five wave modes with n=2, which cannot interchange energy amongst themselves, must be stable to small amplitude disturbances by

Fjørtoft's anti-cascade property because they cannot transfer energy to modes (n=0 or n=1) with lower total wavenumbers. Baines found all waves with $n > 2$ to be unstable when their amplitudes were sufficiently large. Baines and Hoskins found that certain very coarse truncations could give moderately accurate representations of the amplitudes of marginal stability and Baines reported that convergence of the solutions was achieved using 30 perturbation modes.

Hoskins (1973 appendix) and Plumb (1977) have considered the stability of barotropic Rossby waves (6) (of gravest meridional structure) in a Cartesian channel of width L and infinite length on a β-plane. Hoskins showed that a resonant triad instability involving a zonal perturbation and harmonics can only occur when $rk > 2$. The truncation used by Hoskins is only valid when M is small. Plumb, studying the case $M \ll 1$, found that there were no (second order) resonant triad instabilities on waves with $rk < 0.681$. These waves, however, were shown to be unstable to (weaker) side-band instabilities. Plumb (1976) has also discussed the stability of small amplitude waves in finite periodic channels for which there are, in general, no precisely resonant triad instabilities.

The stability of Rossby waves of the form (6) in a bounded zonally periodic Cartesian channel on an f-plane appears to have escaped attention. This may be in part because the concept of triad resonances breaks down on an f-plane; all Rossby waves have zero phase velocity and hence all triads are resonant so the resonance criterion does not pick out a small subset of interacting modes. Perhaps investigators have also been deterred by numerical problems in achieving adequate resolution of the instabilities or considered the problem to be of minor importance compared to that on the sphere.

Numerical results presented in section 4.3 suggest that waves (6) with $k \leq 4$ are linearly stable to normal mode disturbances providing that $\psi_1 = \exp ir\pi x \sin 2\pi y$ has a larger total wavenumber than (6); i.e. that

$$r^2(k^2-1) \leq 3 . \qquad (4.1.8)$$

Hence there are stable waves of more than one azimuthal wavenumber in long narrow channels and satisfaction of the Fjørtoft anti-cascade condition is not sufficient for a wave to be unstable to normal modes. The results also show that extremely fine spectral truncations are required to resolve normal mode perturbations to the waves and that in coarse truncations the stability depends substantially on whether the truncation is "even" or "odd". The formulation and tests of the code used to obtain these results are described briefly in section 4.2. The code implements the 3D spectral model introduced in section 2.4. Different versions of the code can be used to perform non-linear time integrations, linearised time integrations or matrix eigenvalue problems.

Section 4.3 also reports results on the stability of two free modes each of which contains an axisymmetric mode and a wave mode. The first free mode is barotropic; the second is baroclinic and hence of particular interest for annulus flows. Approximate values of the amplitude ratios of the free mode components at marginal stability are established from the numerical results.

Contending interpretations and techniques for the proof of the Rossby waves' stability are surveyed in section 4.4. A qualitative analytical study of singularities imposed on a neutral mode at values of r bordering those defined by (8) is also presented because it is not entirely clear from the numerical results that (8) marks the stability transition. Some suggestions and criticisms concerning the interpretation of large amplitude steady waves in rotating annulus experiments in terms of free modes conclude the discussion.

Section 4.2 DESCRIPTION OF THE NUMERICAL SPECTRAL MODEL

The formulation, testing and documentation of a numerical model is an exacting and lengthy task; more detailed descriptions of the model can be obtained from the author on request.

There are three versions of the model. The one developed first performs time integrations of the full (non-linear) equations. It allows investigation of almost any truncation but is only respectably efficient for coarse truncations. It may be used to study the evolution of waves subject to dissipation on axisymmetric flows tending to relax to non-zero amplitudes. The second version performs time integrations for linear perturbations to a single large amplitude wave mode and/or a single axisymmetric mode. The third version finds the normal modes of such linearised problems by formulating a matrix eigenvalue problem.

All versions of the model use the same subroutines to specify the modes included in the calculation, to determine the non-zero overlap integrals and to calculate their values. Descriptions of these shared elements of the model are followed by notes on the time integration and the matrix formulation and solution. Summaries of the methodical tests of the model conclude the section.

Non-dimensionalisation and mode definition

The non-dimensionalisation used in the model is the same as that used in section 2.1 and defined by (2.1.12), except that in place of $x^* = L\,x$, L being the channel gap width, we use

$$x^* = L_x/2 \,.\, x \,, \qquad (4.2.1)$$

L_x being the channel's periodic repeat length; the domain extends from $x = 0$ to $x = 2$ as a result. The model assumes that N^2 is independent of z so the wave modes are given by simple trigonometric functions

$$\psi_i = \begin{matrix}\cos\\ \sin\end{matrix} k_i \pi x \quad \sin l_i \pi y \quad \cos m_i \pi z \,. \tag{4.2.2}$$

Their total wavenumbers are

$$\lambda_i = \left(r^2 K_i^2 + l_i^2 + m_i^2/B \right) \pi \tag{4.2.3}$$

in which

$$r = 2L/L_x \qquad \text{and} \qquad B = N^2H^2/(f^2L^2) \,. \tag{4.2.4}$$

r will be referred to as the gap width ratio. It is small for long narrow channels. The axisymmetric modes are of the form

$$\psi_i = \cos m_i \pi z \; f_i(y) \,. \tag{4.2.5}$$

The functions $f_i(y)$ are, of course, eigenfunctions of d^2/dy^2;

$$d^2 f_i / dy^2 = \eta f_i \,.$$

The gravest cross stream mode is allowed to take any of the following forms;

$$\begin{aligned} f_1 &= \sin \pi y && \text{choice A} \\ f_1 &= \sinh \alpha\pi(y-1/2) \,, \ \alpha>0 \ ; \ f_1 = y-1/2 && \text{choice B} \\ f_1 &= \sin \alpha\pi(y-1/2) \,, \ 0<\alpha<1 && \text{choice B} \end{aligned} \tag{4.2.6}$$

Choice A is used to investigate the stability of barotropic Rossby waves. Choice B produces axisymmetric flows ($\bar{u}=-\partial\psi/\partial y$) which are symmetric about mid-channel. Choice of one of these forms and a value of α implies a choice of the Sturm - Liouville b.c. coefficients

$$a f_i + b \, df_i/dy = 0 \qquad \text{at } y=0,1 \,. \tag{4.2.7}$$

This ratio is taken to be the same for all axisymmetric modes; no provision is made to allow different values for different vertical wavenumbers. The structure and lateral wavenumber of the higher lateral wavenumbers is hence determined. The modes are of the form

$$f_n = \sin n\pi y \qquad \text{choice A}$$

$$f_n = \cos \alpha_n \pi (y - 1/2) \quad \text{and} \quad f_n = \sin \beta_n \pi (y - 1/2) \qquad \text{choice B} \tag{4.2.8}$$

Basic model equation and truncations

The present model formulation is for an f plane only. It can be used to investigate truncations consisting of arbitrary finite combinations of wave and axisymmetric modes. Extension of the model to the β plane would require a mode $\sin k_i \pi x \cos m_i \pi z \sin l_i \pi y$ to be included in a truncation if and only if $\cos k_i \pi x \cos m_i \pi z \sin l_i \pi y$ were included. A straightforward adaptation of the model for use on a β plane has been devised but not implemented. For a truncation with N modes,

$$\Psi = \sum_{i=1}^{N} a_i \psi_i , \tag{4.2.9}$$

the model equations are (cf (2.4.11))

$$\lambda_i \langle \psi_i \psi_i \rangle \, da_i/dt = \sum_{j>k} r(\lambda_j - \lambda_k) a_j a_k \langle \psi_i J(\psi_j, \psi_k) \rangle - \mu_i (a_i - \bar{a}_i) \langle \psi_i \psi_i \rangle . \tag{4.2.10}$$

The second term on the r.h.s. of (10) represents diffusive dissipation of waves and forcing of the axisymmetric flow; $\bar{a}_i$ is zero for wave modes but may be chosen to be non-zero for each axisymmetric mode. There is some freedom of choice in the dependence of μ_i on λ_i, k_i, l_i and m_i and the constants of proportionality, μ_i, for axisymmetric modes and wave modes are chosen separately.

Calculation of overlap coefficients

The overlaps $\langle \psi_i J(\psi_j, \psi_k) \rangle$ are calculated using selection rules and formulae for the non-zero integrals. The interactions between one axisymmetric mode and two wave modes are calculated by one subroutine and interactions between three wave modes by a

second. Each overlap involving an axisymmetric mode ψ_0 and wave modes ψ_1 and ψ_2 may be deduced from

$$\langle \psi_1, J(\psi_2, \psi_0) \rangle = \iiint \psi_1 \, \partial\psi_2/\partial x \, \partial\psi_0/\partial y \, dx\,dy\,dz \qquad (4.2.11)$$

using the triple product relation. This integral splits naturally into the product of three integrals over x, y and z respectively. The vertical integral

$$I_z = \int_0^1 \cos m_1\pi z \, \cos m_2\pi z \, \cos m_0\pi z \, dz \qquad (4.2.12)$$

$$= 0 \quad \text{unless} \quad m_1 = m_2 + m_0 \text{ or } m_2 = m_0 + m_1 \text{ or } m_0 = m_1 + m_2 .$$

Expressions for the integrals and combinations involved in the overlap sums are written down in the model documentation. The selection rules limiting interactions between modes, which are used in the construction of the truncations in the next section, are stated below. They are easily inferred from simple symmetries and trigonometric formulae.

i) wave - wave overlap $\langle \psi_1 \; J(\psi_2, \psi_3) \rangle$ is zero unless

$k_1 = k_2 + k_3$ or $k_2 = k_3 + k_1$ or $k_3 = k_1 + k_2$

& $l_1 = l_2 + l_3$ or $l_2 = l_3 + l_1$ or $l_3 = l_1 + l_2$

& $m_1 = m_2 + m_3$ or $m_2 = m_3 + m_1$ or $m_3 = m_1 + m_2$

& the overlap contains an even number of cos kx modes

ii) axisymmetric(mode 0) - wave (modes 1 and 2) interactions are zero unless

one wave mode $\propto \cos k_1 x$ and the other wave $\propto \sin k_1 x$

& $m_1 = m_2 + m_3$ or $m_2 = m_3 + m_1$ or $m_3 = m_1 + m_2$

& for choice A axisymm. modes

$l_1 = l_2 + l_3$ or $l_2 = l_3 + l_1$ or $l_3 = l_1 + l_2$

or for choice B axisymm. modes

$l_1 + l_2$ is even(odd) if the axisymmetric mode is even(odd) about mid-channel

The self overlaps $\langle \psi_i, \psi_i \rangle$ are determined similarly (and more easily).

Time integration

Initial values of the mode amplitudes are specified as the modes are chosen for inclusion. The non-zero overlap integrals are, of course, selected and evaluated once and for all before integration begins. The rate of change of each amplitude at each time step is found by summation of overlaps multiplied by mode amplitudes and integration achieved by a simple leapfrog scheme. A Robert filter can be used to enhance numerical stability. The integration is started by Miyakoda's method of a short forward step followed by leapfrog steps up to the chosen time step. For numerical stability it seems to be sufficient to satisfy $K\ U\ \Delta t \ll 1$, K being the largest azimuthal wavenumber, U the maximum flow velocity and Δt the time step length.

With N_w and N_A denoting the number of wave modes and axisymmetric modes respectively, the number of wave-wave overlaps is $\leq 8\ N_w^2$, whilst the number of wave-axisymmetric overlaps is $\leq N_w.N_A$. Time stepping by summation of wave-wave interactions becomes uneconomical in non-linear integrations when more than about 160 wave modes are present. Wave-wave overlaps at finer resolutions would be more efficiently calculated by transforming the wave fields of ψ and $q = \nabla^2\psi$ onto a uniform spatial grid, calculating $J(\psi, q)$ on the grid and transforming back to spectral space. The wave-axisymmetric interactions could continue to be calculated using the overlap summations since $N_A \ll N_w$ at fine truncations (when K is appropriately large). Time integrations linearised about a Rossby wave or free mode are as efficient as they would be with a mixed spectral gridpoint model. Code to allow more than one large amplitude wave mode and/or axisymmetric mode has been devised but not implemented.

Eigenvalue analysis

On setting $d/dt(\lambda_i\ a_i) = \Gamma\ (\lambda_i\ a_i)$, (10) reduces to an eigenvalue problem which may be written in the form

$$\left(\underline{\underline{A}} - \Gamma \underline{\underline{I}} \right) \underline{b} = \underline{0} \ , \qquad (4.2.13)$$

$\underline{\underline{A}}$ being an N x N unsymmetric square matrix with real coefficients and $\underset{\sim}{b}$ the vector with elements $b_i = \lambda_i a_i$. When $N >> L$, L being the number of large amplitude modes, $\underline{\underline{A}}$ is a sparse matrix. No advantage of this fact is taken by the NAG routines which are used to find the eigenvalues and eigenvectors. (The eigenvalues are determined by reducing the matrix $\underline{\underline{A}}$ first to upper Hessenberg form by elementary Householder similarity transformations and then to triangular form by the Francis QR algorithm.) The code assumes that the fluid is inviscid, a restriction which could be alleviated relatively easily.

Tests of the models

After the first version of the model was built a systematic test of the selection rules and evaluations of the overlap summations was undertaken. The success of this test gives considerable confidence that the code calculates these overlaps according to the formulae established in the documentation. Checks of the total rate of change of "energy" (2.4.13) and enstrophy (2.4.14) show the double precision version of the model to conserve these quantities to very high accuracy (consistent with the precision of the calculations) in runs which include all axisymmetric - wave overlaps of type B and all types of wave-wave overlaps. This check is a valuable cross-check on the formulae established in the model documentation. After type A axisymmetric modes were introduced and the second and third versions of the model completed, intercomparisons were made. The non-linear model continued to conserve "energy" and enstrophy and for linear perturbations (of amplitude 10^{-10}) to a two component free mode, containing a wave and axisymmetric components, produced results in acceptable agreement with the linearised time stepping model. The most unstable eigenvalue of the matrix analysis agreed to within 0.1% with that inferred from the linear time stepping integration after 40 non-dimensional time units.

Section 4.3 THE STABILITY OF BAROTROPIC ROSSBY WAVES AND OTHER FREE MODES: NUMERICAL RESULTS

This section commences with results, obtained with the numerical model described in the previous section, concerning the stability of barotropic Rossby waves,

$$\Psi_0 = \frac{1}{\pi} \cos K\pi x \sin \pi y , \tag{4.3.1}$$

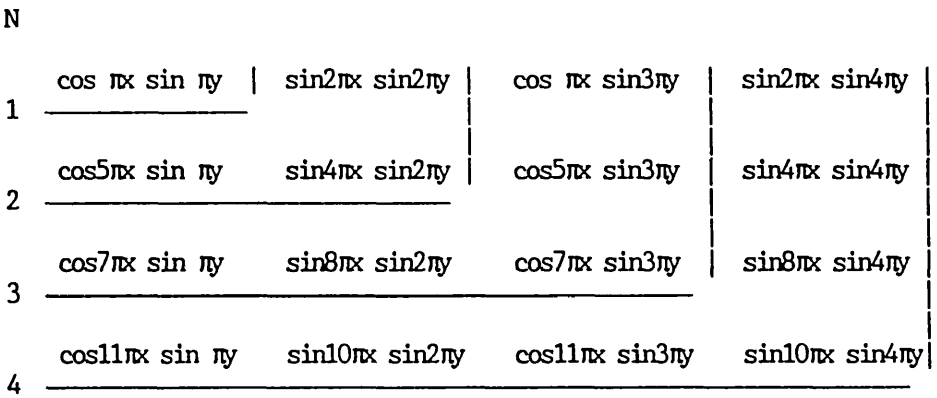
in periodic f-plane Cartesian channels against normal mode disturbances. Only waves Ψ_0 with k = 2,3 and 4 are discussed. These are amongst the modes (which satisfy the boundary conditions) with the smallest total wavenumbers . According to the anti-cascade argument unstable perturbations must include at least one mode of smaller total wavenumber than the main wave; Ψ_0 with k=2, for example, can only be unstable to perturbations which include $\psi = \exp i\pi x \sin \pi y$ or $\psi = \sin \pi y$. We shall say that a perturbation is <u>based</u> on the gravest mode contained in it. Results are presented first for perturbations based on wave modes and then for perturbations based on the axisymmetric mode $\psi = \sin \pi y$. Results for a barotropic free mode and then a baroclinic free mode follow. Suggestions for future work conclude the section.

Definition (4.2.1) for x will be used in this and the following section so streamfunctions will have cos kπx (rather than cos rkπx) dependence. Note also that the velocity field of the Rossby wave (1) is

$$(u_g , v_g) = (- \cos K\pi x \cos \pi y , -rK \sin K\pi x \sin \pi y). \tag{4.3.2}$$

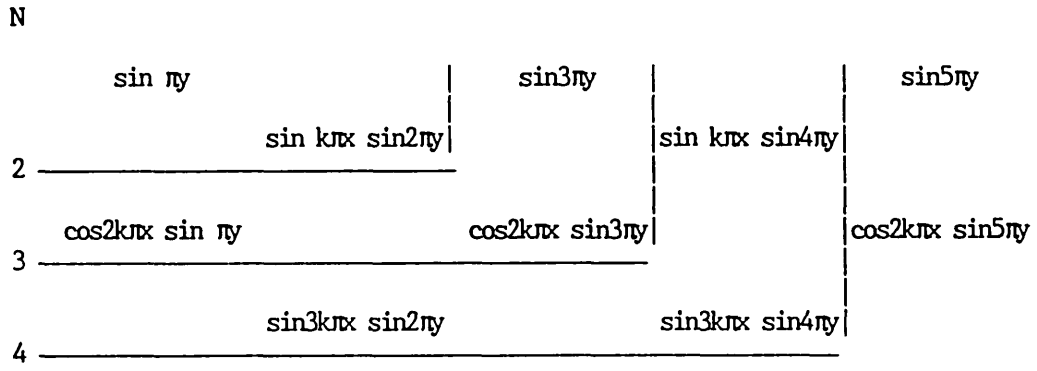
So if rk < 1 then the maximum velocity of the Rossby wave is 1, which is a convenient value for the assessment of growth rates, but if rk > 1 then the maximum value of $|v_g| = rk > 1$.

Figure 4.2



The modes included in a perturbation to $\Psi_0 = \cos 3\pi x \sin \pi y$ based on $\psi = \cos \pi x \sin \pi y$.

Figure 4.3



The modes included in a perturbation to $\Psi_0 = \cos k\pi x \sin \pi y$ based on $\psi = \sin \pi y$.

Perturbations to Rossby waves based on wave modes

The interaction of a small amplitude mode, $\psi = \cos m\pi x \sin n\pi y$, with $\Psi = \cos k\pi x \sin n\pi y$ induces small amplitude modes $\sin(k\pm m)\pi x \sin 2\pi y$. The interaction of these with $\cos k\pi x \sin n\pi y$ induces further modes. The net of modes induced is illustrated by fig. 4.2, which applies to a perturbation to $\cos 3\pi x \sin n\pi y$ based on $\cos\pi x \sin n\pi y$. The net extends up to arbitrarily large wavenumbers and must be truncated for the purposes of numerical calculations. The "square" truncation which has been used includes at level N all modes with lateral wavenumber $\leq N$ and azimuthal wavenumber $< kN$ (e.g. $< 3N$ in fig. 4.2). The truncation at level N includes N^2 perturbation modes.

It is not immediately clear that perturbations to $\cos k\pi x \sin n\pi y$ based on $\sin m\pi x \sin n\pi y$ will yield the same eigenvalues as those based on $\cos m\pi x \sin n\pi y$. Comparison of the overlap sums for the two cases, however, reveals a symmetry between them and they do have the same eigenvalues. Pairs and quartets of normal mode solutions of the stability problem under consideration, namely,

$$\Gamma \nabla^2\psi + J(\Psi_0, (\nabla^2 + \lambda_0^2)\psi) = 0$$
$$\psi = 0 \quad \text{on } y = 0,1 \quad ; \quad \psi \text{ periodic in } x, \ 0 \leq x \leq 2, \tag{4.3.3}$$

in which Γ is the eigenvalue and Γ_r the growth rate, are also related. If $(\psi(x,y-\tfrac{1}{2}), \Gamma)$ is a solution of (3) then so are (ψ^*, Γ^*) and $(\psi(-x,-y+\tfrac{1}{2}), -\Gamma)$. So stability transitions occur either by exchange of instabilities (Chandrasekhar 1961) with a pair of conjugate imaginary eigenvalues passing through zero and yielding two real eigenvalues of opposite sign; or by a pair of unstable solutions with conjugate eigenvalues appearing with a similar pair of decaying solutions.

Table 4.1 displays the most unstable eigenvalues of normal mode perturbations to $\Psi_0 = 1/\pi \cos 2\pi x \sin \pi y$ based on $\psi = \cos\pi x \sin n\pi y$ for two values of the aspect ratio, r, (4.2.4). At both these aspect ratios all truncations with odd values of N have no unstable modes and the instability found with even truncations decreases rapidly in

Table 4.1

r = 1/3		r = 1	
Truncation level (N)	Growth rate	Truncation level (N)	Growth rate
2	$1.51.10^{-1}$	2	$7.55.10^{-1}$
3	S	3	S
4	$1.85.10^{-2}$	4	$9.80.10^{-2}$
5	S	5	S
8	$6.39.10^{-4}$	8	$4.47.10^{-3}$
9	S	9	S
		12	$1.75.10^{-4}$
		13	S

Truncation dependence of the growth rate of the most unstable perturbation based on $\psi = \cos\pi x \sin\pi y$ to $\Psi_0 = 1/\pi \cos 2\pi x \sin\pi y$ for r=1/3 and r=1. "S" indicates that all normal modes are (neutrally) stable.

Table 4.2

Azim. wavenos. k	m	r	Γ_r	Γ_i
2	1	1.000	0.0	0.0
2	1	1.005	0.116	0.0
3	1	0.77	0.0	0.077
3	1	0.78	0.065	0.0
3	2	0.610	0.0	0.087
3	2	0.615	0.091	0.0
4	1	0.585	0.0	0.003
4	1	0.59	0.195	0.0
4	3	0.44	0.0	0.153
4	3	0.45	0.097	0.0

Determination of the aspect ratio (r) of marginal stability for wave perturbations based on $\psi = \cos m\pi x \sin\pi y$ to a large amplitude wave $\Psi_0 = 1/\pi \cos k\pi x \sin\pi y$. The truncation level is N=13 in all cases; Γ_r is the perturbation's growth rate.

strength as N increases. We conclude that the main wave is stable to this wave perturbation at both aspect ratios. The significant difference between coarse even and odd truncations is a characteristic of all results presented here. It is somewhat reminiscent of the results for truncated spectral representations of the stability of the $u = -\frac{1}{2}\cos\pi z$ internal jet flow (section 2.4). There stability transitions were reproduced exactly by perturbations with an even number of modes and with gross error by coarse "odd" truncations.

Stability transitions, according to truncations with N=13, for Ψ_0 with k=2, 3 and 4 are tabulated in table 4.2. The wavenumber 2 main wave is evidently marginally stable at $r = 1.0$. At this aspect ratio, $\sin\pi x \sin 2\pi y$, a prominent member in the net of perturbation modes, has the same total wavenumber as the main wave. Some of the interactions with other modes and the corresponding terms in the stability matrix will change sign at this critical aspect ratio. At higher aspect ratios than the critical two modes will have smaller total wavenumber than the main wave so energy cascades will have more freedom. The wavenumber 3 main wave has a smaller critical aspect ratio for perturbations based on $\cos 2\pi x \sin\pi y$ than $\cos\pi x \sin\pi y$. The net of modes based on the former includes $\sin\pi x \sin 2\pi y$, which has a lower total wavenumber than its counterpart $\sin 2\pi x \sin 2\pi y$ in the $\cos\pi x \sin\pi y$ net. $\sin\pi x \sin 2\pi y$ has the same total wavenumber as the main wave when $(9 - 1)r^2 = 3$; i.e. $r = 0.6124$, whilst $\sin 2\pi x \sin 2\pi y$ has the same total wavenumber as the main wave when $(9 - 4)r^2 = 3$; i.e. $r = 0.7746$. Inspection of table 4.2 shows that the transitions for N=13 must be very close to these values if not identical with them. These results suggest that

> a net of perturbation modes for an odd truncation allows instability if and only if two modes in it have smaller total wavenumber than the main wave itself. (4.3.4)

This "hypothesis" also agrees with table 4.2 for k=4 and m=3 ($(16-1)r^2 = 3 \Rightarrow r = 0.4472$). The transition for k=4 and m=1, however, does not agree with (4). (4) predicts that the transition will occur when $(16-9)r^2 = 3$ i.e. $r=0.6547$, whilst the numerical

Table 4.3

r	N=3	N=5	N=13	N=12
1.000	S	S	S	$1.8.10^{-4}$
1.001	0.088	0.070		
1.005		0.156	0.116	0.060
1.01	0.279	0.220	0.163	0.081
.1	0.887	0.685	0.467	0.141
.2	1.205	0.891	0.478	0.262
3	1.299	0.863	S	0.307
4	1.036	0.047	S	S
5	S	S		

dence on the truncation level N and the aspect ratio r of the growth rate, Γ_r, e most unstable normal mode perturbation based on $\psi = \cos\pi x \sin\pi y$ to a large ude wave $\Psi_0 = 1/\pi \cos 2\pi x \sin\pi y$. "S" indicates that all modes are stable.

Table 4.4

r = 0.3		r = 0.6	
Truncation level (N)	Growth rate	Truncation level (N)	Growth rate
2	$1.81.10^{-1}$	2	$4.24.10^{-1}$
4	$2.87.10^{-2}$	4	$1.92.10^{-1}$
6	$1.09.10^{-2}$	6	$1.17.10^{-1}$
8	$1.54.10^{-3}$	8	$8.28.10^{-2}$
10	$3.03.10^{-4}$	10	$6.11.10^{-2}$
12	$8.65.10^{-5}$	12	$4.33.10^{-2}$

Truncation dependence of the growth rate of the most unstable perturbation based on $\psi = \cos 2\pi x \sin\pi y$ to $\Psi_0 = 1/\pi \cos 3\pi x \sin\pi y$ for r=0.3 and r=0.6.

results locate the transition within the range $0.585 \leq r \leq 0.59$. This transition is quite a strong one (much stronger for instance than the k=3, m=1 transition) and cannot be ignored. The reason why it does not conform to (4) is not clear. Certainly (4) cannot be correct as it stands but the present author has been unable to formulate appropriate qualifications to it to make it valid.

The results in table 4.3 show that the transition for the wavenumber 2 flow is located exactly at r = 1.0 by both N=3 and N=5 truncations, but that the growth rate in the unstable region near the transition is over-estimated by these coarse truncations (as is the case for the zonal jet in section 2.4). The transition to healthy instability is remarkably sharp even for the N=13 case. Indeed the eigenvalue appears to depend on the square root of the supercriticality of the aspect ratio (i.e. on $(r-1)^{\frac{1}{2}}$). Table 4.3 also reveals that the instability does not increase monotonically with r; the N=13 truncation shows a return to stability for $1.2 \leq r \leq 1.3$. Similar returns to stability for the N=3 and N=5 truncations occur for $1.4 \leq r \leq 1.5$ and for the N=12 truncation for $1.3 \leq r \leq 1.4$. Unfortunately, however, the aspect ratio at which stability returns and also the maximum growth rate for the continuous (untruncated) problem cannot be inferred from table 4.3 with tolerable reliability. Higher values of r have not been investigated thoroughly but it appears that instability recurs at higher values of r.

"Odd" truncations with N=3 and N=5 have also been found to locate the stability transition according to (4) for perturbations to $1/\pi \cos 3\pi x \sin \pi y$ based on $\cos 2\pi x \sin \pi y$. An indication of the convergence of even truncations of this perturbation is presented in table 4.4. As in table 4.1 one column of table 4.4 has r well below the critical value (r=0.6124) according to (4) and the other has r only slightly sub-critical. The convergence of the first column to zero is rapid but that of the second is at best very slow and much worse than that of table 4.1.

Table 4.5

r = 1/3		r = 1	
Truncation level (N)	Stability	Truncation level (N)	Stability
3	S	3	S
4	$5.16.10^{-2}$	4	1.56
5	S	5	S
6	$5.04.10^{-3}$	6	1.05
9	S	9	S
10	$1.82.10^{-4}$	10	0.532
15	S	15	S
16	$5.36.10^{-7}$	16	0.224

Truncation dependence of the growth rate of the most unstable perturbation based on $\psi = \sin\pi y$ to $\Psi_0 = 1/\pi \cos 2\pi x \sin\pi y$ for r=1/3 and r=1. "S" indicates that all normal modes are (neutrally) stable.

Table 4.6

k	r	N=5	N=7	N=13
2	1.14			(0,0.373)
2	1.145			(0.052,0)
2	1.160		(0,0.111)	
2	1.165		(0.384,0)	
2	1.170	(0,0.330)		
2	1.175	(0.250,0)		

The eigenvalues (Γ_r, Γ_i) of the most unstable normal mode perturbation to $\Psi_0 = 1/\pi \cos k\pi x \sin\pi y$ based on $\psi = \sin\pi y$ as a function of aspect ratio r and truncation level N.

Perturbations to Rossby waves based on axisymmetric modes

These instabilities are based on the $\sin\pi y$ mode which on interaction with the $\cos k\pi x \sin\pi y$ mode induces a $\sin k\pi x \sin 2\pi y$ mode. The net of modes excited by further interactions is illustrated by fig. 4.3 (p158). A "square" truncation of level N is defined to include all of these modes with lateral wavenumbers $\leq N$ and azimuthal wavenumbers $\leq k(N-1)$. Note that all perturbation modes in fig. 4.3 have azimuthal wavenumbers directly proportional to k; the perturbation itself has azimuthal wavenumber k. So waves $\cos k\pi x \sin\pi y$ at aspect ratio r_0 have the same stability properties as $\cos\pi x \sin\pi y$ at aspect ratio kr_0.

Table 4.5 displays the dependence of the stability of $\Psi_0 = 1/\pi \cos 2\pi x \sin\pi y$ on the level of truncation for aspect ratios r=1/3 and r=1. All "odd" truncations are stable and the instability in even truncations with r=1/3 decreases in magnitude rapidly as the truncation level increases. At r=1 the convergence to stability for even truncations is extremely slow (cf tables 4.1 and 4.4). One expects convergence to occur so that very fine even and odd truncations agree closely.

The aspect ratio of the stability transition for odd truncations depends on the truncation level as may be seen from table 4.6; this behaviour is in contrast to the exact location of the transition by coarse odd truncations of wave perturbations. It seems likely that the perturbations to $\cos 2\pi x \sin\pi y$ become unstable for $r > 1.1$. If this is so, then the wave based perturbations involving $\sin\pi x \sin 2\pi y$ to large amplitude waves $\cos k\pi x \sin\pi y$ with k=2, 3 or 4 break out at smaller aspect ratios than the axisymmetric based ones; this would be qualitatively similar to the findings of Baines (1976) concerning the stability of Rossby waves on the sphere. $\psi = \cos\pi x \sin\pi y$ is stable to all wave based perturbations but unstable to axisymmetric based ones when $r \geq 2.2$.

Table 4.7a

N	a_0 stable	unstable
2	unstable	
3	0.39	0.40
4	unstable	
5	0.26	0.27
6	unstable	
8	0.28	0.29
9	0.29	0.30

Table 4.7b

N	a_0 stable	unstable
6	0.080	0.079
8	0.014	0.013
10	0.010	0.009

The dependence on the truncation level N of the lower and upper bounds on the amplitude ratio a_0 at which the barotropic free mode (4.3.5) becomes unstable to perturbations based on a) $\psi = \begin{matrix}\cos\\ \sin\end{matrix} \pi x \sin\pi y$ and b) $\psi = \sin\pi y$.

Perturbations to a barotropic free mode

Barotropic streamfunctions of the form

$$\Psi_o = \frac{a_o}{\pi} \sin 2\pi y + \frac{1-a_o}{\pi} \cos 2\pi x \sin \pi y \qquad (4.3.5)$$

have been investigated with $r=\sqrt{3}/2$. For this value of r, Ψ_0 is a free mode. The net of modes used for the wave perturbation based on $\begin{matrix}\cos\\ \sin\end{matrix}$ $\pi x \sin \pi y$ with truncation level N includes all wave modes with odd azimuthal wavenumber $\leq$ 2N-1 and lateral wavenumber $\leq$ N. The N level truncation hence involves 2 N^2 modes. The dependence of the stability of the free mode on the amplitude ratio a_0 and the truncation level N is indicated in table 4.7a. With odd truncations the free mode is stable provided a_0 is smaller than a critical value which is bounded by the values in the table. At coarse "even" truncations the free mode was unstable at all values of a_0 investigated (a_0 = 0, 0.25, 0.5, 0.75, 1.), but the finest even and odd truncations investigated, namely N=8 and N=9, show good agreement in the location of the transition. It appears that the free mode is stable to these wave perturbations if $a_0 \leq 0.28$. In qualitative terms, the presence of the zonal flow could be said to destabilize the wave or the wave to stabilize the zonal wave.

The net of modes used for the perturbation based on $\psi = \sin \pi y$ at level N includes all modes

$$\begin{aligned} \psi &= \begin{matrix}\cos\\ \sin\end{matrix} \pi \,(\text{even})\, x \;\cdot\; \sin\,(\text{odd})\,\pi y \\ \psi &= \begin{matrix}\cos\\ \sin\end{matrix} \pi \,(\text{odd})\, x \;\cdot\; \sin\,(\text{even})\,\pi y \end{aligned} \qquad (4.3.6)$$

with azimuthal wavenumber $\leq$ 2(N-1) and lateral wavenumber $\leq$ N. From the results of table 4.5 (for the stability of $\Psi_0 = 1/\pi \cos 2\pi x \sin \pi y$ to perturbations based on $\sin \pi y$) odd truncations are stable when a_0 = 0 and even truncations unstable. The N=9 truncation was found to be stable for a_0 = 0, 0.25, 0.5, 0.75 and 1. Table 4.7b shows that the "spurious" instability of coarse even truncations of the zonal flow streamfunction $\sin \pi y$ is stabilised by relatively small additions of the streamfunction $\sin 2\pi y$. So the free mode appears to be stable to streamfunction perturbations based on $\sin \pi y$ at all

TABLE 4.8a

MMAX=2 a_0 / N	0.1	0.15	0.2	0.25	0.3	0.5
4	$6.51.10^{-4}$	$3.53.10^{-4}$	$3.22.10^{-2}$	$7.43.10^{-2}$	0.108	0.234
5	$8.22.10^{-5}$	$1.07.10^{-4}$	$1.16.10^{-3}$	$5.06.10^{-2}$	$7.32.10^{-2}$	0.235
8	$5.54.10^{-6}$	$5.27.10^{-5}$	$1.17.10^{-4}$	$1.73.10^{-2}$	$9.51.10^{-2}$	0.239
9	S	$4.71.10^{-5}$	S	$1.87.10^{-2}$	$8.86.10^{-2}$	0.240

TABLE 4.8b

N=5 a_0 / MMAX	0.1	0.15	0.2	0.25	0.3	0.5
2	$8.22.10^{-5}$	$1.07.10^{-4}$	$1.16.10^{-3}$	$5.06.10^{-2}$	$7.32.10^{-2}$	0.235
3		$3.82.10^{-5}$	$2.23.10^{-4}$	$7.94.10^{-4}$	S	$4.34.10^{-2}$
4	$7.38.10^{-5}$	$1.44.10^{-4}$	$7.07.10^{-3}$	$3.41.10^{-2}$	$6.44.10^{-2}$	0.171
5		$6.29.10^{-5}$	$1.15.10^{-4}$	$6.67.10^{-4}$	$8.31.10^{-3}$	0.147
6	$7.38.10^{-5}$	$1.51.10^{-4}$	$5.48.10^{-3}$	$2.97.10^{-2}$	$5.31.10^{-2}$	0.160

The maximum growth rates of perturbations based on $\psi=\cos\pi x \sin\pi y$ to the baroclinic free mode (4.3.7) - (4.3.9). (a) displays the dependence on the amplitude ratio a_0 and the horizontal truncation level N for a vertical truncation MMAX=2. (b) displays the dependence on a_0 and MMAX for N=5.

values of a_0.

In conclusion, the barotropic free mode is probably stable to all normal modes when $a_0 \leq 0.28$.

Perturbations to a baroclinic free mode based on wave modes

The streamfunction

$$\Psi_0 = \frac{a_0}{\pi} \cos\pi z \sinh\alpha\pi(y-1/2) + \frac{1-a_0}{\pi} \cos 2\pi x \sin\pi y \qquad (4.3.7)$$

with an aspect ratio, r, and a zonal meridional flow profile reasonably representative of steady waves in differentially heated annuli

$$r = 1/3 \quad ; \quad \alpha = 1 \qquad (4.3.8)$$

is a free mode when

$$1/B = (fL/NH)^2 = 2 + 4r^2 = 2.\dot{4} \; . \qquad (4.3.9)$$

The linear stability of this flow has been investigated using a perturbation net based on $\psi = \cos\pi x \sin\pi y$. The 3D net at each cos(even)πz level is the same and may be generated from $\psi = \cos\pi x \sin\pi y \cos(\text{even})\pi z$ using $\psi_0 = \cos 2\pi x \sin\pi y$ alone (see fig. 4.2). The net at each cos(odd)πz level starts from $\sin\pi x \sin\pi y \cos(\text{odd})\pi z$ and may also be generated using only $\psi_0 = \cos 2\pi x \sin\pi y$. Square horizontal truncations with vertical truncations at cos(MMAX-1)πz have been used; an N by MMAX truncation contains MMAX.N^2 modes. The dependence of the stability of (7) on the amplitude a_0 is presented in table 4.8; 4.8a provides the dependence on N for the minimum baroclinic truncation (MMAX=2) and 4.8b the dependence on MMAX for the fixed horizontal truncation N=5.

The results with MMAX=2 for N=8 and N=9 agree well, with the main stability transition occurring near to $a_0 = 0.25$. Both coarser truncations (N=4 & 5) in table 4.8a give transitions at lower values of a_0. In table 4.8b the MMAX = 2 truncation has a sharper transition than those of MMAX = 4 and 6, which agree quite well.

The MMAX = 3 and 5 truncations have larger critical values of a_0 than the even MMAX transitions and there is no quantitative agreement between the even and odd MMAX truncations. Whilst the above results are not conclusive they do suggest that the free mode may be stable to these wave based perturbations when $a_0 \leq 0.25$.

Suggestions for future work

Most of the quantities tabulated in this section converge only slowly as the spectral truncation is relaxed. Indeed investigations with improved truncation are required to answer several questions raised in the text. The maximum growth rate and aspect ratio of the return to stability of wave perturbations to $\Psi_0 = 1/\pi \cos 2\pi x \sin \pi y$, for instance, cannot be inferred from table 4.3. Examination of the maximum growth rates of finer even truncations than those presented in tables 4.4 and 4.5 is also desirable since the convergence to stability assumed in the text is very slow and the growth rates with the finest truncations presented are quite substantial. Pinpointing the aspect ratio of the stability transition for harmonic disturbances (table 4.6) by enhancing the resolution is another important task. Finally the results of tables 4.7 and 4.8 would clearly benefit from extension to finer truncations.

Calculation of the eigenvalues and eigenvectors of a wave perturbation with an N=13 truncation takes an IBM3081 about 30 seconds. The time required for these eigenvalue calculations is approximately proportional to N^6. So whilst eigenvalue calculations with N=25 are feasible (though expensive) calculations with N=49 are not. It appears that time integrations of the linearised problem, for which the calculation time is proportional to N^3, are more appropriate for such fine truncations. Investigations of alternatives to the square truncations used throughout this section may also shed light on the convergence problems.

The spatial structures of the most unstable normal modes, particularly those near stability transitions, is of considerable interest. Maps of the contours of the streamfunctions could be produced with relatively little effort. These would complement the

calculations concerning incipient instabilities presented in section 4.4 (see (4.4.13) - (4.4.22)) and might illuminate the reasons for the different behaviour of even and odd truncations. Maps of the incipient k=4, m=1 instability (see table 4.2), which does not conform to "hypothesis" (4), might also be valuable.

The dependence of the stability of Rossby waves in periodic β-plane channels on $M \equiv UK^2/\beta$ could also be investigated. For small M a Rossby wave will be stable except in (the minority) of channels which support almost resonant triads. It would be interesting to see whether or not waves always become more stable as M is decreased from its infinite value on an f-plane. This study would complement those of the stability of (5) and (7) which suggest that zonal flows destabilize barotropic Rossby waves in f-plane channels.

Section 4.4 DISCUSSION

The numerical results of the previous section suggest that there are some Rossby waves in narrow f plane Cartesian channels, other than the channel's gravest modes, that are stable against normal mode perturbations. The results do not prove this, however, and certainly require further interpretation. Analytical studies are needed to explore the senses in which the waves are stable and the location and nature of the stability transitions.

The most enticing approach is to seek conservation relations which account for the stability in the spirit of Fjørtoft's argument or angular momentum conservation (see section 4.1). It is worth noting that if there is such a conservation relation it cannot be satisfied by even truncations of the spectral model (see table 4.1). Furthermore the stability of wave modes other than gravest modes cannot be inferred from the criterion based on local potential vorticity conservation derived by McIntyre & Shepherd (1987). This point is most easily made by considering the integral discussed by Blumen (1968) for the case of an internal baroclinic jet

$$I \equiv \int_V (\partial\psi/\partial x)^2 + (\partial\psi/\partial y)^2 + 1/B\,(\partial\psi/\partial z)^2 + 2H(q)\,dV. \qquad (4.4.1)$$

I is a constant of the motion for any (multi-valued) function H of q with well defined local horizontal derivatives ($\nabla_h H = \partial H/\partial q\ \nabla_h q$). The dependence of I on small perturbations to the streamfunction $\Psi_s(q)$ of a free mode may be investigated by setting $\psi = \Psi_s + \delta\psi$. Blumen shows that on choosing

$$\partial H/\partial q = \Psi_s \qquad (4.4.2)$$

$I - I_s$ reduces to

$$I - I_s = \int_V (\partial/\partial x\,\delta\psi)^2 + (\partial/\partial y\,\delta\psi)^2 + 1/B\,(\partial/\partial z\,\delta\psi)^2 + \partial\Psi_s/\partial q\,\delta q^2\,dV. \qquad (4.4.3)$$

McIntyre & Shepherd (1987) show that if

$$-\partial\Psi_s/\partial q \geq c > 0 \quad \text{throughout domain} \tag{4.4.4}$$

and

$$\int_V (\nabla_h^2 \chi + {}^1\!/\!_B\, \partial^2\chi/\partial z^2)^2 \geq \kappa_0^2 \int_V (\partial\chi/\partial x)^2 + (\partial\chi/\partial y)^2 + {}^1\!/\!_B (\partial\chi/\partial z)^2 dV \tag{4.4.5}$$

for any χ satisfying b.c.s on $\delta\psi$

then I is a maximum at I_s (for arbitrary perturbation fields δq) if $c\kappa_0^2 > 1$. Andrews (1984) has shown that severe restrictions on the applicability of similar results for containers with azimuthal symmetry can be found by considering the effect on I of the perturbations resulting from zonal displacements of the field Ψ_s. By (1), I is clearly unchanged by such displacements. Writing $\omega = \partial\Psi_s/\partial x$ one finds that (3), (4) and (5) imply that

$$\int_V (\partial\omega/\partial x)^2 + (\partial\omega/\partial y)^2 + {}^1\!/\!_B (\partial\omega/\partial z)^2 dV = \int_V -\partial\Psi_s/\partial q\,(\nabla_h^2\omega + {}^1\!/\!_B\,\partial^2\omega/\partial z^2)^2 dV$$

$$\geq c \int_V (\nabla_h^2\omega + {}^1\!/\!_B\,\partial^2\omega/\partial z^2)^2 dV$$

$$\geq c\kappa_0^2 \int_V (\partial\omega/\partial x)^2 + (\partial\omega/\partial y)^2 + {}^1\!/\!_B (\partial\omega/\partial z)^2 dV. \tag{4.4.6}$$

Hence free mode internal jet flows with non-axisymmetric components can only be proved to be stable by these methods if $c\,\kappa_0^2 = 1$ and $\partial\Psi_s/\partial x$ includes only the gravest mode(s) of the container (and $c = -\partial\Psi_s/\partial q$ where $\partial\Psi_s/\partial x \neq 0$).

A second avenue of approach involves investigation of the structure of the determinants of the matrices of various spectral truncations. It is possible to investigate extremely coarse truncations involving only four or five triad interactions analytically but the task is tedious and probably of little value. Numerical investigation of alternatives to the square truncations used in section 4.3 may be more rewarding. The importance of the structure of the perturbation net can be illuminated by analytical studies, however, as the following proof of a restriction inherent in normal modes involving only a single line of perturbation modes shows. Consider the linear perturbation net of fig 4.4. Let $\lambda_1 < \lambda_2$ and $\lambda_i > \lambda_2$ for $i>2$ and consider the case when Λ_0, the total

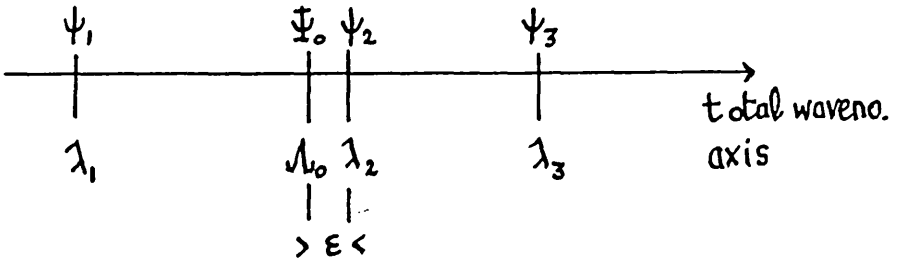


Figure 4.4

A line of modes considered in the linear stability of mode Ψ_0. The total wavenumber Λ_0 of Ψ_0 is considered to be larger than that of only one mode in the line namely ψ_1 and only just smaller than that of the second mode ψ_2.

wavenumber of the main wave Ψ_0, is just less than λ_2. We will show that providing the other λ_i are not close to Λ_0 all even truncations of the perturbation net have an unstable mode. This shows, for example, that a perturbation based on $\cos\pi x \sin\pi y$ to $\cos 2\pi x \sin\pi y$ which includes only a line of modes as in fig. 4.1 will be unstable for values of r just <u>smaller</u> than the critical value of r according to (4.3.4).

Let us denote the mode amplitudes by α_i, the interaction coefficients feeding energy from α_{i+1} into α_i by a_i and those feeding energy from α_{i-1} into α_i by b_{i-1}. Then

$$\begin{aligned} d\alpha_1/dt &= a_1 \alpha_2 \\ d\alpha_i/dt &= b_{i-1}\alpha_{i-1} + a_i \alpha_{i+1} \quad ; \; i > 1 \end{aligned} \tag{4.4.7}$$

and

$$\left.\begin{aligned} a_i &= A_i^2 \langle \phi_i \, J(\Psi_0, \phi_{i+1}) \rangle \, (\Lambda_0 - \lambda_{i+1}) \\ b_i &= B_i^2 \langle \phi_{i+1} J(\Psi_0, \phi_i) \rangle \, (\Lambda_0 - \lambda_i) \end{aligned}\right\} i \geq 1 \tag{4.4.8}$$

Looking for normal modes with growth rate Γ, (7) implies that

$$\det \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & -\Gamma & b_2 & 0 \\ \cdot & a_2 & -\Gamma & b_1 \\ \cdot & 0 & a_1 & -\Gamma \end{pmatrix} = 0 \; . \tag{4.4.9}$$

The lower right hand nxn determinant in (9), D_n, is determined inductively by

$$\begin{aligned} D_1 &= -\Gamma \\ D_2 &= \Gamma^2 - a_1 b_1 \\ D_n &= -\Gamma D_{n-1} - a_{n-1} b_{n-1} D_{n-2} \quad ; \; n > 2 \, . \end{aligned} \tag{4.4.10}$$

Note that when $\Lambda_0 = \lambda_2$, and $\Gamma = 0$, $a_1 = 0$ and $D_{2n} = 0$ for all $n \geq 1$; all even truncations have a neutral mode with $\Gamma = 0$ when $\Lambda_0 = \lambda_2$. Consider now the case when Λ_0 is just less than λ_2 by setting $\lambda_2 - \Lambda_0 = \varepsilon$ ($\varepsilon > 0$ and $\varepsilon \ll 1$). Then a_1 and b_2 are $O(\varepsilon)$ from (8) and from the triple product nature of the overlap integrals

$$a_1 b_1 > 0. \tag{4.4.11}$$

Assuming that all other interaction coefficients a_i and b_i are $O(1)$, and that Γ^2 is of $O(\varepsilon)$ one may prove by induction from (10) that

$$\left.\begin{aligned} D_{2n} &= \text{const.}\,(\Gamma^2 - a_1 b_1) + O(\varepsilon^2) \\ D_{2n+1} &= O(\varepsilon^{3/2}) \end{aligned}\right\} n \geq 1 \tag{4.4.12}$$

with const. of $O(1)$. So that all even truncations have a positive (unstable) eigenvalue with $\Gamma^2 \simeq a_1 . b_1$ for $\lambda_2 - \Lambda_0 = \varepsilon$. This eigenvalue appears to depend little on the level of even truncation. So either the even and odd truncations do not converge as the number of modes included increases or (4.3.4) can only be true for perturbations involving 2D (or 3D) perturbation nets.

When $r^2.(k^2-1) = 3$, $\psi = \cos \pi x \sin 2\pi y$ is a neutral mode on $\psi_0 = \cos k\pi x \sin \pi y$. The instability springing from this mode as r is increased is not captured satisfactorily by the numerical experiments (table 4.3) so a perturbation analysis about the neutral mode comparable with that in section 2.1 for the zonal flow is clearly desirable. The following analysis suggests that the singularity associated with the emerging instability involves areas surrounding the stagnation points of the main Rossby wave. Let $r = r_0 + \delta r$ and Γ again denote the growth rate eigenvalue. The numerical results suggest that $\Gamma^2 = O(\delta r)$. Adopting this scaling we set

$$\begin{aligned} \delta\psi &= \cos \pi x \sin 2\pi y + \Gamma\phi_1 + \Gamma^2\phi_2 + \ldots \\ \Psi_0 &= \cos k\pi x \sin \pi y \quad ; \quad r^2 = r_0^2 + \Gamma^2 r_1^2 + \ldots \end{aligned} \tag{4.4.13}$$

Linearising the potential vorticity equation about Ψ_0 one finds that

$$L\delta\psi = J(\Psi_0, \{\nabla_h^2 + \pi^2(r^2k^2+1)\}\delta\psi) + \Gamma\nabla_h^2\delta\psi = 0. \quad (4.4.14)$$

The boundary conditions for $\delta\psi$ are

$$\delta\psi = 0 \text{ on } y=0,1 \quad ; \quad \delta\psi \text{ periodic in } x. \quad (4.4.15)$$

Integration by parts shows that the adjoint operator for $L\delta\psi$ is

$$M\delta\chi = \{\nabla_h^2 + \pi^2(r^2k^2+1)\} J(\Psi_0, \delta\chi) - \Gamma\nabla_h^2\delta\chi \quad (4.4.16)$$

with $\delta\chi$ satisfying the same boundary conditions as $\delta\psi$.

One suspects that $\Gamma\nabla_h^2\delta\psi$ will play an important role in the solution of (14) near the stagnation points of the streamfunction Ψ_0 (i.e. where $\partial\Psi_0/\partial x = \partial\Psi_0/\partial y = 0$). The series solution (13) is likely to be improved by retaining this term even when it is formally negligible. To order Γ, (14) then yields

$$J(\Psi_0, \{\nabla_h^2 + \pi^2(r_0^2k^2+1)\}\phi_1) + \Gamma\nabla_h^2\phi_1 = -\nabla_h^2(\cos\pi x \sin 2\pi y); \quad (4.4.17)$$

in which we have retained $\Gamma\nabla_h^2\phi_1$ for the reasons just indicated. On the expectation (which is re-evaluated below) that $|\nabla_h^2\phi_1| >> |\phi_1|$ in the region of the stagnation points one may replace $\Gamma\nabla_h^2\phi_1$ by $\Gamma\rho_1$ where

$$\rho_1 = \{\nabla_h^2 + \pi^2(r_0^2k^2+1)\}\phi_1. \quad (4.4.18)$$

The l.h.s. of (17) then reduces to

$$\mathcal{L}\rho_1 \equiv J(\Psi_0, \rho_1) + \Gamma\rho_1 \quad (4.4.19)$$

and the perturbation expansion to a series of problems of the form

$$\mathcal{L}\rho_n = \alpha_n(x,y) \quad (4.4.20)$$

in which α_n is a known function of $\phi_1, \ldots \phi_{n-1}$. Solutions of the

first order partial differential equation (20) are well known (Courant & Hilbert 1962 vol II, ch. I §5 and ch. II §1). The solutions may be found by integrating

$$dx/dt = -\partial\Psi_0/\partial y \; ; \; dy/dt = \partial\Psi_0/\partial x \; ; \; d\rho_n(x,y)/dt = -\Gamma\rho_n + \alpha_n \; . \qquad (4.4.21)$$

In words, following the flow defined by the streamfunction Ψ_0, ρ_n tends to decay exponentially at the rate Γ and is forced by the source term α_n. Solution fields for ρ_n must suffer no net change on being integrated once round a circuit. So if α_n is non-zero at a stagnation point the solution ρ_n must have $\rho_n \simeq \alpha_n/\Gamma$ over a wide neighbourhood surrounding the stagnation point. The r.h.s. of (17) is zero at the stagnation points so ρ_1 is zero there. But ϕ_1 need not be zero at the stagnation points and appears in the forcing of ρ_2

$$J(\Psi_0, \rho_2) + \Gamma\rho_2 = -r_1^2 J(\Psi_0, \pi^2 k^2 \cos\pi x \sin 2\pi y) + \pi^2 (r_0^2 k^2 + 1)\phi_1 \; . \qquad (4.4.22)$$

The dependence of Γ on δr, which is contained in r_1^2 (see (13)), is determined by the requirement that the r.h.s. of (22) be orthogonal to the null solutions of the adjoint problem. Similar analysis to that presented above shows that some of the null solutions of the adjoint problem also contain singularities (for $\Gamma \to 0$) much like those of ρ_n of (21) with $\alpha_n \neq 0$. These singularities are probably responsible for the sensitive dependence of Γ on δr in the odd spectral truncations (table 4.3). But explicit calculation of this dependence through equations (17) and (22) appears a discouragingly heavy task.

It is apparent from the fact that $\rho_1=0$ at the stagnation point that $|\phi_1|$ there is comparable with $|\nabla_h^2\phi_1|$. Since neither are large, however, (20) for n=1 is formally valid. Whether $|\phi_2| << |\nabla_h^2\phi_2|$ and $\Gamma|\phi_2|<<1$ is less clear. It seems that explicit calculations would be necessary to establish whether the proposed series solution would be of any value (e.g. asymptotically valid).

Baroclinic free modes of the form

$$\Psi_0 = \frac{a_0}{\pi} \cos\pi z \sinh \alpha\pi(y-1/2) + \frac{1-a_0}{\pi} \cos k\pi x \sin \pi y$$
$$1/\beta^2 = r^2k^2 + 1 + \alpha^2 \qquad (4.4.23)$$

(cf. (4.3.7) - (4.3.9)) are convenient for stability studies but have some deficiencies as representations of the steady waves found in the differentially heated rotating annulus experiments; the wave component of (23) being barotropic has a null thermal field and fails to lean back with height in a realistic manner (see Hignett et. al. 1985 fig. 7). There are some difficulties also in constructing free modes with reasonable meridional profiles for the zonal flow at the smaller values of the Burger number typical of much of the steady wave regime. For with $(f\ L\ /\ N\ H)^2 = 10$, $r^2 = 0.1$ and $k \leq 6$, (23) yields $\alpha^2 \geq 5.4$ so that $\cosh \alpha\pi/2 \geq 19$; the zonal flow is 19 times as strong at the sidewalls as at mid-channel. This last calculation is very sensitive to the assumed width of the geostrophic interior. If 4L/5 is taken to be appropriate, for example, repeating the above calculation one finds that $\alpha^2 \geq 3.1$ and $\cosh\alpha\pi/2.5 \geq 4.6$, whilst if 3L/5 is taken to be appropriate wavenumber 2 states (k=2) are found to have $\alpha^2 \simeq 2.5$ and $\cosh\ 0.3\alpha\pi \simeq 2.3$.

At the moderate Burger number $\simeq 0.4$ used in section 4.3, (4.3.9), the meridional variation of the zonal flow is less intense but of some importance in assessing values of a_0 representative of steady waves in the laboratory. The ratio of the maximum zonal velocity of the wave to the zonal flow for (23) is

$$\frac{\text{max. zonal flow of wave}}{\text{max. axisymm. flow speed}} = \frac{1-a_0}{a_0} \quad \frac{1}{\alpha \cosh \alpha\pi/2} \qquad (4.4.24)$$

For $\alpha=1$, as in (4.3.8), $\cosh\alpha\pi/2 = 2.5$, so the maximum axisymmetric flow at upper levels is larger than that of the zonal flow of the wave field when $a_0 \geq 0.28$. So baroclinic modes of the form (23) have realistically strong wave components when a_0 is as small as 0.25. The restriction of the stable free modes to those with $a_0 \leq 0.25$ (section 4.3) evidently does not render the free mode

simulation of the laboratory flows totally unrealistic.

The spectral model of section 4.2 could be used to look for steady wave solutions subject to weak Newtonian dissipation on zonal flows tending to relax to non-zero values. Such waves would need to extract energy from the zonal flow and would lean back with height (if they fed off the APE rather than the KE of the zonal flow). The sensitivity of the waves to their dissipation and the forcing of the zonal flow could give some insight into the importance of dissipative processes in the waves in the laboratory systems. The sensitivity of the waves' stability to dissipation is also of interest because dissipation will tend to damp higher wavenumbers preferentially and presumably reduce the level of truncation required to resolve instabilities properly.

CHAPTER FIVE

THE METHOD OF DELAYS

Introduction

In a celebrated paper Ruelle & Takens (1971) argued that the onset of unpredictable (turbulent) motions in fluids could be governed by the same types of transitions as occur in simple non-linear deterministic systems such as ordinary differential equations (o.d.e.s) and maps of the plane and the circle. A considerable body of knowledge concerning the routes to chaos in such simple systems - the standard types of transitions (or series of transitions) which generate the unpredictable motions - has built up since the early 1960s. Useful introductions to the subject, listed in order of increasing notational complexity, are given by Abraham & Shaw (1984), Sparrow (1982) and Guckenheimer & Holmes (1983). Some very carefully controlled Rayleigh-Bénard and Taylor-Couette experiments have been found to exhibit such transitions and the unpredictable flows near the transitions to explore only low dimensional manifolds ("surfaces") in phase space (see the reviews by Swinney (1983) and Guckenheimer (1986)).

The belief that the dynamics of o.d.e.s are relevant to such experiments has been strengthened by some relatively recent investigations of a few dissipative 1D partial differential equations (p.d.e.s) including the complex Ginzburg-Landau equation (Doering *et al.* 1988). All possible initial conditions for these equations have been shown to be attracted onto a finite dimensional Lipschitz manifold on which high wavenumber Fourier modes are slaved to low wavenumber Fourier modes. The long time dynamics of these equations are hence governed by sets of o.d.e.s on finite dimensional manifolds which have been called inertial manifolds (see Constantin *et al.* (1988a) for a review). All initial states of

the 2D Navier-Stokes equations subject to arbitrary forcing have also been shown to be attracted into a subset in the Hilbert phase space which has a finite upper bound on its fractal and Hausdorf dimensions proportional to $G^{2/3}$ $(1 + \log G)^{1/3}$, G being the Grashof number (Constantin et al. (1988b) and Temam 1988).

Recent experimental investigations of baroclinic flows (Guckenheimer & Buzyna (1983) and Hart(1985)) have sought to uncover low dimensional behaviour in the irregular wave regime similar to that found in the Taylor-Couette and Rayleigh-Bénard cells.

The main technique (Broomhead & King 1986) for the reconstruction from experimental measurements of trajectories on the manifold in phase space (to which the flow is assumed to be confined) relies on a method suggested by Ruelle (Guckenheimer 1986) called the method of delays. Results justifying the use of the method have been given by Takens (1981) and Mañé (1981) (see also Eckmann & Ruelle 1985). Its use has unfortunately been hampered by several factors;

i) uncertainty over its conditions of validity
ii) lack of insight into the result of applying the method to a flow to which the theorem does not apply
iii) the fact that the method is purely qualitative; no quantitative estimates of errors in the final construction due to original measurement errors are available
iv) a lack of methods for verifying that trajectories lie on manifolds of moderate dimension (i.e. greater than 3 and smaller than 10).

The aim of this chapter is to provide explicit proofs of a statement of the method of delays similar to that given by Takens and of a related method and to discuss the result of injudicious application of the methods to low dimensional flows. The definitions of the terms and theorems (C.1 to C.5) used in the proofs are presented in appendix C.

An Explanation of the Method of Delays

The method of delays may be used to reconstruct a trajectory, restricted to a compact manifold M of dimension m in phase space, from a single series of (scalar) measurements f(t) of the trajectory by combining measurements at times t, $t+\tau$, $t+2\tau$,.., , $t+2m\tau$, where t is a variable time and τ a constant period, into a (2m+1) dimensional vector $\underset{\sim}{f}(t)$;

$$\underset{\sim}{f}(t) = (f(t), f(t+\tau), \ldots f(t+2m\tau)). \qquad (5.1.1)$$

Takens' theorem establishes conditions under which this trajectory lies on a manifold M^* in (2m+1) dimensional Euclidean space which is diffeomorphic to M in the original phase space (i.e. related to M by a smooth coordinate transformation with a smooth inverse). M^* is said to be an embedding of M. The method is argued to be reliable for most measurement functions, f, vector fields, $\underset{=}{X}$, which generate the flow in phase space and time delays, τ, because the conditions on f, $\underset{=}{X}$ and τ are satisfied by most (a generic subset of) combinations of f, $\underset{=}{X}$ and τ. The other method discussed here requires a time series of 2m+1 simultaneous independent measurements $\{f_i,\ 1 \leq i \leq 2m+1\}$. The trajectory vector

$$\underset{\sim}{f}(t) = (f_1(t), f_2(t), \ldots f_{2m+1}(t)) \qquad (5.1.2)$$

is proved to lie on a manifold M^* in 2m+1 dimensional space which is diffeomorphic to M in phase space for most (a dense and open subset) set of measurement functions $\{f_i\}$.

The proofs of these results are essentially transversality arguments. Given p independent N-1 dimensional surfaces in N dimensional space (p<N) the set of points common to all surfaces will usually lie on surfaces of dimension N-p (i.e. of codimension p). At a typical point $\underset{\sim}{x}$ on an m dimensional manifold M the value of an m dimensional function $\underset{\sim}{f}$ will usually be identical with $\underset{\sim}{f}(\underset{\sim}{x}')$(i.e. $\underset{\sim}{f}$ evaluated at other points) only for isolated points $\underset{\sim}{x}'$. To reduce the surfaces whose points share their value of $\underset{\sim}{f}$ with different points to codimension p, a function of m + p independent

components will be required in general. For a function with 2m+1 components all points will usually have unique values of $\underset{\sim}{f}$. The transversality theorem (see appendix C) is a subtle and rigorous formulation of these simple ideas. In the case of the method of delays points which are advected by the flow (i.e. not stationary) sample the same function at different points. Denoting the position on M of $\underset{\sim}{x}$ after advection for time τ by $\phi(\underset{\sim}{x})$ one may view $\{f(\phi^i(\underset{\sim}{x}), 1 \leq i < 2m+1\}$ as independent measurement functions at such moving points. (1) will provide an embedding provided there is only a finite number of fixed points of the flow.

The conditions on $\underset{\sim}{f}$ for which (2) provides an embedding are presented first (result 1). Most of this result follows from the proof of Whitney's theorem (C.3) (Hirsch 1976 p35). The proof is included because it forms a good introduction to the proof of result 2 which establishes conditions on $\underset{\sim}{f}$ for which (1) provides an embedding, given certain assumptions about $\underset{=}{X}$ and τ. The set of vector fields $\underset{=}{X}$ and times τ for which these assumptions hold is stated directly after the second result.

Result 1

Denote the set of all C^r $(r \geq 1)$ smooth maps $\underset{\sim}{f} : M \to \mathbb{R}^p$ of a compact m-dimensional manifold M into $\mathbb{R}^p$ by P^r and for points $\underset{\sim}{x}$ in M write

$$\underset{\sim}{f}(\underset{\sim}{x}) = (f_1(\underset{\sim}{x}), f_2(\underset{\sim}{x}), \ldots f_p(\underset{\sim}{x})) . \qquad (5.1.3)$$

A dense and open subset of maps in P^r embed all of M in $\mathbb{R}^p$ if $p \geq 2m + 1$.

Proof

It is sufficient to prove that the sets of maps for which

(A) $\underset{\sim}{f}(\underset{\sim}{x}) \neq \underset{\sim}{f}(\underset{\sim}{y})$ if $|\underset{\sim}{x} - \underset{\sim}{y}| > \varepsilon$, for any $\varepsilon > 0$

(B) $\underset{\sim}{f}$ is immersive

are open and dense in P^r if $p \geq 2m+1$. Since, according to the local immersion theorem (C.5), a map which is immersive at any point $\underset{\sim}{x}$ is

injective for all points, $\underset{\sim}{y}$, within some finite neighbourhood of $\underset{\sim}{x}$ and the intersection of two sets which are both open and dense is itself both open and dense, the set of embeddings will hence be both open and dense in P^r.

Part A: $\underset{\sim}{f}(\underset{\sim}{x}) \neq \underset{\sim}{f}(\underset{\sim}{y})$ if $|\underset{\sim}{x} - \underset{\sim}{y}| > \varepsilon$, for any $\varepsilon > 0$

Consider $\underset{\sim}{g}$: $M \times M \to \mathbb{R}^p$ defined by

$$\underset{\sim}{g}(\underset{\sim}{x},\underset{\sim}{y}) = \underset{\sim}{f}(\underset{\sim}{x}) - \underset{\sim}{f}(\underset{\sim}{y}) . \tag{5.1.4}$$

$\underset{\sim}{f}$ fails to be injective at any different points $\{\underset{\sim}{x}, \underset{\sim}{y}$ in $M : \underset{\sim}{x} \neq \underset{\sim}{y}\}$ where $\underset{\sim}{g}(\underset{\sim}{x},\underset{\sim}{y}) = \underset{\sim}{0}$. By the theorem of common zeros (C.1), the set of such points is a manifold of dimension $2m - p$ for $p \leq 2m$, or an empty set for $p \geq 2m + 1$, if at each point in $\underset{\sim}{g}^{-1}(\underset{\sim}{0})$, excepting those where $\underset{\sim}{x} = \underset{\sim}{y}$, $\underset{\sim}{g}$ is submersive.

$\underset{\sim}{g}$ is submersive at each point in the pre-image of $\underset{\sim}{0}$ (i.e. $\underset{\sim}{g}^{-1}(\underset{\sim}{0})$) if and only if $\underset{\sim}{g}$ is transversal to the origin in $\mathbb{R}^p$. By the homotopy stability of transversal intersections (theorem C.2) the set of maps for which $\underset{\sim}{g}$ is transversal to the origin in $\mathbb{R}^p$ is open in P^r.

To complete part A of the proof we prove that for a dense set of maps $\underset{\sim}{f}$ in P^r, $\underset{\sim}{g}$ is submersive at all points $(\underset{\sim}{x}, \underset{\sim}{y})$ if $\underset{\sim}{x}$ and $\underset{\sim}{y}$ are not arbitrarily close together. We need the following additional notation. Let $\underset{\sim}{i}_1, \underset{\sim}{i}_2, \ldots \underset{\sim}{i}_{2m+1}$ be $2m+1$ mutually orthogonal axes in $\mathbb{R}^{2m+1}$. Denote by $\underset{=}{Q}^j$ $(2 \leq j \leq 2m+1)$ the "rotation" matrix which maps $\underset{\sim}{i}_1$ to $-\underset{\sim}{i}_j$ and $\underset{\sim}{i}_j$ to $\underset{\sim}{i}_1$ and leaves the other basis vectors unchanged and by $\underset{=}{Q}^1$ the identity matrix. Also let $\underset{\sim}{h}$

$$\underset{\sim}{h} = (h_1, h_2, \ldots h_{2m+1}) \; ; \; \underset{\sim}{h} : M \to \mathbb{R}^{2m+1} \tag{5.1.5}$$

map M diffeomorphically onto its image in $\mathbb{R}^{2m+1}$. There must be such a map available by Whitney's theorem (theorem C.3) . For each pair $(\underset{\sim}{x}, \underset{\sim}{y})$, $\underset{\sim}{x} \neq \underset{\sim}{y}$ in $M \times M$, at least one of the vectors $\underset{=}{Q}^j [\, \underset{\sim}{h}(\underset{\sim}{x}) - \underset{\sim}{h}(\underset{\sim}{y}) \,]$, $1 \leq j \leq 2m+1$, call it $\underset{\sim}{z}$, will have a non zero projection on axis $\underset{\sim}{i}_1$. $\{\underset{=}{Q}^j \underset{\sim}{z}$, $1 \leq j \leq p\}$ will form a set of p vectors

which span $\mathbb{R}^p$. We apply the transversality theorem (C.4) to the map

$$\underset{\sim}{G}(\underset{\sim}{x},\underset{\sim}{y},\underline{\underline{b}}) = \underset{\sim}{F}(\underset{\sim}{x},\underline{\underline{b}}) - F(\underset{\sim}{y},\underline{\underline{b}}) \quad ; \quad G : M\times M\times B^{4m+2} \to \mathbb{R}^p \qquad (5.1.6)$$

defined using

$$F_i(\underset{\sim}{x},\underline{\underline{b}}) = f_i(\underset{\sim}{x}) + b_{jk}\, O^j_{i\ell}\, O^k_{\ell n}\, h_n(\underset{\sim}{x}) \qquad (5.1.7)$$

in which $O^j_{i\ell}$ are the elements of the matrix $\underline{\underline{O}}^j$, the summation convention is employed for repeated indices and $\underline{\underline{b}}$ belongs to B^{4m+2}, the open unit ball in $\mathbb{R}^{4m+2}$. $\underset{\sim}{G}$ is a submersion for all different points $\underset{\sim}{x}$, $\underset{\sim}{y}$ because, for fixed different values of $\underset{\sim}{x}$ and $\underset{\sim}{y}$, the Jacobian derivative of $\underset{\sim}{G}$ with respect to $\underline{\underline{b}}$ is

$$\partial G_i/\partial b_{jk} = O^j_{i\ell}\, O^k_{\ell n}[h_n(\underset{\sim}{x}) - h_n(\underset{\sim}{y})] = O^j_{i\ell}\{\underline{\underline{O}}^k(\underline{\underline{h}}(\underset{\sim}{x}) - \underline{\underline{h}}(\underset{\sim}{y}))\}_\ell \; . \qquad (5.1.8)$$

One of the vectors $\underline{\underline{O}}^k$ [$\underline{\underline{h}}(\underset{\sim}{x})$ - $\underline{\underline{h}}(\underset{\sim}{y})$] will have non zero projection on the axis $\underset{\sim}{i}_1$ for each $\underset{\sim}{x}$ and $\underset{\sim}{y}$ so the Jacobian will be of rank p.

Direct application of the transversality theorem shows that $\underset{\sim}{g}_b$ is a submersion for almost every $\underline{\underline{b}}$ in the 4m+2 dimensional open unit sphere and hence that $\underset{\sim}{g}$ is submersive at all points where $|\underset{\sim}{x}-\underset{\sim}{y}| > \varepsilon$ for a dense set of maps $\underset{\sim}{f}$ in P^r.

Part B: $\underset{\sim}{f}$ is immersive

To be immersive at $\underset{\sim}{x}$ the linear map $\underline{\underline{T}}\,\underset{\sim}{f} : T(M) \to \mathbb{R}^p$

$$\underline{\underline{T}}\,\underset{\sim}{f}(\underset{\sim}{x},\underset{\sim}{v}) = df_{\underset{\sim}{x}}(\underset{\sim}{v}) \quad ; \quad \underset{\sim}{v} \in T_x(M) \qquad (5.1.9)$$

must be injective at $\underset{\sim}{x}$. Any linear map $\underline{\underline{L}}(\underset{\sim}{v})$ is injective if and only if $\underline{\underline{L}}(\underset{\sim}{v}) = \underset{\sim}{0}$ implies $\underset{\sim}{v} = \underset{\sim}{0}$. Hence by theorem C.1 the set of points $(\underset{\sim}{x},\underset{\sim}{v})$ in T(M) at which $\underset{\sim}{f}$ is not immersive is a manifold of codimension p, if at each point in $(\underline{\underline{T}}f)^{-1}(\underset{\sim}{0})$, with $|\underset{\sim}{v}| = 1$, $\underline{\underline{T}}\,\underset{\sim}{f}$ is submersive. By theorem C.2 the set of maps which satisfy the latter condition are open in P^r. Use of the "homotopy"
$\underline{\underline{T}}\,\underset{\sim}{F} : T(M) \times B^{4m+2} \to \mathbb{R}^p$

$$\underline{\underline{T}}\,\underline{F}(\underset{\sim}{x},\underset{\sim}{v}) = d\underset{\sim}{F}_x(\underset{\sim}{v}) \quad , \quad \underset{\sim}{v} \in T_x(M) \tag{5.1.10}$$

in which $\underset{\sim}{F}$ is defined by (7), in the same manner as for part A, completes the proof (because $\underset{\sim}{h}$ is an immersion of M in $\mathbb{R}^{2m+1}$).

If $\underset{\sim}{f}$ is not immersive at $(\underset{\sim}{x},\underset{\sim}{v})$ in T(M) it will not be immersive at $(\underset{\sim}{x}, t\underset{\sim}{v})$ in T(M) for all real values of t. Hence $\underset{\sim}{f}$ cannot fail to be immersive on a submanifold of T(M) of codimension 2m (i.e. at an isolated point).

The proof for result 2 is similar to that for result 1 except that a new function is sought to apply the transversality theorem in place of (7).

Result 2

Let M be a compact m dimensional manifold, $\underline{\underline{X}}$ a vector field on M with only a finite number of hyperbolic fixed points and ϕ a map of M, given by the flow generated by $\underline{\underline{X}}$ applied for a fixed time τ, which has no orbits of period less than or equal to 2m+1 (i.e. if $\phi^j(\underset{\sim}{x}) = \underset{\sim}{x}$ for any $\underset{\sim}{x}$ in M and $1 \leq j \leq 2m+1$ then $\underline{\underline{X}}(\underset{\sim}{x}) = \underset{\sim}{0}$) and for which all the eigenvalues of the map linearised about all the fixed points are different. Denote the set of all C^r ($r \geq 1$) smooth maps $f : M \to \mathbb{R}$ by T^r and for points $\underset{\sim}{x}$ in M write

$$\underset{\sim}{f}_\phi(\underset{\sim}{x}) = \left(f(\underset{\sim}{x}), f(\phi(\underset{\sim}{x})), \ldots . f(\phi^{p-1}(\underset{\sim}{x})) \right) . \tag{5.1.11}$$

An open and dense subset of T^r generate maps $\underset{\sim}{f}_\phi(\underset{\sim}{x})$ which embed all of M in $\mathbb{R}^p$ if $p \geq 2m + 1$.

Proof

Part A: $\underset{\sim}{f}_\phi(\underset{\sim}{x}) \neq \underset{\sim}{f}_\phi(\underset{\sim}{y})$ if $|\underset{\sim}{x} - \underset{\sim}{y}| > \varepsilon$, for any $\varepsilon > 0$

Let the distance function ρ be a metric for the manifold. We consider first whether $\underset{\sim}{f}_\phi$ is injective ($\underset{\sim}{f}_\phi(\underset{\sim}{x}) \neq \underset{\sim}{f}_\phi(\underset{\sim}{y})$) for points $\underset{\sim}{z}=(\underset{\sim}{x},\underset{\sim}{y})$ in a set S of restricted areas of M x M. The restrictions,

namely (12), (13) and (16), are expressed in terms of a distance s which is assumed to be finite but allowed to become arbitrarily small later in the argument. For each $\underset{\sim}{z}$ in S the points $\underset{\sim}{x}$ and $\underset{\sim}{y}$ must be at least distance s apart and $\phi^q(\underset{\sim}{x})$ must be at least distance s from $\phi^q(\underset{\sim}{y})$ for each q, $0 \leq q \leq p-1$;

$$\rho(\phi^q(\underset{\sim}{x}), \phi^q(\underset{\sim}{y})) > s \quad \text{for } 0 \leq q \leq p-1. \tag{5.1.12}$$

The images of each point $\underset{\sim}{x}$ under p-1 iterations of ϕ must also be at least distance s apart and the same condition is required on each point $\underset{\sim}{y}$;

$$\rho(\phi^q(\underset{\sim}{x}), \phi^t(\underset{\sim}{x})) > s \quad \text{for } 0 \leq q,t \leq p-1 \text{ unless } q=t \tag{5.1.13a}$$

$$\rho(\phi^q(\underset{\sim}{y}), \phi^t(\underset{\sim}{y})) > s \quad \text{for } 0 \leq q,t \leq p-1 \text{ unless } q=t. \tag{5.1.13b}$$

For finite s, since M is compact, it may be covered by a finite number, n, of overlapping m dimensional submanifolds N_i each of which is connected and lies within some m dimensional sphere of radius

$$R = s/4. \tag{5.1.14}$$

The overlap between the manifolds must be greater than η and smaller than ν at all points on the submanifold boundaries ($\eta < \nu$). The distance η can be made as small as one chooses. Let $\{ \theta_i, \ 1 \leq i \leq n \}$ be a set of smooth functions with

$$\theta_i = 0 \quad \text{outside } N_i; \tag{5.1.15a}$$

$$\theta_i = 1 \quad \text{inside } N_i \text{ distance} > \eta \text{ from its bdy}. \tag{5.1.15b}$$

Let U_j, V_j be the set of manifolds N_i for which $\theta_i(\underset{\sim}{z}) \neq \underset{\sim}{0}$ for $\underset{\sim}{z} = \phi^{j-1}(\underset{\sim}{x})$ and for $\underset{\sim}{z} = \phi^{j-1}(\underset{\sim}{y})$ respectively. The final restriction on S concerns the interlocking of these manifolds and their images after one application of ϕ;

$$\phi(U_j) \cap U_K = 0 \quad \text{for} \quad 1 \leq K \leq j \leq p \tag{5.1.16a}$$

$$\phi(V_j) \cap V_K = 0 \quad \text{for} \quad 1 \leq K \leq j \leq p . \tag{5.1.16b}$$

This restriction requires $\phi^j(\underset{\sim}{x})$ not to belong to U_1 for $1 \leq j \leq p$.

When s is sufficiently small and $p \leq 2m+1$, since ϕ has no orbits of period $\leq 2m+1$, all of M x M other than small neighbourhoods of i) $\underset{\sim}{x}=\underset{\sim}{y}$, ii) fixed points $\underset{\sim}{x}$ (for any $\underset{\sim}{y}$) and iii) fixed points $\underset{\sim}{y}$ (for any $\underset{\sim}{x}$) will satisfy all these conditions ((12), (13) and (16)).

By construction

$$U_j \cap V_j = 0 \qquad 1 \leq j \leq p \tag{5.1.17}$$

$$U_j \cap U_K = 0 \qquad 1 \leq j,K \leq p \quad \text{unless} \quad j=K \tag{5.1.18a}$$

$$V_j \cap V_K = 0 \qquad 1 \leq j,K \leq p \quad \text{unless} \quad j=K. \tag{5.1.18b}$$

(17) follows from (12) and (18) from (13). Furthermore from (13) and (16)

$$\{ U_K, U_\ell, \ldots . \} \neq \{ V_K, V_\ell, \ldots . \} \quad \text{for} \quad 1 \leq K, \ell \ldots . \leq p \tag{5.1.19}$$

i.e. the manifolds N_i visited by the image of $\underset{\sim}{x}$ under a selection of diffeomorphisms formed by iterating ϕ up to p-1 times, is not identical with the set visited by $\underset{\sim}{y}$ under the <u>same</u> selection of maps.

Let

$$F(\underset{\sim}{x}, \underset{\sim}{b}) = f(\underset{\sim}{x}) + \sum_{i=1}^{n} b_i \theta_i(\underset{\sim}{x}) \tag{5.1.20}$$

where $\underset{\sim}{b}$ belongs to B^n, and construct $\underset{\sim}{G}(\underset{\sim}{x},\underset{\sim}{b})$ by (6) as before. Consider the Jacobian derivative of $\underset{\sim}{G}$, for fixed $\underset{\sim}{x}$ and $\underset{\sim}{y}$, with respect to $\underset{\sim}{b}$

$$\partial G_i / \partial b_j (\underset{\sim}{x}, \underset{\sim}{y}) = \theta_j(\phi^{i-1}(\underset{\sim}{x})) - \theta_j(\phi^{i-1}(\underset{\sim}{y})) . \tag{5.1.21}$$

Take a fixed value of $\underset{\sim}{x}$ and $\underset{\sim}{y}$ and to simplify notation let $\phi^{i-1}(\underset{\sim}{x})$ have $\theta_i=1$ for $1\leq i\leq p$. Then all the diagonal elements of the first p rows of $\partial\underset{\sim}{G}/\partial\underset{\sim}{b}$ are equal to 1 (by (17)). By (18) each of the rows and columns of the Jacobian contain at most 2 elements (+1 on the diagonal and a negative off diagonal element). The first p rows of the Jacobian can then only be linearly dependent if some subset of non zero off diagonal elements $\{g_{ij}\}$ has the same set of row subscripts as column subscripts (as in $\{g_{25}, g_{53}, g_{34}, g_{42}\}$ for which both row and columns have subscripts 2,3,4 and 5). (19) ensures that the first p rows can be chosen so that this is not the case and hence that $\partial\underset{\sim}{G}/\partial\underset{\sim}{b}$ is of rank p.

If $\underset{\sim}{y}$ lies in the neighbourhood of a fixed point (i.e. $\phi(\underset{\sim}{y}) = \underset{\sim}{y}$) but $\underset{\sim}{x}$ does not the only alteration to the above argument that is required is that conditions (13b) and (16b) be dropped and replaced by

$$\rho\left(\phi^{q}(\underset{\sim}{x}), \phi^{t}(\underset{\sim}{y})\right) > \varepsilon \quad \text{for } 0\leq q,t\leq p-1. \tag{5.1.22}$$

A similar argument applies if $\underset{\sim}{x}$ is a fixed point but $\underset{\sim}{y}$ is not.

Let $\{\underset{\sim}{x}_i, 1\leq i\leq I\}$ denote the finite set of fixed points. Then the set of functions with

$$f(\underset{\sim}{x}_i) \neq f(\underset{\sim}{x}_j) \qquad 1\leq i,j\leq I \tag{5.1.23}$$

is also open and dense. Part A of the proof is completed by showing that $\underset{\sim}{f}_\phi$ with p=2m+1 is injective when confined to sufficiently small neighbourhoods of all the fixed points at the end of part B.

<u>Part B:</u> $\underset{\sim}{f}_\phi$ is immersive

The following transversality function can be used to show that $\underset{\sim}{f}_\phi(\underset{\sim}{x})$ of (11) with p=2m+1 is immersive (from which it follows that it is locally injective) for an open and dense subset of T^r provided that $\underset{\sim}{x}$ is not a fixed point.

$$F = \sum_{k=1}^{n}\sum_{j=1}^{2m+1} b_{jk}\ \underline{i}_j\cdot\underline{\underline{I}}_k\ \underline{h}(\underset{\sim}{x}). \tag{5.1.24}$$

Here $\underset{\sim}{h}(\underset{\sim}{x})$ is an embedding of M in $\mathbb{R}^{2m+1}$, $\{ \underset{\sim}{i}_j, \ 1 \leq j \leq 2m+1 \}$ a mutually orthogonal set of vectors in $\mathbb{R}^{2m+1}$ (as in the proof of the first result) and $I_k : \mathbb{R}^{2m+1} \to \mathbb{R}^{2m+1}$ is the identity matrix inside the image of N_k and the zero matrix outside it

$$\underset{\sim}{\underset{=}{I}}_K = \underset{\sim}{\underset{=}{0}} \quad \text{on } \underset{\sim}{h}(M) \text{ outside } \underset{\sim}{h}(N_K)$$

$$\underset{\sim}{\underset{=}{I}}_K = \underset{\sim}{\underset{=}{I}} \quad \text{inside } \underset{\sim}{h}(N_K) \text{ distance} > \eta \text{ from bdy of } N_K. \tag{5.1.25}$$

To prove that $\underset{\sim}{f}_\phi$ is immersive at a fixed point, $\underset{\sim}{x}_1$, when $p \geq m+1$, we infer from

$$d\underset{\sim}{f}_\phi(\underset{\sim}{z}) = \underset{\sim}{0} \tag{5.1.26}$$

that $\underset{\sim}{z}=\underset{\sim}{0}$. Let $\underset{\sim}{\underset{=}{L}}\underset{\sim}{z}$ represent $d\phi(\underset{\sim}{z})$ at $\underset{\sim}{x}_1$. By assumption $\underset{\sim}{\underset{=}{L}}$ has m different eigenvalues and eigenvectors $(\lambda_j, \underset{\sim}{e}_j)$. Writing $\underset{\sim}{z} = z_j \underset{\sim}{e}_j$ and $df(\underset{\sim}{z}) = \partial f/\partial e_j \, z_j$, (26) implies that

$$\sum_{j=1}^{m} \partial f/\partial a_j \ \lambda_j^i \ z_j = 0 \qquad \text{for } 0 \leq i \leq m. \tag{5.1.27}$$

(27) may be written as

$$\underset{\sim}{\underset{=}{P}} \underset{\sim}{\delta f} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ \lambda_1 & \lambda_2 & \dots & \lambda_m \\ \vdots & \vdots & & \vdots \\ \lambda_1^m & \lambda_2^m & \dots & \lambda_m^m \end{pmatrix} \begin{pmatrix} \partial f/\partial e_1 z_1 \\ \partial f/\partial e_2 z_2 \\ \vdots \\ \partial f/\partial e_m z_m \end{pmatrix} = \underset{\sim}{0} \tag{5.1.28}$$

The determinant of P is zero if $\lambda_i = \lambda_j$ for any $i \neq j$. Indeed

$$\det \underset{\sim}{\underset{=}{P}} = \text{const.} \prod_{i,j=1;\, j>i}^{m} (\lambda_i - \lambda_j)$$

so det P is non zero given that the λ_i are all different. For an open and dense subset of T^r, $\partial f/\partial e_j \neq 0$ for $1 \leq j \leq m$, so $\underset{\sim}{z}=0$ and $\underset{\sim}{f}_\phi$ is immersive at each fixed point. The local immersion theorem (C.5) then implies that $\underset{\sim}{f}_\phi$ is injective within a sufficiently small but finite neighbourhood of each fixed point.

To show that $\underset{\sim}{f}_\phi$ is injective in a neighbourhood of all fixed points when $p=2m+1$ consider points $\underset{\sim}{x}+\underset{\sim}{x}_1$ and $\underset{\sim}{y}+\underset{\sim}{x}_2$ in the neighbourhood of fixed points $\underset{\sim}{x}_1$ and $\underset{\sim}{x}_2$ respectively and let $\Delta f = f(\underset{\sim}{x}_2)-f(\underset{\sim}{x}_1)$. $\underset{\sim}{f}_\phi$ is injective unless

$$\Delta f = \sum_{j=1}^{m} \left(\partial f/\partial e_j \, \lambda_j^i x_j - \partial f/\partial c_j \, \eta_j^i y_j \right) \qquad (5.1.29)$$

where $(\lambda_j, \underset{\sim}{e}_j)$ and $(\eta_j, \underset{\sim}{c}_j)$ are the eigenvalues and vectors of $d\phi$ linearised about $\underset{\sim}{x}_1$ and $\underset{\sim}{x}_2$ respectively. Subtraction of (29) with $i=1$ from (29) for each of i, $1 \leq i \leq 2m$, yields

$$\begin{pmatrix} \lambda_1 - 1 & \cdots & \lambda_m - 1 & \eta_1 - 1 & \cdots & \eta_m - 1 \\ \vdots & & \vdots & \vdots & & \vdots \\ \lambda_1^{2m} - 1 & \cdots & \lambda_m^{2m} - 1 & \eta_1^{2m} - 1 & \cdots & \eta_m^{2m} - 1 \end{pmatrix} \begin{pmatrix} \partial f/\partial e_1 x_1 \\ \vdots \\ \partial f/\partial c_m y_m \end{pmatrix} = \underset{\sim}{0} \, . \qquad (5.1.30)$$

For hyperbolic fixed points with different eigenvalues ($\lambda_i \neq \eta_j$) and $\partial f/\partial e_i \neq 0$, $\partial f/\partial c_i \neq 0$, (30) implies that $\underset{\sim}{x}=\underset{\sim}{y}=\underset{\sim}{0}$. Then (29) for $i=0$ implies that $\Delta f = 0$. So by (23) $\underset{\sim}{f}_\phi$ is injective for an open and dense subset of T^r.

The requirements on $\underline{\underline{X}}$ and τ specified in the statement of result 2 are satisfied by open and dense subsets of vector fields and finite time delays for a natural topology on the space of smooth vector fields on a compact manifold (Palis & de Melo (1982), hereafter PM, ch. 1 § 2). The subset of vector fields, $\underline{\underline{X}}$, with only a finite number of fixed points is open and dense (PM ch. 2, § 3, proposition 3.3) and the subset for which the linearised flows about all the fixed points have different eigenvalues is also open and dense. The subset of vector fields with only a finite number of closed (periodic) orbits of period $\leq T$ (where T is any finite number) is also open and dense (PM ch. 3, § 2). So any τ from an open and dense subset of real numbers on the interval $0<\tau<T$ will provide a map ϕ with no periodic points of period $\leq 2m+1$ other than the fixed points of $\underline{\underline{X}}$.

Discussion

The proof of result 2 is very similar to that given by Takens (1981), the main difference being the length of presentation. There are only four substantial points of comparison worth noting. Takens' theorem 2 asserts that the method of delays provides an embedding only for a generic choice of vector field, measurement function and delay time. It appears from result 2 and the theorems of Palis & de Melo (1982) that the method of delays applies to open and dense subsets of vector fields, finite delay times and measurement functions. Secondly the text of Takens proof contains a typing error in the specification of conditions on the open subsets covering M. Condition (ii) of Takens (p370) should read "for each $i,j=1,..N$ and $k,\ell=0,1,..2m$, $\overline{\phi}^{-k}(U_i) \cap U_j \neq 0$ and $\overline{\phi}^{-\ell}(U_i) \cap U_j \neq 0$" (i.e. $\overline{\phi}^{-\ell}$ not $\overline{\phi}^{\ell}$). Thirdly Takens' statement of the conditions on the covering sets is simpler than that given above. Finally Takens' proof that $\underset{\sim}{f}_\phi$ is injective in the joint neighbourhood of two fixed points is inexplicit.

As mentioned in the introduction, proofs of the method of delays provide no indication of how measurement errors may amplify and corrupt the embedding. Indeed since no measurement process strictly provides a function (i.e. a unique value according to position in phase space) let alone a smooth function the value of the theorem for real measurements is open to question. Furthermore as the justification of the technique does not claim to distinguish mountains from mole-hills its practical value is unclear. Concern over these points is heightened by the almost arbitrarily flexible nature of the function f used in the proof. A partial response to the first criticism is provided by the formulation of the sub-manifolds N_i and the sets of points $(\underset{\sim}{x},\underset{\sim}{y})$ in M x M for which $\underset{\sim}{f}_\phi$ is injective in terms of a variable distance s. To capture the main features of the manifold it is essential that widely separated points should be widely separated in the embedding space. This will depend largely on two things; i) the size of the submanifolds, N_i, which can be chosen for a given minimal separation of points $\underset{\sim}{x}$ and $\underset{\sim}{y}$ and ii) the "articulation" of the measurement function f. A poor choice of time delay (close to a small multiple of a natural period)

could lead to $\phi(\underset{\sim}{x})$ being close to $\underset{\sim}{x}$ (for some j, $1 \leq j \leq p$) for a large fraction of the points on the attractor. This would result in the size of the manifolds N_i which are adequate for any separation being significantly reduced. But it seems plausible that a mapping will embed a given scale of features if a representative variation in f between neighbouring manifolds N_i of an adequate set (for that scale) considerably exceeds the noise level. So given that one chooses a suitable time delay, in which "spurious" nearly periodic orbits are avoided, one can reasonably expect some or many features of the attractor to be revealed according to the signal to noise ratio of the measurements.

There has been considerable speculation over the restrictions on the time delay which should be employed partly because in practice the method of delays is disconcertingly sensitive to it. If τ is chosen equal to the period of an orbit, all points $\underset{\sim}{x}$ on the orbit will be liable to have $\underset{\sim}{f}_\phi(\underset{\sim}{x})=\underset{\sim}{f}_\phi(\underset{\sim}{y})$ for some other point on the orbit. If thrice the delay time equals the period of an orbit, however, $\underset{\sim}{f}_\phi$ is unlikely to fail to be injective, when confined to the orbit itself (assuming it to be isolated). Nevertheless $\underset{\sim}{f}_\phi$ will probably fail to be immersive on the orbit; whilst one can derive an expression somewhat similar to (28) for any point on the orbit, one cannot infer that $\underset{\sim}{z}=\underset{\sim}{0}$ because $\partial f/\partial e_i$ will vanish for some point on the orbit for too large a set of functions f.

The proofs of the method of delays show that if τ is extremely short or $i.\tau$, $1 \leq i \leq 2m+1$, is close to an orbit's period the size of the manifolds N_i required to make $\underset{\sim}{f}_\phi$ an embedding for most points may be considerably reduced and hence the method ineffective. If τ is very long then divergence of nearby points in phase space could also reduce the size of the manifolds N_i and produce poor results (Ruelle 1977). It is not clear, however, that these points account entirely for the perceived sensitivity to τ. Certainly felicitous choices of τ could give rise to embeddings being achieved in smaller dimensions than 2m+1. The conditions for such felicitous choices must not be confused with the conditions for the validity of the method.

As suggested in the introduction the method of delays will provide an embedding for most points on M if $m+1<p<2m+1$. This is important because most visual Poincaré sections are produced using $p<2m+1$. The dimension of the "surfaces" on M x M at which $\underset{\sim}{f}_\phi(\underset{\sim}{x}) = \underset{\sim}{f}_\phi(\underset{\sim}{y})$ for $\underset{\sim}{x} \neq \underset{\sim}{y}$ and on T(M) at which $\underset{\sim}{f}_\phi$ fails to be immersive in M are of dimension 2m-p. The typical shapes of the surfaces in M or $\mathbb{R}^p$ on which the embedding fails may be worth investigation.

Many recent investigations have used the method of delays to reconstruct phase portraits because the quality of a single measurement time series was significantly superior to that of combined measurements of several quantities. Where joint measurements of comparable quality are available it is probably better to use them as in equation (2). This method avoids the various symmetries imposed on the phase space by the method of delays (such as the confinement of fixed points to the diagonal in $\mathbb{R}^p$) and the problems of sensitivity to the time delay. A combination of the two methods using a delay time and several measurements could be used if the number of joint measurements were less than 2m+1. Combinations of measurements using unrelated delay times $T_1, T_2, \ldots, T_{2m}$

$$\underset{\sim}{f}(t) = \left(f(t), f(t+T_1), \ldots f(t+T_{2m}) \right) \qquad (5.1.31)$$

(Eckmann & Ruelle 1985) may also be more effective than use of related delays (i.e. $T_n = \tau_n$). This method can be justified using arguments similar to those given above.

REFERENCES

Abraham, R. H., and C. D. Shaw, 1984: Dynamics-the geomety of behavior; Part One: Periodic behavior; Part Two: Chaotic behavior; Part Three: Global behavior. Santa Cruz, Aerial Press.

_____, J. E. Marsden, and T. Ratiu, 1983: Manifolds, Tensor Analysis and Applications. London, Addison-Wesley, 582 pp.

Andrews, D. G., 1983: A conservation law for small-amplitude quasi-geostrophic disturbances on a zonally asymmetric basic flow. J. Atmos. Sci., 40, 85-90.

_____, 1984: On the existence of nonzonal flows satisfying sufficient conditions for stability. Geophys. Astrophys. Fluid Dyn., 28, 243-256.

Arnol'd V. I., 1965: Conditions for nonlinear stability of stationary plane curvilinear flows of an ideal fluid. Dokl. Akad. Nauk. SSSR, 162, 975-978. (English transl.: Soviet Math., 6, 773-777 (1965).)

_____, 1966: On an a priori estimate in the theory of hydrodynamic stability. Izv. Vyssh. Uchebn. Zaved. Matematika, 54, no. 5, 3-5. (English transl.: Amer. Math. Soc. Transl. Series 2, 79, 267-269 (1969).)

_____, 1978: Mathematical Methods of Classical Mechanics. Springer-Verlag, 462 pp.

Baines, P. G., 1976: The stability of planetary waves on a sphere. J. Fluid Mech., 73, 193-213.

Barcilon, V., 1964: Role of the Ekman layers in the stability of the symmetric regime obtained in a rotating annulus. J. Atmos. Sci., 21, 291-299.

Batchelor, G. K., 1967: An Introduction to Fluid Dynamics. London, Cambridge Univ. Press, 615 pp.

Bell, M. J., and A. A. White, 1988a: The stability of internal baroclinic jets: some analytical results. J. Atmos. Sci., 45, 2571-2590.

_____, and A. A. White, 1988b: Spurious stability and instability in N-level quasi-geostrophic models. J. Atmos. Sci., 45, 1731-1738.

Blumen, W., 1968: On the stability of quasi-geostrophic flow. J. Atmos. Sci., 25, 929-931.

Bretherton, F. P., 1966a: Critical layer instability in baroclinic flows. Quart. J. Roy. Meteor. Soc., 92, 325-334.

_____, 1966b: Baroclinic instability and the short wavelength cut-off in terms of potential vorticity. Quart. J. Roy. Meteor. Soc., 92, 335-345.

Broomhead, D. S., and G. P. King, 1986: Extracting qualitative dynamics from experimental data. Physica, 20, D, 217-236.

Chandrasekhar, S., 1961: Hydrodynamic and hydromagnetic stability. Oxford, Clarendon Press, 652 pp.

Charney, J. G., 1971: Geostrophic turbulence. J. Atmos. Sci., 28, 1087-1095.

_____, 1973: Planetary fluid mechanics. Dynamical Meteorology, Ed. Morel, P., Dordrecht, Reidel, 97-351.

_____, and M. E. Stern, 1962: On the stability of internal baroclinic jets in a rotating atmosphere. J. Atmos. Sci., 19, 159-172.

Constantin, P., C. Foias, B. Nicolaenko and R. Temam, 1988a: Integral and inertial manifolds in dissipative nonlinear partial differential equations. Springer series in Appl. Maths, 70, New-York, Springer-Verlag, 165 pp.

_____, C. Foias, and R. Temam, 1988b: On the dimension of the attractors in 2-dimensional turbulence. Physica, 30, D, 284-296.

Courant, R., and D. A. Hilbert, 1953: Methods of mathematical physics, Vol. I. London, Interscience, 561 pp.

_____, and D. A. Hilbert, 1962: Methods of mathematical physics, Vol. II. London, Interscience, 830 pp.

Dieudonné, J., 1960: Foundations of Modern Analysis. New York, Academic Press, 361 pp.

Doering, C. R., J. D. Gibbon, D. D. Holm and B. Nicolaenko, 1988: Low-dimensional behaviour in the complex Ginzburg-Landau equation. Nonlinearity, 1, 279-309.

Drazin, P. G., 1970: Non-linear baroclinic instability of a continuous zonal flow. Quart. J. Roy. Meteor. Soc., 96, 667-676.

_____, and W. H. Reid, 1981: Hydrodynamic stability. London, Cambridge Univ. Press, 525 pp.

Eady, E.T., 1949: Long waves and cyclone waves. Tellus, 1, 33-52.

Eckmann, J.-P., and D. Ruelle, 1985: Ergodic theory of chaos and strange attractors. Rev. Modern Phys., 57, 617-656.

Eliassen, A., 1983: The Charney-Stern theorem on barotropic-baroclinic instability. Pure Appl. Geophys., 121, 563-572.

Fjørtoft, R., 1951: Stability properties of large-scale atmospheric disturbances. Compendium of Meteorology. Ed. Malone, T., Boston, Amer. Meteor. Soc., 454-463.

_____, 1953: On the changes in the spectral distribution of kinetic energy for twodimensional, nondivergent flow. Tellus, 5, 225-230.

Fowlis, W. W., and R. Hide, 1965: Thermal convection in a rotating annulus of liquid: effect of viscosity on the transition between axisymmetric and non-axisymmetric flow regimes. J. Atmos. Sci., 22, 541-558.

Gill, A. E., 1974: The stability of planetary waves on an infinite beta-plane. Geophys. Fluid Dyn., 6, 29-47.

_____, 1982: Atmosphere-Ocean Dynamics. New York, Academic Press, 662 pp.

Green, J. S. A., 1960: A problem in baroclinic stability. Quart. J. Roy. Meteor. Soc., 86, 237-251.

Greenspan, H. P., 1968: The Theory of Rotating Fluids. Cambridge, Cambridge Univ. Press, 327 pp.

Guckenheimer, J., 1986: Strange attractors in fluids: another view. Ann. Rev. Fluid Mech., 18, 15-31.

_____, and G. Buzyna, 1983: Dimension measurements for geostrophic turbulence. Phys. Rev. Lett., 51, 1438-1441.

_____, and P. Holmes, 1983: Non-linear Oscillations, Dynamical Systems and Bifurcations of Vector Fields. New York, Springer-Verlag, 459 pp.

Guillemin, V., and A. Pollack, 1974: Differential Topology. London, Prentice-Hall, 222 pp.

Haltiner, G. J., and R. T. Williams, 1980: Numerical prediction and dynamic meteorology. New York, Wiley, 477pp.

Harrold, T. W., and K. A. Browning, 1969: The polar low as a baroclinic disturbance. Quart. J. Roy. Meteor. Soc., 95, 710-723.

Hart, J. E., 1972: A laboratory study of baroclinic instability. Geophys. Fluid Dyn., 3, 181-209.

_____, 1979: Finite amplitude baroclinic instability. Ann. Rev. Fluid Mech., 11, 147-172.

_____, 1985: A laboratory study of baroclinic chaos on the f-plane. Tellus, 37A, 286-296.

Haynes, P. H., 1985: Nonlinear stability of a Rossby-wave critical layer. J. Fluid Mech., 161, 493-511.

Hide, R., 1953: Some experiments on thermal convection in a rotating liquid. Quart. J. Roy. Meteor. Soc., 79, 161.

_____, 1967: Theory of axisymmetric thermal convection in a rotating fluid annulus. Phys. Fluids, 10, 56-68.

_____, 1969: Some laboratory experiments on free thermal convection in a rotating fluid subject to a horizontal temperature gradient and their relation to the theory of the global atmospheric circulation. The Global Circulation of the Atmosphere, Ed. Corby, G. A., London, Roy. Meteor. Soc., 196-221.

_____, 1977: Experiments with rotating fluids. Quart. J. Roy. Meteor. Soc., 103, 1-28.

_____, and P. J. Mason, 1975: Sloping convection in a rotating fluid: a review. Advances in Phys., 24, 47-100.

_____, and P. J. Mason, 1978: On the transition between axisymmetric and non-axisymmetric flow in a rotating liquid annulus subject to a horizontal temperature gradient. Geophys. Astrophys. Fluid Dyn., 10, 121-156.

Hignett, P., 1985: Characteristics of amplitude vacillation in a differentially heated rotating annulus. Geophys. Astrophys. Fluid Dyn., 31, 247-281.

_____, A. A. White, R. D. Carter, W. D. N. Jackson and R. M. Small, 1985: A comparison of laboratory measurements and numerical simulations of baroclinic wave flows in a rotating cylindrical annulus. Quart. J. Roy. Meteor. Soc., 111, 131-154.

Hirsch, M. W., 1976: Differential Topology. Graduate texts in Mathematics, 33, New York, Springer-Verlag, 221 pp.

_____, and S. Smale, 1974: Differential Equations, Dynamical Systems and Linear Algebra. Academic Press, New York, 359 pp.

Holm, D. D., J. E. Marsden, T. Ratiu, and A. Weinstein, 1985: Nonlinear stability of fluid and plasma equilibria. Phys. Rep., 123, 1-116.

Holopainen, E. O., 1961: On the effect of friction in baroclinic waves. Tellus, 13, 363-367.

Hoskins, B. J., 1973: Stability of the Rossby-Haurwitz wave. Quart. J. Roy. Meteor. Soc., 99, 723-745.

_____, 1975: The geostrophic momentum approximation and the semi-geostrophic equations. J. Atmos. Sci., 32, 233-242.

_____, and A. J. Simmons, 1975: A multi-layer spectral model and the semi-implicit method. Quart. J. Roy. Meteor. Soc., 101, 637-655.

_____, M. E. McIntyre, and A. W. Robertson, 1985: On the use and significance of isentropic potential vorticity maps. Quart. J. Roy. Meteor. Soc., 111, 877-946.

Howard, L. N., 1964: The number of unstable modes in hydrodynamic stability problems. J. Mech., 3, 433-443.

James, I. N., 1987: Suppression of baroclinic instability in horizontally sheared flows. J. Atmos. Sci., 44, 3710-3720.

Jeffreys, H., 1926: On the dynamics of geostrophic winds. Quart. J. Roy. Meteor. Soc., 52, 85-104.

Jonas, P. R., 1980: Laboratory experiments and numerical calculations of baroclinic waves resulting from potential vorticity gradients at low Taylor number. Geophys. Astrophys. Fluid Dyn., 15, 297-315.

_____, 1981: Some effects of boundary conditions and fluid properties on vacillation in thermally driven rotating flow in an annulus. Geophys. Astrophys. Fluid Dyn., 18, 1-23.

Killworth, P. D., and M. E. McIntyre, 1985: Do Rossby-wave critical layers absorb, reflect, of over-reflect? J. Fluid Mech., 161, 449-492.

Kolmogorov, A. N., and S. V. Fomin, 1957: Elements of the Theory of Functions and Functional Analysis. Vol. 1: Metric and normed spaces. Transl. Boron, L. F., Rochester, New York, Graylock Press, 129 pp.

Kreyszig, E., 1979: Advanced Engineering Mathematics. New York, Wiley, 939 pp.

Lorenz, E. N., 1967: The nature and theory of the general circulation of the atmosphere. World Meteor. Organisation, 161 pp.

_____, 1972: Barotropic instability of Rossby wave motion. J. Atmos. Sci., 29, 258-264.

Mañé, R., 1981: On the dimension of the compact invariant sets of certain nonlinear maps. Dynamical Systems and Turbulence, Warwick 1980, Eds. Rand D., and L. S. Young, Lecture Notes in Mathematics, 898, New York, Springer-Verlag, 230-242.

Mansfield, D. A., 1974: Polar lows: the development of baroclinic disturbances in cold air outbreaks. Quart. J. Roy. Meteor. Soc., 100, 541-554.

Maslowe, S. A., 1981: Shear flow instabilities and transition. Hydrodynamic Instabilities and the Transition to Turbulence. Ed. Swinney, H. L, and J. P. Gollub, New York, Springer-Verlag, 181-228.

_____, 1986: Critical layers in shear flows. Ann. Rev. Fluid Mech., 18, 405-432.

Mason, P. J., 1975: Baroclinic waves in a container with sloping end walls. Phil. Trans. Roy. Soc. Lond., A 278, 397-445.

McIntyre, M. E., 1970: On the non-separable baroclinic parallel-flow instability problem. J. Fluid Mech., 40, 273 - 306.

_____, and T. N. Palmer, 1983: Breaking planetary waves in the stratosphere. Nature, 305, 593-600.

_____, and T. N. Palmer, 1984: The 'surf zone' in the stratosphere. J. Atmos. Terres. Phys., 46, 825-849.

_____, and T. N. Palmer, 1985: A note on the general concept of wave breaking for Rossby and gravity waves. Pure Appl. Geophys., 123, 964-975.

_____, and T. G. Shepherd, 1987: An exact local conservation theorem for finite amplitude disturbances to non-parallel shear flows, with remarks on Hamiltonian structure and on Arnol'd's stability theorems. J. Fluid Mech., 181, 527-565.

Palis, J., and W. de Melo, 1982: Geometric Theory of Dynamical Systems - An Introduction. Transl. Manning, A. K., New York, Springer-Verlag, 198 pp.

Pedlosky, J., 1970: Finite amplitude baroclinic waves. J. Atmos. Sci., 27, 178-200.

_____, 1982a: Geophysical Fluid Dynamics. New York, Springer-Verlag, 624 pp.

_____, 1982b: A simple model for nonlinear critical layers in an unstable baroclinic wave. J. Atmos. Sci., 39, 1229-2127.

Phillips, N. A., 1963: Geostrophic motion. Reviews of Geophyics, 1, 123-176.

Plumb, R. A., 1976: Wave interactions in a bounded fluid. Unpublished typescript.

_____, 1977: The stability of small amplitude Rossby waves in a channel. J. Fluid Mech., 80, 705-720.

Read, P. L., 1985: Finite amplitude, neutral baroclinic waves and mean flows in an internally heated rotating fluid. 1. Numerical simulations and quasi-geostrophic 'free modes'. Dyn. Atmos. Oceans, 9, 135-207.

Reed, R. J., 1987: Polar lows. The nature and prediction of extra-tropical weather systems. ECMWF Seminar Proceedings.

Roache, P. J., 1976: Computational fluid dynamics. Alberqueque, Hermosa Publ., 446 pp.

Ruelle, D., 1977: Sensitive dependence on initial condition and turbulent behavior of dynamical systems. Ann. NY Acad. Sci., 316, 408-416.

_____, and F. Takens, 1971: On the nature of turbulence. Comm. Math. Phys., 20, 167-192.

Simmons, A. J., and B. J. Hoskins, 1980: Barotropic influences on the growth and decay of nonlinear baroclinic waves. J. Atmos. Sci., 37, 1679-1684.

Sparrow, C., 1982: The Lorenz Equations, New York, Springer-Verlag, 269 pp.

Stuart, J. T., 1960: On the non-linear mechanics of wave disturbances in stable and unstable parallel flows. J. Fluid Mech., 9, 353-370.

Stewartson, K., 1978: The evolution of the critical layer of a Rossby wave. Geophys. Astrophys. Fluid Dyn., 9, 185-200.

Swinney, H. L., 1983: Observations of order and chaos in nonlinear systems. Physica, 7D, 3-15.

Takens, F., 1981: Detecting strange attractors in turbulence. Dynamical Systems and Turbulence, Warwick 1980, Eds. Rand, D., and L. S. Young, Lecture Notes in Mathematics, 898, New York, Springer-Verlag, 366-381.

Temam, R., 1988: Infinite Dimensional Dynamical Systems in Mechanics and Physics. Springer Series in Appl. Maths., 68, New York, Springer-Verlag, 615 pp.

Van Dyke, M., 1975: Perturbation methods in fluid mechanics. Stanford, Parabolic Press, 271 pp.

Warn, T. and H. Warn, 1978: The evolution of a nonlinear critical layer. Studies in Applied Maths., 59, 37-71.

Watson, J., 1960: On the non-linear mechanics of wave disturbances in stable and unstable parallel flows. Part 2. The development of a solution for plane Poiseuille flow and for plane Couette flow. J. Fluid Mech., 9, 371-389.

White, A. A., 1977: Modified quasi-geostrophic equations using geometric height as vertical coordinate. Quart. J. Roy. Meteor. Soc., 103, 383-396.

_____, 1986: Finite amplitude, steady Rossby waves and mean flows: analytical illustrations of the Charney-Drazin non-acceleration theorem. Quart. J. Roy. Meteor. Soc., 112, 749-773.

_____, 1987: Approximating the equations governing rotating fluid motion: a case study based on a quasi-geostrophic model. J. Fluid Mech., 182, 315-334.

_____, 1988: The dynamics of rotating fluids: numerical modelling of annulus flows. Meteor. Mag., 117, 54-63.

APPENDIX A

FREDHOLM'S SOLVABILITY CONDITION

We present here a brief statement of Fredholm's criterion for a linear partial differential equation to be soluble and the manipulations used to derive the solvability condition, (2.2.16), for (2.2.15).

Continuous functions, ϕ and ψ, with continuous derivatives defined on the interval $0 \leq y,z \leq 1$ may be viewed as the elements of a Hilbert space in which the scalar product $\langle \phi, \psi \rangle$ is defined by

$$\langle \phi, \psi \rangle \equiv \int_z \int_y \phi^* \psi \, dy \, dz . \tag{A.1}$$

Let L be a linear differential operator, such as

$$L\psi = \partial^2\psi/\partial y^2 - K^2\psi + \partial/\partial z\,(1/\beta\, \partial\psi/\partial z) + q_{yy}\psi/(u-C) \tag{A.2}$$

and F a linear operator which supplies appropriate homogeneous boundary conditions for L such as

$$\begin{aligned} F\psi &= \psi = 0 \quad \text{on} \quad y = 0,1 \\ F\psi &= (u-C)\,\partial\psi/\partial z - (\partial u/\partial z - ip^2\epsilon\chi)\psi = 0 \quad \text{on} \quad z = 0,1. \end{aligned} \tag{A.3}$$

Then the adjoint of L subject to F, which we call M subject to G, is defined implicitly by the requirement that

$$\langle \phi, L\psi \rangle = \langle M\phi, \psi \rangle \tag{A.4}$$

for all ψ satisfying the homogeneous boundary condtions (A.3) in F and all ϕ satisfying homogeneous boundary conditions in G. For a given function W in the Hilbert space. Fredholm's necessary and sufficient condition (Courant & Hilbert 1953) for there to be a solution ψ (belonging to the Hilbert space) of

$$L\psi = W, \tag{A.5}$$

subject to (A.3), is that W be orthogonal to all members of the kernel of the adjoint problem;

$$\langle \phi, W \rangle = 0 \quad \text{for all } \phi \text{ such that } M\phi = 0 \text{ and } G\phi = 0. \tag{A.6}$$

The adjoint problem to (A.2) and (A.3) is found by integrating $\langle \phi, L\psi \rangle$ by parts:

$$\langle \phi, L\psi \rangle = \int_z \int_y \phi^* \{ \partial^2\psi/\partial y^2 - k^2\psi + \partial/\partial z(1/B\, \partial\psi/\partial z) + \frac{q_y \psi}{u-c} \} \, dy\, dz$$

$$= \int_z \int_y \{ \partial^2\phi/\partial y^2 - k^2\phi + \partial/\partial z(1/B\, \partial\phi/\partial z) + q_y\phi/(u-c^*) \}^* \psi \, dy\, dz$$

$$+ \int_z [\phi^* \partial\psi/\partial y]_0^1 \, dz + \int_y \left[\{ (\partial u/\partial z + ik^2 \epsilon X)\phi - (u-c^*)\partial\phi/\partial z \}^* \frac{\psi}{B(u-c)} \right]_0^1 dy$$

from which it is clear that the adjoint problem is

$$M\phi = \partial^2\phi/\partial y^2 - k^2\phi + \partial/\partial z(1/B\, \partial\phi/\partial z) + q_y\phi/(u-c^*) \tag{A.7}$$

$$\begin{aligned} &\phi = 0 \quad \text{on} \quad y = 0, 1 \\ &G\phi = (u-c^*)\partial\phi/\partial z - (\partial u/\partial z + ik^2\epsilon X)\phi = 0 \quad \text{on } z = 0, 1. \end{aligned} \tag{A.8}$$

It is readily verified that ϕ_0 is a member of the kernel of M subject to G (i.e. $M\,\phi_0 = 0$) if and only if $\psi = \phi_0^*$ is a member of the kernel of L subject to F.

Equation (2.2.15) is not of the form (A.5) because $\psi = \delta\phi$ in (2.2.15) does not satisfy the homogeneous boundary condition (A.3). To find the condition for there to be a solution of

$$L\rho = W \; ; \quad \rho = 0 \text{ on } y = 0, 1 \; ; \quad F\rho = V \text{ on } z = 0, 1 \tag{A.9}$$

it is useful to let g be some element in the Hilbert space for which

$$g = 0 \text{ on } y = 0, 1 \quad ; \quad Fg = V \text{ on } z = 0, 1. \tag{A.10}$$

Then $\psi \equiv \rho - g$ satisfies (A.3), and (A.9) has a solution if and only if

$$\int_z \int_y \psi_0 (W - Lg) \, dy \, dz = 0 \quad \text{for each } \psi_0 \text{ such that } L\psi_0 = 0 \;\&\; F\psi_0 = 0. \tag{A.11}$$

As noted by McIntyre (1970), $< \psi_0^*, \; Lg >$ may be calculated by integration by parts

$$\int_z \int_y \psi_0 Lg \, dy \, dz = \int_z \int_y (L\psi_0) g \, dy \, dz + \int_y \left[(\psi_0 \partial g/\partial z - \partial \psi_0/\partial z \, g)/B \right]_0^1 dy$$

$$= \int_y \left[\psi_0 V / B(u-C) \right]_0^1 dy. \tag{A.12}$$

Hence (A.9) has a solution if and only if

$$\int_z \int_y \psi_0 W \, dy \, dz - \int_y \left[\psi_0 V / B(u-C) \right]_0^1 dy = 0$$

$$\text{for each } \psi_0 \text{ such that } L\psi_0 = 0 \text{ and } F\psi_0 = 0. \tag{A.13}$$

APPENDIX B

VERTICAL INTEGRATION COEFFICIENTS

This appendix summarises the method used to find the spectral coefficients of $R \equiv \partial q/\partial y\ \phi_n / (u-c)$ from the coefficients Φ_n^m in the calculations of the growth rates of normal modes on laterally sheared flows described in sections 2.6 and 2.7. The method assumes that u is symmetric about mid-channel (see (2.6.6) and the last paragraph of example 1 in section 2.7).

Any flow which is even about mid-channel has fields of $\partial q/\partial y$ and

$$S \equiv \partial q/\partial y/(u-c) = \partial q/\partial y\ \frac{(u-c_r+ic_i)}{(u-c_r)^2+c_i^2} \tag{B.1}$$

which are also even about mid-channel. $S_n(y)$ (i.e. S at level n) can be expanded as a half range cosine expansion on the interval $1/2 \le y \le 3/2$

$$S_n(y) = \frac{1}{\sqrt{2P}} \sum_{p=0}^{P} a_n^p \cos p\pi(y-1/2). \tag{B.2}$$

The P used here takes the same value in the code as the P in (2.6.14). The well known formulae for the a_n^p are

$$\begin{aligned} a_n^0 &= 1/2\sqrt{2P} \int_{-1/2}^{+3/2} S_n(y-1/2)\,dy \\ a_n^p &= \sqrt{2P} \int_{-1/2}^{+3/2} S_n(y-1/2) \cos p\pi(y-1/2)\,dy, \quad 1 \le p \le P. \end{aligned} \tag{B.3}$$

These may be reduced to standard FFT form by using a variable η with $\eta = 0$ at $y = 1/2$ and $\eta = 1$ at $y = 3/2$

$$2\eta = y + 1/2 \quad ; \quad S_n'(\eta) = S_n(y-1/2).$$

Representing S_n' at 2P points on the interval $0 \le \eta \le 1$, and writing $S_n''(j) = S_n'(j\ /\ 2P)$ the FFT forms of (B.3) are

$$\begin{aligned} a_n^0 &= 1/\sqrt{2P} \sum_{j=0}^{2P-1} S_n''(j) \\ a_n^p &= 2(-1)^p/\sqrt{2P} \sum_{j=0}^{2P-1} S_n''(j) \cos\{2p\pi j/2P\}, \quad 1 \le p \le P. \end{aligned} \tag{B.4}$$

To calculate the half range transform coefficients of $R_n = S_n \cdot \phi_n$ it is convenient to re-express (B.2) as

$$S_n(y) = 1/\sqrt{2P}\left\{\sum_{p=0}^{[P/2]} (-1)^p a_n^{2p} \cos 2p\pi y + \sum_{p=0}^{[(P-1)/2]} (-1)^p a_n^{2p+1} \sin(2p+1)\pi y\right. \tag{B.5}$$

where [A] is the largest integer smaller than or equal to A. The coefficient R_n^q of the lateral spectral expansion of R_n is given by

$$R_n^q = 2 \int_0^1 \sin q\pi y \; S_n(y) \sum_{m=1}^{M} \Phi_n^m \sin m\pi y \, dy . \tag{B.6}$$

Tedious manipulations show that when q,m and p are positive

$$\begin{aligned} I_{qpm} &\equiv 2 \int_0^1 \sin q\pi y \cos 2p\pi y \sin m\pi y \, dy \\ &= 1/2 \{ \delta(m+2p-q) + \delta(m-2p-q) - \delta(m-2p+q)\} \end{aligned} \tag{B.7}$$

$$\begin{aligned} J_{qpm} &\equiv 2 \int_0^1 \sin q\pi y \sin(2p+1)\pi y \sin m\pi y \, dy \\ &= -\frac{8\, q\, (2p+1)\, m}{\pi (q^2-\alpha^2)(q^2-\beta^2)} \qquad \text{if } q+m \text{ is even} \\ &= 0 \qquad \text{if } q+m \text{ is odd} \end{aligned} \tag{B.8}$$

where $\alpha = 2p+1+m$, $\beta = 2p+1-m$.

So the integration formula for laterally sheared flows is

$$\Phi_{n+1}^q = 2\Phi_n^q - \Phi_{n-1}^q + B/N^2 \{ (K^2 + \pi^2 q^2)\Phi_n^q - R_n^q\} \tag{B.9}$$

in which R_n^q is given by

$$R_n^q = \sum_{m=1}^{M} \left\{ \sum_{p=0}^{[P/2]} (-1)^p a_{2p} \Phi_n^m I_{qpm} + \sum_{p=0}^{[(P-1)/2]} (-1)^p a_{2p+1} \Phi_n^m J_{qpm} \right\} \tag{B.10}$$

APPENDIX C

METRIC SETS, MAPS AND MANIFOLDS

This appendix defines the terms and states the theorems used in chapter 5. The definitions are grouped into three sets under the titles; sets in metric spaces, maps and manifolds, and stable mappings and intersections. A reader completely unfamiliar with the concepts behind these definitions will find them developed (and the proofs of all the theorems) in §§1 - 19, §21 and §24 of Kolmogorov and Fomin (1957) and chapter 1 and §§1 - 3 of chapter 2 of Guillemin and Pollack (1974). The definitions given here follow the conventions of those authors closely. Dieudonné (1960) chs. 1 - 7 and Abraham et. al. (1983) chs. 1 - 3 provide alternative accounts of the material.

Sets in metric spaces

1. Metric spaces

A metric space R is a set T of elements, $\underset{\sim}{x}$, called points, together with a distance function (i.e. a single-valued non-negative real function) $\rho(\underset{\sim}{x},\underset{\sim}{y})$, which for all points $\underset{\sim}{x}$ and $\underset{\sim}{y}$ in T satisfies three conditions:

a) $\rho(\underset{\sim}{x},\underset{\sim}{y}) = 0$ if and only if $\underset{\sim}{x} = \underset{\sim}{y}$.

b) $\rho(\underset{\sim}{x},\underset{\sim}{y}) = \rho(\underset{\sim}{y},\underset{\sim}{x})$

c) $\rho(\underset{\sim}{x},\underset{\sim}{y}) + \rho(\underset{\sim}{y},\underset{\sim}{z}) \geq \rho(\underset{\sim}{x},\underset{\sim}{z})$

An example of a metric space is Euclidean n - space $\mathbb{R}^n$ for which T is the set of ordered n-tuples of real numbers $\underset{\sim}{x} = (x_1, x_2, \ldots, x_n)$ and the distance function is

$$\rho(\underset{\sim}{x}, \underset{\sim}{y}) = \sum_{k=1}^{n} (x_k - y_k)^2$$

2. Open sets

An open ε neighbourhood of the point $\underset{\sim}{x}_0$ (in R) is an open sphere of radius ε and centre $\underset{\sim}{x}_0$ (i.e. it consists of all points $\underset{\sim}{x}$ which satisfy $\rho(\underset{\sim}{x},\underset{\sim}{x}_0) < \varepsilon$, $\varepsilon > 0$). A set P is defined to be open if each of its points has an open neighbourhood contained entirely within P.

3. Closed and dense sets

A point $\underset{\sim}{x}$ (in R) is called a contact point of a set P if every neighbourhood of $\underset{\sim}{x}$ contains at least one point of P. The set of all contact points of P is called the closure of P. P is defined to be closed if it coincides with its closure. If the closure of P coincides with the whole space then P is said to be dense in R.

4. Generic sets

If P and Q are open and dense in R then so is $P \cap Q$. To see this note that all and only sets which are dense in R have non zero intersection with every non-empty open set in R. Now if P and Q are open and dense subsets of R and S is any non-empty open set then $P \cap S$ is a non-empty open set and so is $Q \cap (P \cap S) = (Q \cap P) \cap S$. Hence $Q \cap P$ is dense (and of course open) in R. If P is the intersection of a countable number of sets which are both open and dense in R, it is said to be generic in R.

5. Compact sets

A set P in R is compact if every infinite sequence of elements in P contains a subsequence which converges to some $\underset{\sim}{x}$ in R. It can be shown that any subset of Euclidean n-space (n being finite) which can be enclosed within a sufficiently large cube (whose edges are of finite length) is compact. [Some subsets of infinite dimensional spaces are compact. In fact a subset P of a complete metric space R is compact if and only if P is totally bounded (§13 and §16 of Kolmogorov and Fomin).] Furthermore any collection of open sets $\{U_\alpha\}$ which covers a compact set in a metric space contains a finite sub-collection which covers the compact set (Kolmogorov & Fomin §18).

Maps and manifolds

1. Maps

A mapping $\underset{\sim}{f}$ which assigns to each element in a set P an element in a set Q will be denoted by $\underset{\sim}{f} : P \to Q$. A map from an open set U in $\mathbb{R}^n$ into $\mathbb{R}^m$ is said to be C^r smooth if it has continuous partial

derivatives at all orders $\leq r$. For a point $\underset{\sim}{q}$ in Q the pre-image of $\underset{\sim}{q}$, $f^{-1}(\underset{\sim}{q})$, is the set of all points $\underset{\sim}{p}$ in P for which $\underset{\sim}{f}(\underset{\sim}{p}) = \underset{\sim}{q}$.

2. Diffeomorphisms

$\underset{\sim}{f} : P \to Q$ is said to be one-one (or injective) if no more than one element in P is mapped by $\underset{\sim}{f}$ onto any one element of Q. On the other hand $\underset{\sim}{f} : P \to Q$ is onto (or surjective) if it assigns to each element of Q at least one element of P. A C^r smooth map $\underset{\sim}{f} : P \to Q$ which is both one-one and onto and whose inverse $\underset{\sim}{f}^{-1} : Q \to P$ is C^r smooth is a C^r diffeomorpism.

3. Manifolds

Suppose that X is a subset of a large "ambient" Euclidean space [or some Banach space]. X is a k-dimensional manifold if each point $\underset{\sim}{x}$ in X has a neighbourhood V within X which is diffeomorphic to an open set U of $\mathbb{R}^k$. (This definition is somewhat simpler than that adopted by most authors eg Hirsch (1976), Arnol'd (1978).) The letters X and Y are reserved to denote manifolds in this appendix.

4. Derivative maps, tangent spaces and tangent bundles

Let U be an open set of $\mathbb{R}^k$ and W a subset of $\mathbb{R}^m$ and consider a map $\underset{\sim}{f} : U \to W$. Take any vector $\underset{\sim}{h}$ in $\mathbb{R}^k$ and let t belong to $\mathbb{R}^1$. The derivative of $\underset{\sim}{f}$ in the direction $\underset{\sim}{h}$, taken at the point $\underset{\sim}{x}$ within U, is defined to be

$$d\underset{\sim}{f}(\underset{\sim}{h}) = \lim_{t \to 0} \{ (\underset{\sim}{f}(\underset{\sim}{x} + t\underset{\sim}{h}) - f(\underset{\sim}{x})) / t \}$$

Consider a neighbourhood V of a point $\underset{\sim}{x}$ within a k-dimensional manifold X (which itself lies in $\mathbb{R}^N$) and a (i.e. any) diffeomorphism $\underset{\sim}{\phi} : U \to V$, U being an open set in $\mathbb{R}^k$. Assume for convenience that $\underset{\sim}{\phi}(\underset{\sim}{0}) = \underset{\sim}{x}$. The best linear approximation to $\underset{\sim}{\phi} : U \to V$ at $\underset{\sim}{0}$ is

$$\underset{\sim}{u} \to \underset{\sim}{\phi}(\underset{\sim}{0}) + d\underset{\sim}{\phi}_0(\underset{\sim}{u}) = \underset{\sim}{x} + d\underset{\sim}{\phi}_0(\underset{\sim}{u}).$$

The tangent space of X at $\underset{\sim}{x}$ is defined to be the image of the map $d\underset{\sim}{\phi}_0 : \mathbb{R}^k \to \mathbb{R}^N$. It has dimension k and is denoted by $T_x(X)$. The

tangent bundle T(X) of a manifold X in $\mathbb{R}^N$ is defined by

$$T(X) = \{ (\underset{\sim}{x}, \underset{\sim}{v}) \text{ in } X \times \mathbb{R}^N ; \ \underset{\sim}{v} \in T_x(X) \}.$$

It may be shown (Guillemin & Pollack pp 50-51) that the tangent bundle of a manifold is a manifold and that dim T(X) = 2 dim X.

5. Open sets of functions

Let U and V be open sets in $\mathbb{R}^k$ and $\mathbb{R}^m$ respectively, and T^r be the set of all C^r smooth functions f : U → V. A natural definition of an open ε neighbourhood of f_0 in T^r is the set of functions f for which

$$\| d^K \underset{\sim}{f}(\underset{\sim}{x}) - d^K \underset{\sim}{f}_0(\underset{\sim}{x}) \| < \epsilon \quad \text{for all } \underset{\sim}{x} \text{ in } U \text{ and } 0 \leq K \leq r,$$

(||A|| denotes the norm of A); f and f_0 are close only if $f(\underset{\sim}{x})$ and $f_0(\underset{\sim}{x})$ and their first r derivatives are close at all $\underset{\sim}{x}$ in U. A similar definition of open neighbourhoods is easily given for maps on compact manifolds.

Stable maps and intersections

1. Immersions and Submersions

Let X and Y be manifolds with dim X ≤ dim Y and $\underset{\sim}{f}$: X → Y. If $d\underset{\sim}{f}_x : T_x(X) \to T_y(Y)$ is injective (i.e. one-one) $\underset{\sim}{f}$ is said to be immersive at $\underset{\sim}{x}$. If $\underset{\sim}{f}$ is immersive at all points $\underset{\sim}{x}$ in X it is called an immersion.

Suppose instead that dim X ≥ dim Y. If $d\underset{\sim}{f}_x : T_x(X) \to T_y(Y)$ is surjective (i.e. onto) $\underset{\sim}{f}$ is submersive at $\underset{\sim}{x}$. If $\underset{\sim}{f}$ is submersive at all points $\underset{\sim}{x}$ in X it is a submersion.

2. Embeddings

A map $\underset{\sim}{f}$: P → Q is said to be proper if the pre-image of any compact set in Q is compact in P. Any map from a compact manifold is obviously proper. An embedding is an immersion that is injective and proper. The main theorem for embeddings is that any embedding $\underset{\sim}{f}$: X → Y maps X diffeomorphically onto its image in Y. [The

definition of embeddings varies between authors but because of this theorem all definitions are equivalent.]

3. Transversal intersections and codimension

Consider two manifolds X and Y, a third manifold Z lying within Y (i.e. a submanifold of Y), and a smooth map $\underset{\sim}{f} : X \to Y$. $\underset{\sim}{f}$ and Z are said to be transversal if at any point $\underset{\sim}{x}$ in X whose image, $\underset{\sim}{y}$, under $\underset{\sim}{f}$, lies within Z, the image in $T_y(Y)$ of the tangent map of $\underset{\sim}{f}$ at $\underset{\sim}{x}$ (i.e. $d\underset{\sim}{f}_{\underset{\sim}{x}} : T_x(X) \to T_y(Y)$) and the tangent space to Z at $\underset{\sim}{y}$, $T_y(Z)$, together span $T_y(Y)$.

Two manifolds X and Z lying within Y are said to intersect transversally if the identity map $\underset{\sim}{i} : X \to Y$ ($\underset{\sim}{i}(\underset{\sim}{x}) = \underset{\sim}{x}$) and Z are transversal. A manifold X within Y is said to have codimension p within Y if at each point $\underset{\sim}{x}$ in X the smallest number of vectors required in addition to those in $T_x(X)$ to span $T_x(Y)$ is p.

4. Homotopies and stability

Let $\underset{\sim}{f}_0 : X \to Y$ and $\underset{\sim}{f}_1 : X \to Y$ be smooth maps and I denote the closed unit interval [0,1] in $\mathbb{R}^1$. $\underset{\sim}{f}_0$ and $\underset{\sim}{f}_1$ are said to be homotopic if there is a smooth map $\underset{\sim}{F} : X \times I \to Y$ with $\underset{\sim}{F}(\underset{\sim}{x},0) = \underset{\sim}{f}_0(\underset{\sim}{x})$ and $\underset{\sim}{F}(\underset{\sim}{x},1) = \underset{\sim}{f}_1(\underset{\sim}{x})$.

A property of a map is said to be stable if whenever $\underset{\sim}{f}_0 : X \to Y$ possesses the property and $\underset{\sim}{f}_1 : X \to Y$ is a (i.e. any) homotopy of $\underset{\sim}{f}_0$ then for some $\varepsilon > 0$ each $\underset{\sim}{f}_t$ with $0 \leq t < \varepsilon$ also possesses the property.

Theorems

Theorem C.1 : Theorem of common zeros

If the map $\underset{\sim}{g}$

$$\underset{\sim}{g} = (g_1, g_2, \cdots g_k) \quad ; \quad \underset{\sim}{g} : X \to \mathbb{R}^k$$

of a manifold X into k dimensional Euclidean space is submersive at each point $\underset{\sim}{x}$ in X where $\underset{\sim}{g}(\underset{\sim}{x}) = (0,0,\ldots\ 0)$ then the set Z of common zeros (i.e. $\underset{\sim}{g}^{-1}(\underset{\sim}{0})$) is a submanifold of X of codimension k.

Theorem C.2 : Homotopy stability of transversal intersections

The property of being transversal to any specified closed submanifold Z of Y is stable for C^r ($r \geq 1$) smooth maps $\underset{\sim}{f}$: X → Y of a compact manifold X into a manifold Y.

Theorem C.3 : Whitney embedding theorem

Every k dimensional manifold admits a C^r ($r \geq 1$) one-one immersion in $\mathbb{R}^{2k+1}$.

Theorem C.4 : Transversality theorem

Let

i) X, S and Y be manifolds, X x S be the Cartesian product of X and S and Z be a submanifold of Y

ii) $\underset{\sim}{F}$: X x S → Y be a smooth map

iii) $\underset{\sim}{f}_s$: X → Y be defined by $\underset{\sim}{f}_s(\underset{\sim}{x}) = \underset{\sim}{F}(\underset{\sim}{x},\underset{\sim}{s})$ for all points $\underset{\sim}{s}$ in S and $\underset{\sim}{x}$ in X.

If $\underset{\sim}{F}$ is transversal to Z then, for almost every $\underset{\sim}{s}$ in S, $\underset{\sim}{f}_s$ is transversal to Z. (Almost every $\underset{\sim}{s}$ in S means for all points in S other than some set of measure zero.)

Theorem C.5 : Local immersion theorem

If X is a k dimensional manifold, $\underset{\sim}{f}$: X → Y an immersion at $\underset{\sim}{x}$, and $\underset{\sim}{y} = \underset{\sim}{f}(\underset{\sim}{x})$, then there is a set of local coordinates around $\underset{\sim}{x}$ and $\underset{\sim}{y}$ in which $\underset{\sim}{f}(x_1, \ldots, x_k) = (x_1, \ldots x_k, 0, .. 0)$.