Part III Mathematics
Large-scale Atmosphere-Ocean Dynamics
Michaelmas Term 2003

lecturer: Peter Haynes (phh@damtp.cam.ac.uk)

10 January 2004

1 Introduction

1.1 Scientific background

This course will discuss fluid dynamics with combined ingredients of rotation and density stratification. This is basis for understanding fluid flow in atmospheres and oceans (Earth and elsewhere) and is also relevant to interior motion of planets and stars.

It almost goes without saying that this subject is vital part of understanding and quantitative modelling of the dynamical, chemical and physical processes governing the atmosphere and the ocean. This understanding and modelling capability has already resulted in good weather forecasting in the short and medium term. Future challenges are prediction on seasonal and longer time scales, including more reliable prediction of anthropogenic effects on climate and coupled biological and chemical processes.

For reviews of recent progress and current questions see:

Meteorology at the Millennium, ed. R. P. Pearce; London, Academic Press and Royal Meteorol. Soc., 330 pp. (2002).

'Nature Insights' on *Climate and Water*, Nature, 419, 6903 (also available at
http://www.nature.com/nature/insights/6903.html).

Three topics that serve as good motivation are:

(i). **Stratospheric ozone depletion**: Mild reductions in stratospheric ozone were predicted to result from emissions of long-lived halogen compounds (CFCs etc). In late 1970s scientists at British Antarctic Survey discovered significant ozone reduction in Southern Hemisphere in late winter in spring (the 'Antarctic ozone hole'). Ozone levels in SH late winter and spring reduced through 1980s and in 1990s. There was some evidence for Arctic depletion, but much smaller and much more variable from year to year. In 2002 the Antarctic circulation was very different from usual and the ozone depletion was much less than in previous years. 2003 shows large depletion ('the 2nd largest ozone hole ever observed'). Questions: Why strong ozone depletion in southern hemisphere, not northern hemisphere? Why was 2002 so different to other recent years? Why will ozone levels take 50-100 years to return to pre-industrial levels?

(ii). **Extratropical effects of El Nino**: El Nino is a change in the sea surface temperature distribution in the equatorial Pacific that occurs every few years. It results in changes in patterns of tropical convection (large-scale thunderstorm complexes etc) in the Pacific, but also in (simultaneous) changes in weather patterns elsewhere in tropics (e.g. Indian Ocean) and in extratropics (e.g. North America, Australia). There is evidence that El Nino events may case changes in the ocean temperatures and circulation in the North Pacific that are felt up to a decade later. Questions: Why are some remote effects of El Nino felt instantaneously and others many years later? Why does El Nino occur in the equatorial Pacific but not the equatorial Atlantic? What determines the frequency of El Nino events?

(iii). **Quasi-biennial oscillation (QBO)**: The QBO is a regular oscillation in the winds in the tropical stratosphere, with period about 28 months. It is arguably the most predictable atmospheric phenomenon apart from the annual cycle. The QBO was discovered only about 40 years ago and the mechanism explained only about 30 years ago. Climate models have begun to simulate a realistic QBO only in the last 5 years or so. But a simple laboratory experiment can be used to demonstrate the essential mechanism. Questions: Why is the primary QBO confined to the tropics? Why has the QBO been so difficult for climate models to simulate?

1.2 Basic equations and boundary conditions

[Based on notes prepared for NERC GEFD Summer School by M.E.McIntyre.]

The book *An Introduction to Fluid Dynamics* by G.K. Batchelor includes a very thorough discussion of the governing equations. But in this course we will essentially note the form of the equations and move on to solving them.

(The Eulerian description is used, so that the material derivative $D/Dt = \partial/\partial t + \mathbf{u}\cdot\nabla$.)

Mass conservation, or "continuity":

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\mathbf{u}) = 0 \quad \text{or} \quad \frac{D\rho}{Dt} + \rho\nabla.\mathbf{u} = 0 \tag{1.1}$$

where ρ is density and $\mathbf{u}$ is velocity.

For incompressible flow:

$$\nabla\cdot\mathbf{u} = 0 \quad . \tag{1.2}$$

associated boundary condition:

$$\mathbf{u}\cdot\mathbf{n} = (\mathbf{u}\cdot\mathbf{n})_{\text{boundary}}$$

($\mathbf{n}$ is a unit vector perpendicular to boundary; right-hand side (hereafter RHS) = motion of boundary normal to itself *or* ρ^{-1} times rate at which mass crosses a permeable boundary, e.g. in dye experiments);

$$\mathbf{u}\cdot\mathbf{n} = 0 \text{ at a stationary rigid boundary}$$

Newton's second law: Acceleration = force/mass.

In an inertial frame,

$$\text{LHS} = \frac{D\mathbf{u}}{Dt} = \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u}\cdot\nabla\mathbf{u} = \frac{\partial \mathbf{u}}{\partial t} + \boldsymbol{\zeta}\times\mathbf{u} + \frac{1}{2}\nabla(|\mathbf{u}|^2) \qquad (1.3)$$

where $\boldsymbol{\zeta} = \nabla\times\mathbf{u} =$ vorticity.

$$\text{RHS} = -\frac{1}{\rho}\nabla p + \mathbf{g} + \mathbf{F} \quad , \qquad (1.4)$$

where p is total pressure, including hydrostatic, and $\mathbf{g}$ and $\mathbf{F}$ are, respectively, the gravitational and frictional forces per unit mass. For spatially uniform dynamical viscosity μ (see e.g. Batchelor's textbook for more general cases),

$$\mathbf{F} = \frac{\mu}{\rho}\nabla^2\mathbf{u} \ .$$

associated boundary conditions: define $\mathbf{u}_\parallel = \mathbf{u} - (\mathbf{u}\cdot\mathbf{n})\mathbf{n}$, tangential velocity. First,

$$\mathbf{u}_\parallel = (\mathbf{u}_\parallel)_{\text{boundary}} \qquad \text{(no-slip condition)}$$

if boundary is solid with prescribed tangential velocity. Second, if the tangential stress (friction force $\boldsymbol{\tau}$ per area) is prescribed at an otherwise rigid *plane* boundary (sometimes useful in thought-experiments) then

$$\mu\frac{\partial \mathbf{u}_\parallel}{\partial n} = \boldsymbol{\tau}$$

In more general cases (curved boundaries, free boundaries, etc.) it is necessary to deal with the full expression for total stress due to viscosity *and* pressure; again see Batchelor's textbook for a clear discussion; Cartesian tensors are required).

Rotating reference frames: To RHS of Newton's second law add:

$$-2\boldsymbol{\Omega}\times\mathbf{u} - \boldsymbol{\Omega}\times(\boldsymbol{\Omega}\times\mathbf{r}) \qquad (1.5)$$

where $\boldsymbol{\Omega}$ is the angular velocity of the reference frame, assumed constant, $\mathbf{r}$ is position relative to any point on the rotation axis, and $\mathbf{u}$ is now velocity relative to the rotating frame. The second and third terms are the Coriolis and centrifugal forces per unit mass, 'fictitious forces' felt by an "observer on a merry-go-round". It is convenient, and conventional, to recognize that both $\mathbf{g}$ and the centrifugal force $-\mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{r})$ are gradients of potentials and that there therefore exists an "effective gravitational potential"

$$\Phi = \Phi_0 - \tfrac{1}{2}|\mathbf{\Omega}|^2 r_\perp^2$$

such that $\mathbf{g}-\mathbf{\Omega}\times(\mathbf{\Omega}\times\mathbf{r}) = -\nabla\Phi$, where Φ_0 is the gravitational potential in the ordinary sense, and $r_\perp$ is perpendicular distance to the axis of rotation. (On the rotating Earth, the level surfaces Φ = constant and Φ_0 = constant are only slightly different; the Φ and Φ_0 surfaces tangent to each other at the north pole are about 11 km apart at the equator.)

Newton's second law can thus be written

$$D\mathbf{u}/Dt + 2\mathbf{\Omega}\times\mathbf{u} = -\frac{1}{\rho}\nabla p - \nabla\Phi + \mathbf{F} \qquad (1.6)$$

It is traditional to put the Coriolis force per mass on the left, even though in the rotating frame it has the role of force/mass, rather than acceleration.

Equation of state: In general we still have one more unknown than the number of equations and need to introduce an equation of state relating pressure and density, typically of the form

$$p = p(\rho, T, \dots), \qquad (1.7)$$

where T is temperature, and the dots indicate that the density may well depend on other physical quantities such as salinity. (If it does then each extra quantity requires an extra equation.)

Thermodynamic equation: the temperature T satisfies the equation

$$\frac{DT}{Dt} = \mathcal{H}_{adiabatic} + \mathcal{H}_{diabatic} \qquad (1.8)$$

where the terms on the RHS represent, respectively, heating associated with reversible (adiabatic) and irreversible (diabatic) processes (see Batchelor - p155-156 for more details).

For compressible flow (1.1) plus (1.6) (3 components) plus (1.7) plus (1.8) form a complete set specifying the evolution of ρ, $\mathbf{u}$ (3 components), T and p. (Note that there are 5 time derivatives.)

For incompressible flow (1.1) plus (1.2) plus (1.6) (3 components) form a complete set specifying the evolution of ρ, $\mathbf{u}$ (3 components) and p. (Note that there are 4 time derivatives.)

In fluid dynamics we are no longer seeking the governing equations, though some fundamental questions remain. We have to find ways of understanding the rich behaviour allowed by these equations.

2 Density stratification

Density stratification is important ingredient of atmospheric and oceanic flows and the effects of density stratification alone could fill a lecture course. (See Environmental Fluid Dynamics course.)

Here we shall be concerned with the combined effects of density stratification and rotation, so we now discuss the basics of density stratification and how to include it in the equations of motion, but will not go into details.

2.1 Hydrostatic equilibrium

In a state of rest there is a balance on the right-hand side of the momentum equations between the pressure gradiant and the imposed force. If the latter is purely gravitational then we have the balance

$$-\nabla p_{\mathrm{e}} + \rho_e \mathbf{g} = 0.$$

This equation expresses *hydrostatic balance.*

We will assume that gravity acts in the vertical and will take the z axis to be the upward vertical. Hence $\mathbf{g} = (0, 0, -g)$ where g is the gravitational acceleration.

Note that there can be equilibrium only if the density depends on height alone (otherwise the vertical gradient of pressure and hence the pressure would vary in the horizontal and the horizontal pressure gradient could not be balanced by gravity).

2.2 Boussinesq approximation

This approximation is valid in the asymptotic limit $\Delta\rho/\rho \ll 1$, $g' = g\Delta\rho/\rho$ finite, all height scales $\ll c_{\mathrm{sound}}^2/g$.

The fluid is assumed to be incompressible, so (1.2) holds and hence from (1.1) density is conserved following fluid particles.

$$\frac{D\rho}{Dt} = 0$$

The density ρ is split into two parts ρ_0 and ρ' with ρ_0 constant and $\rho'/\rho_0 \ll 1$. The pressure is then split into two parts, $p_0(z)$ such that

$$-\frac{dp_0}{dz} - \rho_0 g = 0$$

and remainder p'.

Then the vertical component of the momentum equation is written as follows

$$\frac{Dw}{Dt} = -\frac{1}{\rho_0+\rho'}\frac{\partial p_0}{\partial z} - \frac{1}{\rho_0+\rho'}\frac{\partial p'}{\partial z} - g \tag{2.1}$$

$$= -\frac{1}{\rho_0}\frac{\partial p_0}{\partial z} + \frac{\partial p_0}{\partial z}\left\{\frac{1}{\rho_0} - \frac{1}{\rho_0+\rho'}\right\} - \frac{\partial p'}{\partial z}\frac{1}{(\rho_0+\rho')} - g \tag{2.2}$$

$$\simeq -\frac{1}{\rho_0}\frac{\partial p'}{\partial z} + \frac{\rho'}{\rho_0}g \tag{2.3}$$

neglecting terms of $\mathcal{O}(\rho'/\rho_0)$.

In the horizontal components of the momentum equations $1/\rho$ is is replaced by $1/\rho_0$.

To summarise, under the Boussinesq approximation, the density ρ is replace by the constant value ρ_0, except where it is multiplied by the gravitational acceleration g.

It is clear from the Boussinesq equations that if a fluid parcel is displaced into surroundings that are heavier, it feels an upward force and if it is displaced into surroundings that are lighter, it feels a downward force. Thus if density increases downward then a displaced fluid parcel will tend to return to its original location. Vertical motion is inhibited and the fluid is said to be *statically stable*.

A measure of the stability is the *buoyancy frequency* or *Brunt-Vaisala frequency* N defined, in the Boussinesq system, by

$$N^2 = -\frac{g}{\rho_0}\frac{d\rho'}{dz}. \tag{2.4}$$

This is the frequency of oscillation of vertically aligned slabs of fluid which are displaced vertically relative to their neighbours.

The Boussinesq model might seem a reasonable approximation for the ocean (though in fact because pressure differences are large, the pressure dependence of density needs to be taken into account if considering the whole water column). For the atmosphere compressibility is much more important and the Boussinesq approximation cannot be justified for most quantitative calculations.

2.3 Perfect gas

Now consider effects of compressibility. For a perfect gas equation (1.8) becomes

$$\frac{DT}{Dt} = \frac{\kappa T}{p}\frac{Dp}{Dt} + \mathcal{H}_{diab} \tag{2.5}$$

where κ is ratio of specific heats, or, equivalently,

$$\frac{D\theta}{Dt} = (p/p_*)^{-\kappa}\mathcal{H}_{diab} \tag{2.6}$$

defining $\theta = T(p/p_*)^{-\kappa}$, where p_* is a constant pressure. (It may be convenient to substitute θ for T in the equation of state.) The 'potential temperature' θ is the relevant measure of buoyancy in a stratified compressible fluid.

The corresponding definition of the buoyancy frequency is

$$N^2 = \frac{g}{\theta}\frac{d\theta}{dz} = g\left(\frac{1}{T}\frac{dT}{dz} - \frac{\kappa}{p}\frac{dp}{dz}\right). \tag{2.7}$$

2.4 Vertical stratification in the atmosphere and ocean

Note that in the lower part of the atmosphere (the *troposphere*) up to about 10km the temperature T decreases upwards. However the atmosphere is statically stable (at least on the large scale), since the potential temperature θ increases upwards.

For the troposphere $N^2 \simeq 10^{-4}\mathrm{s}^{-2}$.

From about 10km to about 50km temperature increases with height (the *stratosphere*) and the static stability is greater than in the troposphere.

For the stratosphere $N^2 \simeq 4 \times 10^{-4}\mathrm{s}^{-2}$.

In the upper ocean (the *thermocline*) the buoyancy frequency is similar to that in the troposphere and stratosphere. In the deep ocean it is often a factor of 10 or so less.

In the atmosphere the stable stratification is set up by a balance between

(i). Short-wave heating in the stratosphere (ozone) and at the ground.

(ii). Long-wave heating which transfers heat from the ground to the atmosphere and within the atmosphere.

(iii). Convection and larger-scale dynamical processes which tend to stabilise the temperature profile in the lower part of the atmosphere.

2.5 Summary: density stratification

(i). Density stratification can give stability with respect to vertical displacements. The size of vertical displacements is limited. A parcel can move large distances in the vertical only if it 'forgets' its density.

(ii). Hydrostatic equilibrium is possible only if there are no horizontal density gradients. If there are horizontal density gradients in an initial condition then fluid accelerations will result and an adjustment process will take place.

3 Rotation

It was noted in §1 that the effect of rotation is to add two fictitious forces, the Coriolis force and the centrifugal force, and that the latter can be combined with the gravitational force and represented by an 'effective gravititational potential', Φ.

It is then convenient to work in a coordinate system where 'vertical' means 'perpendicular to surfaces of constant Φ'. These surfaces are so close to spherical that the geometric corrections associated with non-sphericity may be neglected. See Gill §4.12 for more discussion.

3.1 Shallow water equations

To study the effects of rotation and density stratification together we shall consider the simplest possible system that contains both ingredients – a single layer of homogeneous fluid, of density ρ, resting on a horizontal plane and with a free upper surface, in the presence of background rotation about a vertical axis at rate $\frac{1}{2}f$. f is called the *Coriolis parameter*. The system is described using Cartesian coordinates (x, y, z), with x, y horizontal and z vertical.

[In fact if a possible steady state of the system is for the fluid to be at rest with the layer depth constant the lower boundary cannot be quite flat. If the axis of rotation is the line $x = 0$, $y = 0$ then it must have shape $z = f^2(x^2 + y^2)/8g$ (so that it is an equipotential surface of the effective gravitational potential, including both gravity and the centrifugal force). Provided that $x^2 + y^2 \ll g^2/f^4$ the deviation from being exactly flat is very small.]

We take the lower boundary to be $z = -H$, where H is constant, and describe the free surface by $z = \eta(x, y, t)$.

Provided that the layer of fluid is shallow relative to the horizontal length scale of the motion, we may make the *hydrostatic approximation*, assuming that the dominant balance in the vertical momentum equation is pressure gradient balancing gravitational force to deduce that

$$\frac{\partial p}{\partial z} = -\rho g \qquad \Rightarrow \qquad p = p_{\text{atm}} - \rho g(z - \eta) \tag{3.1}$$

where p_{atm} is the pressure in the medium above the free surface (whose density is assumed to be negligible)

If we also assume that the horizontal velocities are independent of depth the horizontal momentum equations become

$$u_t + uu_x + vu_y - fv = -g\eta_x \tag{3.2}$$

$$v_t + uv_x + vv_y + fu = -g\eta_y \tag{3.3}$$

and the integral of the continuity equation implies that

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = -\frac{1}{(H+\eta)}[w]_{z=-H}^{z=\eta} = \frac{1}{(H+\eta)}\frac{D}{Dt}(H+\eta) \tag{3.4}$$

where

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y}$$

These equations (3.2), (3.3) and (3.4) are the governing equations for the shallow-water system.

For the time being we shall assume that the lower boundary is flat, so that $H =$ constant. To make the mathematics of the shallow water equations easier it is useful to assume that the variations in free surface height are small, compared to the depth

$$|\eta| \ll H$$

Since it is the size of η that drives the subsequent motion, we shall assume that u and v are also small, i.e. $u, v \ll \max(c, fL)$ where L is a typical horizontal length scale. The equations (3.2), (3.3) and (3.4) may then be linearised about a state of rest to give

$$u_t - fv = -g\eta_x \tag{3.5}$$

$$v_t + fu = -g\eta_y \tag{3.6}$$

$$\eta_t + H(u_x + v_y) = 0 \tag{3.7}$$

We may construct an energy equation for the linearised shallow water equations by taking (3.5) $\times\, u$ + (3.6) $\times\, v$ + (3.7) $\times\, \eta g/H$, giving

$$\left(\tfrac{1}{2}u^2 + \tfrac{1}{2}v^2 + \tfrac{1}{2}\eta^2\frac{g}{H}\right)_t + (g\eta u)_x + (g\eta v)_y = 0 \tag{3.8}$$

This is an energy conservation equation, allowing a budget of energy to be calculated.

3.2 Rossby adjustment problem

Consider the adjustment of the shallow-water system from an initial condition in which the fluid is at rest, but where there is horizontal variation of the thickness of the fluid layer. It follows that there are horizontal pressure gradients which will accelerate fluid parcels.

For simplicity we consider an initial condition which is y-independent, i.e. we take $\eta(x, y, 0) = \eta_0(x)$. It is convenient to assume that $\eta_0(-\infty) = -\eta_0(\infty) = h$ and that $\eta_0(0) = 0$.

If the initial conditions are y-independent then the solution will be y-independent for all time, so we suppress any dependence on y. We also assume that $|\eta_0(x)| \ll H$ for all x, so that the linearised equations (3.5), (3.6) and (3.7) may be used.

If the Coriolis parameter $f = 0$ then u and v can be eliminated from (3.5), (3.6) and (3.7) to give

$$\eta_{tt} = gH\eta_{xx}. \tag{3.9}$$

This is a 1-D non-dispersive wave equation with wave speed $c = (gH)^{1/2}$.

If u and v are zero initially then it follows from (3.7) that $\eta_t = 0$ at $t = 0$. The solution of (3.9) subject to the given initial conditions is

$$\eta(x, t) = \tfrac{1}{2}\left\{\eta_0(x - ct) + \eta_0(x + ct)\right\}.$$

For the assumed form of the initial condition, at time t there is a region of size $2ct$ centred on $x = 0$ in which $\eta = 0$. The fluid in this region is in motion with constant velocity U say, with $U \simeq gh/c$. At the edge of this region there are outward propagating wave fronts each travelling with speed c.

After sufficient time has elapsed, in any finite region $-l < x < l$, the potential energy has been reduced to zero and there is a corresponding amount of kinetic energy. *All the potential energy in the initial state is converted to potential energy.* Globally, there is an infinite amount of potential energy in the initial condition and the amount converted to kinetic energy increases with time without bound.

Now consider the rotating case, with $f \neq 0$.

We shall be interested in the final state, after any adjustment has taken place, so note first that in the steady state (3.5), (3.6) and (3.7) reduce to

$$-fv = -g\eta_x \tag{3.10}$$

$$fu = -g\eta_y \tag{3.11}$$

$$H(u_x + v_y) = 0 \tag{3.12}$$

(3.10) and (3.11) express the condition of *geostrophic balance* in which the pressure gradient balances the Coriolis force.

However we cannot determine the nature of the steady state solely from (3.10), (3.11) and (3.12), since any one of these equations may be derived from the other two.

For the non-rotating case we deduce the final state by solving the equation governing the time evolution. With $f \neq 0$ we may deduce a corresponding equation from (3.5), (3.6) and (3.7) as

$$\eta_{tt} - gH(\eta_{xx} + \eta_{yy}) = -fH(v_x - u_y). \tag{3.13}$$

This is not as simple as in the $f = 0$ case because of the term on the right-hand side.

Consider also the vorticity equation, formed by $\partial/\partial x$ (3.6) minus $\partial/\partial y$ (3.5). (Note that we are considering only the vertical component of the vorticity.) This implies that

$$\zeta_t = v_{xt} - u_{yt} = -f(u_x + v_y) = \frac{f\eta_t}{H}$$

i.e. that

$$\tilde{P} = \zeta - \frac{f\eta}{H} = v_x - u_y - \frac{f\eta}{H} \quad \text{is independent of time.} \tag{3.14}$$

This is an approximate form of the result that for the full shallow-water equations, i.e. for (3.2), (3.3) and (3.4), the quantity

$$P = \frac{\zeta + f}{H + \eta}$$

is conserved (in the absence of frictional forces etc.) following fluid particles. P is called the potential vorticity. The particular expression for P given here holds for the shallow-water equations. There are corresponding expressions for potential vorticity in other systems of equations. [Derivation of conservation of P and demonstration of equivalence to (3.14) in the appropriate limit is left as an exercise.]

Adding $f^2\eta$ to the right-hand side of (3.13) and using (3.14) it follows that

$$\eta_{tt} - gH(\eta_{xx} + \eta_{yy}) + f^2\eta = -fH(v_x - u_y - f\eta/H) = -fH\tilde{P} = -fH\tilde{P}|_{t=0}. \tag{3.15}$$

This is the generalisation of (3.9) to the case $f \neq 0$ and may be solved given initial conditions on η and η_t. Note that this is a dispersive wave equation and that the initial conditions are 'remembered' through the forcing term on the right-hand side.

We now seek a steady-state solution assuming that at $t = 0$ $u = v = 0$ and $\eta = \eta_0(x)$. As before we assume that the absence of any y-dependence in the initial conditions implies that the solution may be taken to be y-independent for all time.

To fix ideas we consider the case $\eta_0(x) = h\,\mathrm{sgn}(x)$. It follows that $-fHP|_{t=0} = -f^2h\,\mathrm{sgn}(x)$. The steady state is therefore described by the solution of

$$-gH\frac{d^2\eta}{dx^2} + f^2\eta = -f^2h\,\mathrm{sgn}(x). \tag{3.16}$$

We seek solutions that are bounded as $|x| \to \pm\infty$ and continuous at $x = 0$. The relevant solution of (3.16) is

$$\eta(x) = h\,\mathrm{sgn}(x)\left\{1 - \exp(-|x|/L_R)\right\},$$

where $L_R = (gH)^{1/2}/f$.

From (3.10) and (3.11), i.e. the statement that the steady-state flow is in geostrophic balance, it follows that

$$u(x) = 0$$

and

$$v(x) = (g/H)^{1/2} h \exp(-|x|/L_R).$$

The length scale L_R is called the *Rossby radius of deformation.*

The adjustment problem just considered is called the *Rossby adjustment problem.* It describes how a fluid with horizontal gradients in thickness (or density) adjusts to a geostrophically balanced steady state. What have we learned so far?

(i). Rotation (i.e. $f \neq 0$) allows a steady state in which η varies in the horizontal and in which the corresponding horizontal pressure gradients are balanced by Coriolis forces.

(ii). Adding rotation inhibits the release of the potential energy available in the initial configuration. Note that the change in potential energy ΔV is given by

$$\Delta V = \int_{-\infty}^{\infty} \tfrac{1}{2} \frac{g}{H} [\eta^2]_{t=0}^{t=\infty} = -\frac{gh^2}{H} \int_0^{\infty} \{1 - (1 - e^{-x/L_R})^2\} dx = -\tfrac{3}{2} \frac{gh^2}{H} L_R \qquad (3.17)$$

and is finite, in contrast to the case where $f = 0$.

The corresponding change in kinetic energy ΔT is given by

$$\Delta T = \int_{-\infty}^{\infty} \tfrac{1}{2} \frac{gh^2}{H} e^{-2|x|/L_R} dx = \tfrac{1}{2} \frac{gh^2}{H} L_R \qquad (3.18)$$

and is also finite.

But note that $\Delta T = -\Delta V/3$ and $\Delta T + \Delta V < 0$. Where has the extra energy gone?

(iii). In order to determine the steady state we had to retain information about the initial conditions using the property of conservation of potential vorticity. In the linearised problem, because fluid particle displacements are small, the disturbance potential vorticity $\tilde{P}$ remains constant at each location.

(iv). There is an important role for the length scale L_R. This can be regarded as the length scale on which the effects of density stratification and rotation balance. Note, for example that L_R is the distance over which gravity waves in the shallow-water system propagate in a time f^{-1}.

3.3 The dynamics of adjustment

The equation (3.15) is linear with a steady forcing term on the right-hand side, so the time evolution is governed by the unforced equation

$$\eta_{tt} - gH(\eta_{xx} + \eta_{yy}) + f^2\eta = 0 \tag{3.19}$$

As noted earlier this is a dispersive wave equation. Behaviour of solutions may be conveniently analysed by considering solutions of the form $\eta(x, y, t) = \mathrm{Re}(\hat{\eta}e^{ikx-i\omega t})$ where k is a constant wavenumber and ω is a constant frequency and deriving a dispersion relation giving ω as a function of k.

It is useful to represent other variables by similar expressions, i.e. $u(x, y, t) = \mathrm{Re}(\hat{u}e^{ikx-i\omega t})$ and $v(x, y, t) = \mathrm{Re}(\hat{v}e^{ikx-i\omega t})$. (3.6) and (3.7) then imply that $\hat{u} = \omega\hat{\eta}/kH$ and $\hat{v} = -if\hat{u}/\omega$.

The dispersion relation is

$$\omega^2 = f^2 + gHk^2, \text{ i.e. } \omega = \pm(f^2 + gHk^2)^{1/2} \ . \tag{3.20}$$

The two branches of the dispersion relation correspond to waves with phase propagation in two different directions. Note that the relative size of the two terms appearing in the expression for ω is determined by the relative size of k to $f/(gH)^{1/2} = L_R^{-1}$.

For $k \gg L_R^{-1}$ (i.e. lengthscales much smaller than L_R) we have $\omega \simeq \pm(gH)^{1/2}k$. This is just the dispersion relation for non-dispersive gravity waves in the absence of rotation. In this limit the effects of rotation are very small. Note that $|\hat{v}| \ll |\hat{u}|$, i.e. the horizontal velocity is almost parallel to the x-direction, i.e. the direction of propagation.

For $k \ll L_R^{-1}$ (i.e. lengthscales much larger than L_R) we have $\omega \simeq \pm f(1 + \frac{1}{2}gHk^2/f^2 \ldots)$. This describes oscillations with frequency close to f. The dependence of ω on k is weak, so that the group speed is very small (approximately $gHk/f = (gH)^{1/2}kL_R$).

These oscillations are called *inertial oscillations*. They are the natural oscillations about a state of rest in a rotating system and result from a balance between acceleration and Coriolis force. They are captured by setting $\eta = 0$ in (3.5) and (3.6) to give

$$\begin{aligned} u_t - fv &= 0 \\ v_t + fu &= 0. \end{aligned}$$

In this limit, therefore, the effects of stratification (i.e. the effects in the shallow-water system of variations in η) may be neglected. Note that $\hat{v} \simeq \mp i\hat{u}$, i.e. the y-velocity is equal in magnitude to and $\frac{1}{2}\pi$ out of phase with the x-velocity. Particles travel in 'inertial circles.

In the intermediate regime, $k \sim L_R^{-1}$, both stratification and rotation are equally important. $\hat{v} = -if\hat{u}/\omega$, i.e. the y-velocity is has magnitude less than and is $\frac{1}{2}\pi$ out of phase with the x-velocity. Particles travel in ellipses whose eccentricity increases as ω increases.

The waves admitted by (3.19), i.e. shallow-water gravity waves modified (sometime strongly) by rotation, are called *Poincaré waves*.

The Rossby adjustment problem demonstrates the adjustment of an arbitrary initial condition to a state of geostrophic balance. (The process is sometimes called *geostrophic adjustment*.)

(Everyone accepts that the wind blows in a direction parallel to isobars. The Rossby adjustment problem tells us how it got to be that way.)

The adjustment is achieved by the propagation of Poincaré waves away from the region where there was initially horizontal variation in η. The short wavelength waves have fast group velocities and therefore propagate away quickly. The longer waves have slower group velocities and therefore take longer to propagate away. It is the longer waves (almost inertial oscillations) that dominate at large times (and the decay of the oscillations is only algebraic in time, since the group velocity vanishes as $k \to 0$). Gill, Chapter 7, shows detailed solutions constructed from Bessel functions.

The energy that appears to go 'missing' when comparing the initial and final states is carried away by the waves.

In the atmosphere and the ocean the large-scale slowly varying part of the flow is observed to be close to a state of geostrophic balance. (We are more precise about what is meant by 'large-scale' and 'slowly varying' later.) We might interpret this as being, in part, a result of continuous adjustment in response to slowly changing forcing.

The dispersion relation derived above for the Poincaré waves has two branches, i.e. there are two frequencies, one positive, one negative, for each value of k. But the original shallow-water equations (3.2), (3.3) and (3.4), and also the linearised forms (3.5), (3.6) and (3.7), had three time derivatives, so we might have expected three branches for the dispersion relation. It is important not to forget the steady geostrophically balanced flow that is left after the adjustment. This corresponds to the third branch, which, in this particular problem, has zero frequency.

The third branch of the dispersion relation does not always have zero frequency – it occurs in the problem considered above because of the assumption of constant thickness η and hence constant potential vorticity P in the undisturbed state about which the equations are linearised.

3.4 The effect of potential vorticity variations (and boundaries) on adjustment

We now consider small amplitude waves propagating in a different system — namely a channel running in the x-direction with rigid boundaries at $y = 0$ and $y = L$, with the additional important modification that the bottom of the channel is taken to be slightly sloping. The potential vorticity in the basic state is then no longer constant. This will highlight two features:

(i). the modification to gravity-like waves from the channel geometry;

(ii). the presence of a wave of non-zero frequency associated with the potential vorticity gradient in the basic state.

The modification to the form of the linearised shallow water equation used previously is that H is a function of y, so (3.7) now becomes

$$\eta_t + H(u_x + v_y) + H_y v = 0 \tag{3.21}$$

If $H_y \neq 0$ it is not possible to show that the linearized expression for the potential vorticity is constant in time. Instead one has an equation for the rate of change of this quantity in time.

$$\tilde{P}_t = v_{xt} - u_{yt} - \frac{f}{H}\eta_t = \frac{vf}{H}H_y = -\frac{v}{H}\left(\frac{f}{H}\right)_y .$$

This expresses, within the approximation of linearization, that the potential vorticity at a particular point may change as a result of advection across the potential vorticity gradient in the basic state.

Eliminating u, then v, leaves the equation.

$$\eta_{ttt} + f^2\eta_t - gH\left(\eta_{xxt} + \eta_{yyt}\right) - gH_y\eta_{yt} + fg\eta_x H_y = 0. \tag{3.22}$$

This equation has y-dependent coefficients, so the equation for y structure will, in general, be non-trivial. The boundary conditions are that $v = 0$ at $y = 0$ and L. In order to express these in terms of η it is useful to note that $v_{tt} + f^2 v = -g\eta_{yt} + fg\eta_x$.

Now assume that $H = H_0(1 - \epsilon y/L)$, where H_0 is constant and ϵ is small, and seek wave solutions of the form $\eta = Re\left(\hat{\eta}(y)e^{ikx-i\omega t}\right)$. Substituting into (3.22) gives

$$\omega((\omega^2 - f^2)\hat{\eta} + gH(\hat{\eta}_{yy} - k^2\hat{\eta})) + g\frac{\epsilon H_0}{L}\omega\hat{\eta}_y + \epsilon kfg\hat{\eta}\frac{H_0}{L} = 0 \tag{3.23}$$

with the boundary conditions

$$\omega\hat{\eta}_y + kf\hat{\eta} = 0 \quad \text{on} \quad y = 0, L. \tag{3.24}$$

Posing the expansion $\hat{\eta} = \hat{\eta}_0 + \epsilon\hat{\eta}_1 \ldots$ it follows at leading order in ϵ that

$$\omega\left\{(\omega^2 - f^2 - gH_0k^2)\hat{\eta}_0 + gH_0\hat{\eta}_{0yy}\right\} = 0 \tag{3.25}$$

There are two distinct possibilities for solutions.

(i). $\omega = 0$ (at leading order in ϵ), with $\hat{\eta}_0$ arbitrary – except that $\hat{\eta}_0(0) = \hat{\eta}_0(L) = 0$. These solutions represent 'quasi-geostrophic' wave modes and it will be necessary to go to higher order in ϵ to obtain any useful information about them.

(ii). **either** $(\omega^2 - f^2 - gH_0k^2)\hat{\eta}_0 + gH_0\hat{\eta}_{0yy} = 0$ so $\hat{\eta}_0 = A_+e^{ily} + A_-e^{-ily}$ with $\omega^2 = f^2 + gH_0(k^2 + l^2)$. The boundary conditions imply that $l = \frac{n\pi}{L}$, where n is a positive integer, and give a relation between A_+ and A_-. [The corresponding solution is a superposition of Poincaré waves such that the boundary condition is satisfied.]

or $\omega\hat{\eta}_{0y} + kf\hat{\eta}_0 \equiv 0$ implying that $l = \pm ikf/\omega$ and $\omega^2 - f^2 + gH_0k^2\left(1 - f^2/\omega^2\right) = \left(\omega^2 - f^2\right)\left(1 - gH_0k^2/\omega^2\right) = 0$. hence $\omega = \pm k\sqrt{gH_0}$ and $\hat{\eta}_{0y} = \mp(f/\sqrt{gH_0})\eta_0$ everywhere in the channel i.e., $\hat{\eta} \propto e^{\pm fy/\sqrt{gH_0}} = e^{\pm y/L_R}$. $\omega = +k\sqrt{gH_0}$ corresponds to a wave confined to the $y = 0$ wall, $\omega = -k\sqrt{gH_0}$ to a wave confined to the $y = L$ wall. These two waves are called *Kelvin waves.* The Kelvin wave is in geostrophic balance across the channel, $fu = -g\eta_y$, and behaves like a non-dispersive (non-rotating) gravity wave in propagating along the channel. The Kelvin waves are truly confined to one boundary or the other when the channel is broad, i.e. $L \gg L_R$. [Note that we have ignored the root $\omega = \pm f$. Further investigation shows that this is not acceptable unless $k = \sqrt{gH_0}/f$, when it simply gives the Kelvin wave solution.]

We now consider the effect of the bottom slope and therefore expect to have to go to higher order in the perturbation series in ϵ. For the Kelvin and Poincaré modes, this will simply give corrections to the structure of the mode and to the frequencies. However, with regard to the geostrophic modes we have as yet no information about the structure. All we have is the 'trivial' $\omega = 0$ root.

Returning to the equation (3.23) for $\hat{\eta}$, we see that there is a balance possible if $\omega = \mathrm{O}(\epsilon)$, in which case the dominant terms are

$$\omega\left(\hat{\eta}_{0yy} - k^2\hat{\eta}_0 - \frac{f^2}{gH_0}\hat{\eta}_0\right) - \frac{k\epsilon}{L}\hat{\eta}_0 = 0 \tag{3.26}$$

with the boundary condition (3.24) reducing to $\hat{\eta} = 0$ at $y = 0, L$.

It follows that $\hat{\eta} = \sin n\pi y/L$ with n a positive integer and

$$\omega = -\frac{f\epsilon k}{L}\frac{1}{(k^2 + \frac{f^2}{gH} + \frac{n^2\pi^2}{L^2})} \tag{3.27}$$

The geostrophic modes are thus resolved as a set of wave modes with frequency $\mathrm{O}(\epsilon)$. They have frequency $< \epsilon f$ and phase speed $< \epsilon\sqrt{gH_0}$

The phase speed of these modes is negative if ϵ is positive, i.e. if the channel is shallower with increasing y. The direction of phase propagation is thus with shallow water to the right (in the Northern Hemisphere). These modes are referred to as *topographic Rossby waves*. In the ocean these waves propagate along topographic slopes associated with the continental shelves.

Kelvin waves are important in the ocean in propagating information along coastlines (in a direction with the coast to the right in the Northern Hemisphere, to the left in the Southern Hemisphere). Examples are the propagation of storm surges (e.g. around the North Sea) or propagation of changes in the temperature (e.g. northwards along the west coast of Central and North America after an El Nino event).

3.5 Fast modes and slow modes

We have seen above that if there is variation of potential vorticity in the basic state then there are two waves with frequencies greater than or equal to f (two 'fast' modes) and one wave with frequency much less than f (if the proportional variation in potential vorticity across the flow is small) (one 'slow' mode).

Without solving an adjustment problem in detail, we might therefore expect it to be two-stage

(i). Rossby adjustment to a state close to geostrophic balance

(ii). (if potential vorticity is not constant) slow adjustment, e.g. through the propagation of topographic Rossby waves.

The flow in the second stage is 'balanced', i.e. close to geostrophic balance, since the frequency of the waves is small and therefore $|u_t| \ll |fv|$ etc. It follows that, in this second stage, at any instant if we know one of u, v and η (most obviously η) then we know the other two. The time evolution should therefore be governed by single 'prognostic' equation (i.e. an equation containing a $\partial/\partial t$ term) rather than three.

In fact this extends to classes of flow that are more general than linear waves.

3.6 A prognostic equation for the slow modes

We take U, L and T to be respectively estimates of horizontal velocities, horizontal length scales and time scales, and estimate the different terms on the left-hand side of the momentum equation.

$$
\begin{array}{cccll}
u_t + & uu_x + vu_y & -fv & = -g\eta_x & \\
v_t + & uv_x + vv_y & +fv & = -g\eta_x & \\
 & & & & \\
UT^{-1} & U^2L^{-1} & fU & \text{estimated magnitudes} & \\
(fT)^{-1} & U(fL)^{-1} & 1 & \text{relative magnitudes} &
\end{array}
\tag{3.28}
$$

The dominant balance in the horizontal momentum equation is geostrophic balance if

$$
\max\left\{\frac{1}{fT}, \frac{U}{fL}\right\} \ll 1 \tag{3.29}
$$

The dimensionless quantities appearing on the left-hand side are both possible definitions of the *Rossby number*, *Ro*. Which is appropriate depends on what quantities are naturally fixed for a specific problem and what emerge as part of the solution. For example if the problem involves a forcing with a specified frequency then $(fT)^{-1}$ is the natural definition. If the problem involves flow at a specified rate over topography with a specified horizontal length scale then $U(fL)^{-1}$ is the natural definition.

The flow is close to geostrophic balance if the Rossby number, *Ro*, is small.

Now consider the magnitude of the varying part of the free-surface height, η, relative to a typical average thickness H_0. Geostrophic balance implies that

$$
\frac{\eta}{H_0} \sim \frac{fUL}{gH_0} = \frac{U}{fL}\frac{L^2}{L_R^2}. \tag{3.30}
$$

Hence $\eta/H_0 \ll 1$ if $Ro \ll L_R^2/L^2$.

For present purposes we shall take $L \sim L_R$, hence if $Ro \ll 1$ then relative variations in the free surface height are small.

Under these assumptions, in each of the shallow-water governing equations (3.2), (3.3) and (3.4) the terms involving time derivatives are small and other terms dominate. The time evolution of the flow is therefore governed by small deviations from the leading-order balance (of geostrophic balance and non-divergence of the horizontal flow).

The equation governing the time evolution can be derived by a systematic asymptotic calculation based on the Rossby number as a small parameter. We do this later for the three-dimensional continuously stratified case. Here, for the single-layer system, we adopt a more heuristic approach.

We have already seen the important role played by the potential vorticity P. Therefore consider how P can be approximated when $Ro \ll 1$.

We have $P = (v_x - u_y + f)/(H + \eta)$. Geostrophic balance implies that

$$u \simeq u_g = -g\eta_y/f = -\psi_y \text{ and } v \simeq v_g = g\eta_x/f = \psi_x \tag{3.31}$$

where $\psi = g\eta/f$ is a streamfunction for the leading-order horizontal flow. Now write $H = H_0 + H_1$, where H_0 is constant and H_1 is spatially varying, and $|H_1| \ll H_0$, i.e. spatial variations in the undisturbed layer depth are relatively weak. It follows that

$$P \simeq (\psi_{xx} + \psi_{yy} + f)/(H_0 + H_1 + f\psi/g) \tag{3.32}$$

$$\simeq \frac{f}{H_0} + \frac{\psi_{xx} + \psi_{yy}}{H_0} - \frac{f}{H_0^2}(H_1 + \frac{f\psi}{g}) \tag{3.33}$$

$$\simeq \frac{f}{H_0} + \frac{P_g}{H_0} \tag{3.34}$$

(3.35)

where

$$P_g = \psi_{xx} + \psi_{yy} - \frac{f^2\psi}{gH_0} - \frac{fH_1}{H_0}. \tag{3.36}$$

P_g is called the *quasi-geostrophic potential vorticity.*

The leading- order form of the equation $DP/Dt = 0$ is then

$$\left(\frac{\partial}{\partial t} + \mathbf{u}_g.\nabla\right) P_g = 0, \tag{3.37}$$

where $\mathbf{u}_g = (u_g, v_g) = (-\psi_y, \psi_x)$. This eqation is called the *quasigeostrophic potential vorticity equation.*

Together (3.36) and (3.37) form a self-contained evolution equation describing the time-evolution of the 'slow', 'balanced' flow. (3.37) gives a rule for updating P_g. (3.36) allows the new ψ (and hence the new u, v and η) to be determined from the new P_g. The corresponding operator acting on P_g is called the 'inversion ' operator. Evaluation of the inversion operator acting on P_g requires the solution of an elliptic equation and therefore requires suitable boundary conditions.

Here are some important points that arise from the equations (3.36) and (3.37).

(i). **Generic form of the equations**: The generic form of the equations is (a) an evolution equation

$$\frac{DP}{Dt} = 0 \tag{3.38}$$

describing the material conservation of a quantity P (a suitable potential vorticity) and (b) an inversion operator which determines instantaneously all other flow quantities given P. (Note that in the presence of frictional or diabatic terms effects or similar, e.g. a mass source in the shallow-water equations, the right-hand side will be non-zero.) A familiar set of equations with this structure are the equations for incompressible two-dimensional flow, in the form of a equation describing material conservation of vorticity, plus an equation expression vorticity in terms of streamfunction, or vice-versa,

$$\frac{D\zeta}{Dt} = \frac{\partial \zeta}{\partial t} + u\frac{\partial \zeta}{\partial x} + v\frac{\partial \zeta}{\partial y} = 0 \tag{3.39}$$

and

$$\nabla^2 \psi = \zeta \tag{3.40}$$

with ψ the streamfunction, $u = -\partial\psi/\partial y$ and $v = \partial\psi/\partial x$.

(ii). **'Slow' vs 'fast'**: The specific form of the equations (3.36) and (3.37), or the generic form just described, contain only one time derivative, whereas the shallow water equations from which they were derived contained three time derivatives. Therefore, for the case of small-amplitude disturbances about a steady basic state, the equations (3.36) and (3.37) allow only one wave (at given horizontal wavenumber) whereas the original shallow water equations allowed three waves. The wave allowed is precisely that which appeared as a low-frequency mode in the analysis of the channel with weakly varying depth. But the reduction in the number of time derivatives from three to one is important in more general contexts than small-amplitude waves. The reduced equations are sometimes said to allow 'slow' motion or 'slow modes' that have a vortex-like character, to be distinguished from 'fast modes' analogous to Poincaré waves. For more discussion see the McIntyre article on 'Balanced motion' in the Encyclopedia of Atmospheric Sciences (available from `http://www.atm.damtp.cam.ac.uk/mcintyre/papers/ENCYC/`.

(iii). **Rossby adjustment**: We may now summarise the method of finding the steady state in the Rossby adjustment problem as (i) determine the distribution of P_g in the initial state and deduce the distribution of P_g in the final state (trivial in the linearised problem) and (ii) solve (3.36) to deduce ψ and hence all other flow variables. Note that there are sometimes subtleties, e.g. for the adjustment problem in a flat-bottomed channel $0 < y < L$, since arbitrary multiples of the functions $\exp(y/L_R)$ and $\exp(-y/L_R)$ may be added to ψ without violating the boundary condition $\partial\psi/\partial x = 0$ on $y = 0$ and $y = L$. The coefficients of these functions can be determined only by considering the time-dependent adjustment and are set by Kelvin waves that propagate away during the adjustment. (See Gill §10.7.) The Kelvin waves therefore play an implicit role in the balanced problem.

(iv). **Rossby waves**: The low-frequency waves found in the analysis of the channel problem and allowed by the equations (3.36) and (3.37) may generically be called *Rossby waves*. They depend on the existence of a potential vorticity gradient in the basic state and their dynamics may be understood in terms of potential vorticity advection and the resulting changes in circulation.

Consider a resting basic state $\psi = 0$ with $H_1 = -\epsilon H_0 y/L$. According to (3.36), if $f > 0$ (Northern Hemisphere like), P_g increases in the positive y direction. Now consider the effect of displacing a fluid parcel in the positive y direction. P_g is conserved following this parcel, H_1 is decreased from its value at the original position of the parcel, therefore $\psi_{xx} + \psi_{yy} - \frac{f^2\psi}{gH_0}$ decreases, implying a circulation in the clockwise (anticyclonic) sense about the displaced parcel. Displacing a fluid parcel in the negative y direction correspondingly gives a circulation in the anti-clockwise (cyclonic) sense.

Now consider a pattern, extending in the x-direction, of parcels displaced alternately in the positive and negative y-directions. The sense of the resulting circulations is move the pattern in the negative x-direction and also to limit the magnitudes of the displacements.

The gradient of potential vorticity therefore gives rise to a restoration mechanism, the *Rossby restoration* mechanism, and to waves with phase propagation in the negative x-direction, when the potential vorticity increases in the positive y-direction. (There is an analogy with the restoration mechanism associated with a stable density gradient in the vertical, though there are significant differences between the two mechanisms. Note that in the Rossby case a displacement implies a velocity, whereas in the density-stratified case a displacement implies an acceleration.)

4 Effects of rotation and stratification in 3-D flow

4.1 The Taylor-Proudman theorem

(This is noted in passing for historical completeness.) For a rapidly rotating *homogeneous* fluid the dominant balance in the momentum equation is

$$2\mathbf{\Omega} \times \mathbf{u} = -\frac{1}{\rho_0}\nabla p \qquad (4.1)$$

where ρ_0 is the (constant) density. Assume that $\mathbf{\Omega}$ is in the z-direction, then (4.1) implies $p = p(x, y, t)$, hence $u = u(x, y, t)$ and $v = v(x, y, t)$.

The flow in the plane perpendicular to $\mathbf{\Omega}$ is therefore independent of z. This is the **Taylor-Proudman theorem**.

The theorem has remarkable consequences, e.g. if a sphere is towed horizontally through a fluid that is rapidly rotating in the vertical, then the fluid above and below the body (forming a 'Taylor column' must move with it). There are many beautiful laboratory studies of this and related phenomenon.

However the relevance to the atmosphere and the ocean is somewhat limited, because of the presence of density stratification.

4.2 Some basic facts about rotation and stratification in 3-D

(i). In a model problem with stratification represented by constant buoyancy frequency N and with rotation $\frac{1}{2}f$ about the vertical axis the dispersion relation for small amplitude waves is

$$\omega^2 = \frac{N^2k^2 + f^2m^2}{k^2 + m^2} \tag{4.2}$$

where k is the horizontal wavenumber and m is the vertical wavenumber. It follows that the relative strength of stratification vs rotation is N/(horizontal length scale) vs f/(vertical length scale).

(ii). In the atmosphere and the ocean N is typically much larger than f. f is $O(10^{-4}\text{s}^{-1})$. N is $O(10^{-3}\text{s}^{-1})$ (deep ocean) or $O(10^{-2}\text{s}^{-1})$ (upper ocean and atmosphere).

(iii). It follows that rotation is important only if vertical length scales are much less than horizontal length scales, but this implies that vertical velocities are much less than horizontal velocities and that the hydrostatic approximation is valid.

(iv). Given the above, the Coriolis force may be neglected in the vertical momentum equation and in the horizontal momentum equation only the part of the Coriolis force associated with the horizontal velocity need be included. These approximations are equivalent to replacing the rotation vector by its vertical component only.

(v). The sequence of equations that follow from last two points, plus the geometric simplication that the fluid layer is thin compared to the radius of the Earth, are called the *primitive equations*. These have, until very recently, been widely used as a basis for numerical modelling of atmosphere and ocean. (The most recent models are designed to be valid at very high horizontal resolution, where the hydrostatic approximation may no longer apply.)

(vi). The vertical component of rotation is a function of position. However over a limited region the variation may be neglected and the Coriolis parameter set to a constant value $f = 2\Omega \sin\phi$ where Ω is the rotation rate and ϕ is some suitable latitude. This is called the *f-plane approximation*.

4.3 Effects of spherical geometry: the β-plane

On large scales we need to take account of the sphericity of the Earth's surface. For accurate numerical computation we need to use a spherical coordinate system, however the algebra is complicated by the fact that the coordinates are curvilinear.

It turns out that a model which neglects all effects of sphericity except the latitudinal variation of the Coriolis parameter f captures many of the important phenomenona that result from the fact that the Earth is a rotating sphere.

Consider a narrow region centred on latitude ϕ_0, so

$$f = 2\Omega \sin\phi \simeq 2\Omega \sin\phi_0 + 2\Omega \cos\phi_0\,(\phi - \phi_0). \tag{4.3}$$

Now use Cartesian coordinates with x distance in longitudinal direction and y distance in latitudinal direction, so $y = a(\phi - \phi_0)$, where a is the radius of the Earth. Hence

$$f = f_0 + \beta y \text{ where } f_0 = 2\Omega \sin\phi_0,\ \beta = 2\Omega \cos\phi_0\,/a. \tag{4.4}$$

The approximation (4.4) is conventionally called the *β-plane approximation*. The insight that the latitudinal variation of the Coriolis parameter was the most important effect of sphericity came from Rossby in 1939.

4.4 The primitive equations

For a Boussinesq fluid on a β-plane, the primitive equations take the form

$$\frac{Du}{Dt} - (f_0 + \beta y)v = -\frac{1}{\rho_0}p'_x \tag{4.5}$$

$$\frac{Dv}{Dt} + (f_0 + \beta y)u = -\frac{1}{\rho_0}p'_y \tag{4.6}$$

$$p'_z = -\rho' g \tag{4.7}$$

$$\frac{D\rho'}{Dt} = 0 \tag{4.8}$$

$$\nabla.\mathbf{u} = 0 \tag{4.9}$$

where $\mathbf{u} = (u, v, w)$, $D/Dt = \partial/\partial t + u\partial/\partial x + v\partial/\partial y + w\partial/\partial z$.

We use the Boussinesq form of the primitive equations for simplicity, but for large-scale atmospheric flows it is often necessary to take account of compressibility. This may be done most easily by a change of vertical coordinate to a function of pressure rather than geometric height. Using the equation of state for a perfect gas then gives rise to a very similar set of equations to those above. (See section 4.7).

As is the case for the shallow-water equations, the primitive equations have three prognostic equations and admit both fast and slow modes. Indeed the primitive-equation system behaves as if it consisted of many shallow-water systems stacked on top of each other, communicating through the vertical derivative in the hydrostatic equation.

4.5 Thermal wind equation

When the Rossby number is small we expect the flow to be close to geostrophic balance, so that

$$-fv = -\frac{1}{\rho_0}p'_x, \tag{4.10}$$

$$fu = -\frac{1}{\rho_0}p'_y. \tag{4.11}$$

Then differentiating in the vertical and using the hydrostatic relation, it follows that

$$fv_z = -\frac{g}{\rho_0}\rho'_x \text{ and } fu_z = \frac{g}{\rho_0}\rho'_y \tag{4.12}$$

This is the *thermal wind equation.* (Note that ρ' can be a measure of temperature.)

This equation has been of practical interest since historically it has often been the density or temperature field that is observed and then the thermal wind equation may be used to deduce information about the velocity field. The vertical integration introduces an arbitrary function of x and y. This may be set in the atmosphere by low-level pressure observations or in the ocean by an ad hoc assumption of a 'level of no motion'.

[The modern approach is to use 'data assimilation' – all available observations, of different variables and taken at different positions and times are used as input to a dynamical model and space-time fields of all model variables constructed.]

4.6 Rossby-Ertel potential vorticity

For the shallow-water equations we deduced a prognostic equation for the slow motion from the equation for potential vorticity conservation. Recall that the potential vorticity is conserved exactly, according to the full shallow-water equations, without any assumption that the motion is slow.

The full equations for 3-dimensional density stratified flow also imply, without approximation, material conservation of a potential vorticity – generally called Rossby-Ertel potential vorticity – in the absence of forcing and dissipation. The full approximation for the Rossby-Ertel potential vorticity is

$$P = \frac{1}{\rho}(2\boldsymbol{\Omega} + \boldsymbol{\zeta}).\nabla\alpha \tag{4.13}$$

where α is any materially conserved quantity such that $\nabla\alpha.(\nabla p\times\nabla\rho)=0$. For a 'single-component' compressible fluid with equation of state, $\rho=\rho(p,\theta)$, α may therefore be any function of potential temperature θ. For a Boussinesq fluid α may be taken to be ρ' and the expression for P reduces to

$$P=\frac{1}{\rho_0}(2\boldsymbol{\Omega}+\boldsymbol{\zeta}).\nabla\rho'. \tag{4.14}$$

Finally the hydrostatic equations also give conservation of potential vorticity, provided that $2\boldsymbol{\Omega}+\boldsymbol{\zeta}$ is replaced by $(-v_z,u_z,f+v_x-u_y)$. [This can be verified from the equations.]

Thus for the Boussinesq form of the equations that we are using

$$P=\frac{1}{\rho_0}\left\{(f+v_x-u_y)\rho'_z+u_z\rho'_y-v_z\rho'_x\right\} \tag{4.15}$$

and it may be shown that $DP/Dt=0$ in the absence of forcing and dissipation.

4.7 Pressure coordinates

The hydrostatic equation

$$\frac{\partial p}{\partial z}=-\rho g=-\rho\frac{\partial\Phi}{\partial z}, \tag{4.16}$$

where Φ is the geopotential, may be used as the basis for a transformation of the primitive equations to a form in which the pressure p or its logarithm, in the form $Z=-H_s\log(p/p_*)$ where H_s is a constant 'scale height' and p_* is a constant reference pressure, are used as a vertical coordinate. The geometric height z, or equivalently the geopotential $\Phi=gz$, then becomes a dependent variable in the place of pressure.

The transformation does not involve any change in the reference vectors used to define the components of vectors such as the velocity. Thus u and v are retained as components of the velocity, though their directions are not necessarily perpendicular to p or Z coordinate surfaces. In particular u and v remain equal to the rate of change of x and y following a fluid parcel.

Further details will be given for the case where Z is the vertical coordinate.

The advective derivative D/Dt becomes

$$\frac{D}{Dt}=\frac{\partial}{\partial t}+u\frac{\partial}{\partial x}+v\frac{\partial}{\partial y}+W\frac{\partial}{\partial Z} \tag{4.17}$$

where $W=\dot{Z}$ is the rate of change of Z following a fluid parcel, i.e. the vertical velocity in log-pressure coordinates. Note that the partial derivatives $\partial/\partial x$ and $\partial/\partial y$ are to be evaluated keeping Z, not z, constant.

The transformed momentum equations are obtained simply by taking the original equations for u and v and making a change of independent variable. Apart from the advective derivative dealt with above, the pressure gradient terms must be written in terms of the new dependent variable z. It is helpful to note that

$$dp=p_xdx+p_ydy+p_zdz=p_xdx+p_ydy-\rho d\Phi \tag{4.18}$$

where the second equality follows from the hydrostatic relation. The subscripts are used to denote partial derivatives with x, y and z as independent variables. Note, for example, that the partial derivative p_x is calculated keeping z, not Z, constant.

Reordering, we have that

$$d\Phi = \frac{p_x}{\rho}dx + \frac{p_y}{\rho}dy - \frac{dp}{\rho} = \frac{p_x}{\rho}dx + \frac{p_y}{\rho}dy + \frac{p}{\rho}\frac{dZ}{H}, \qquad (4.19)$$

where the second equality follows from the relation between p and Z.

It follows, considering the coefficients of dx and dy, that the pressure gradient terms appearing in the u and v equations in their original form may be rewritten as

$$\frac{p_x}{\rho} = \frac{\partial \Phi}{\partial x} \quad \text{and} \quad \frac{p_y}{\rho} = \frac{\partial \Phi}{\partial y}. \qquad (4.20)$$

Similarly the coefficient of dZ gives that

$$\frac{\partial \Phi}{\partial Z} = \frac{p}{\rho H} = \frac{RT}{H} \qquad (4.21)$$

where the second equality follows from the perfect gas law. This is the hydrostatic relation in the new coordinate system.

The mass continuity equation may be transformed directly, but its form in the new coordinates follows most easily by rederiving it from scratch. Consider the total mass $\mathcal{M}$ in a volume $\mathcal{V}$ of (x, y, z) space, which transforms to a volume $\mathcal{V}'$ of (x, y, p) space. We have that

$$\mathcal{M} = \int_{\mathcal{V}} \rho dxdydz = -\int_{\mathcal{V}} g^{-1} p_z dxdydz = -g^{-1}\int_{\mathcal{V}'} dxdydp = (gH)^{-1}p_* \int_{\mathcal{V}'} e^{-Z/H} dxdydZ. \qquad (4.22)$$

Thus the 'density' in (x, y, Z) coordinates is equal to $(gH)^{-1}p_* e^{-Z/H}$, which is independent of time. The fluid therefore appears to be quasi-incompressible. (In (x, y, p) coordinates the fluid appears exactly incompressible since the density is then constant in the vertical.) Applying the integral form of the continuity equation to the volume $\mathcal{V}'$ and then using the divergence theorem and the fact that $\mathcal{V}'$ is arbitrary, it follows that the transformed continuity equation is

$$\frac{\partial}{\partial t}(e^{-Z/H}) + \frac{\partial}{\partial x}(ue^{-Z/H}) + \frac{\partial}{\partial y}(ve^{-Z/H}) + \frac{\partial}{\partial Z}(We^{-Z/H}) = 0, \qquad (4.23)$$

which simplifies to

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + e^{Z/H}\frac{\partial}{\partial Z}(We^{-Z/H}) = 0. \qquad (4.24)$$

In summary, the main simplifications of the transformation to log-pressure coordinates are:

(i) the horizontally varying ρ^{-1} factors multiplying the horizontal pressure gradient disappear (4.20)

and

(ii) the density is no longer a function of time and the continuity equation reduces to a form very close to that for an incompressible fluid (4.24).

The equations in log-pressure coordinates therefore have a form which is almost identical to that of the Boussinesq equations - the sole difference being the appearance of factors of $\exp(-Z/H)$ (which may be approximated as 1 if the height scale is sufficiently shallow). However the vertical coordinate is not geometric height and this must be taken into account when, for example, applying a condition at a rigid lower boundary.

Note, finally, that if the height scale H is taken to be about 7km then there is a reasonable correspondence between Z and geometric height above the Earth's surface.

4.8 The quasi-geostrophic equations

We now obtain a prognostic equation for the slow motion from the Boussinesq primitive equations on a β-plane.

We write $\rho'(x,y,z,t) = \rho_s'(z) + \tilde{\rho}(x,y,z,t)$, where $\rho_s'(z)$ represents density variation in a hydrostatically balanced basic state where there is no motion. We therefore expect that $\tilde{\rho}$ is associated with motion of the fluid when it is disturbed from this resting basic state.

We then write pressure as the sum of two terms, $p'(x,y,z,t) = p_s'(z) + \tilde{p}(x,y,z,t)$, where each term is in hydrostatic balance with the corresponding part of the density field, i.e.

$$\tfrac{dp_s'}{dz} = -\rho_s' g \text{ and } \tfrac{\partial p}{\partial z} = -\tilde{\rho} g \tag{4.25}$$

The density equation (4.8) therefore becomes

$$\frac{D\tilde{\rho}}{Dt} + w\frac{d\rho_s'}{dz} = 0 \tag{4.26}$$

The velocity field is divided into a part that is in geostrophic balance with the pressure field (assuming that the Coriolis parameter is the constant f_0) and a remainder, referred to as the ' ageostrophic' velocity, i.e.

$$\mathbf{u} = \mathbf{u}_g + \mathbf{u}_a \quad \text{where} \quad f_0 \mathbf{k} \times \mathbf{u}_g = -\frac{1}{\rho_0}\nabla_h \tilde{p} \tag{4.27}$$

and ∇_h indicates the horizontal part of ∇.

Note that the vertical component of $\mathbf{u}_g$ is zero, and $\nabla.\mathbf{u}_g = 0$.

We shall also assume that the scale L_y in the y direction is sufficiently small that $\beta L_y/f_0 \ll 1$. Then if $Ro \ll 1$ it follows that $|\mathbf{u}_a| \ll |\mathbf{u}_g|$.

The primitive equations may now be written

$$\left\{\frac{\partial}{\partial t} + (\mathbf{u}_g + \mathbf{u}_a).\nabla\right\}(u_g + u_a) - f_0 v_g - f_0 v_a - \beta y (v_g + v_a) = -\frac{1}{\rho_0}\frac{\partial \tilde{p}}{\partial x} \tag{4.28}$$

$$\left\{\frac{\partial}{\partial t} + (\mathbf{u}_g + \mathbf{u}_a).\nabla\right\}(v_g + v_a) + f_0 u_g + f_0 u_a + \beta y (u_g + u_a) = -\frac{1}{\rho_0}\frac{\partial \tilde{p}}{\partial y} \tag{4.29}$$

$$-\frac{\partial \tilde{p}}{\partial z} - \tilde{\rho} g = 0 \tag{4.30}$$

$$\left\{ \frac{\partial}{\partial t} + (\mathbf{u}_g + \mathbf{u}_a).\nabla \right\} \tilde{\rho} + w_a \frac{d\rho'_s}{dz} = 0 \tag{4.31}$$

$$\frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y} + \frac{\partial w_a}{\partial z} = 0 \tag{4.32}$$

Given that $Ro \ll 1$ we may replace $D\mathbf{u}/Dt$ by $D_g\mathbf{u}_g/Dt$, where $D_g/Dt = \partial/\partial t + \mathbf{u}_g.\nabla$, and $\beta y \mathbf{u}$ by $\beta y \mathbf{u}_g$.

It remains to eliminate $\mathbf{u}_a$, which is done by calculating $\partial(4.29)/\partial x$ - $\partial(4.28)/\partial y$ and then using the nondivergence of the geostrophic velocity, to form a vorticity equation

$$\frac{D_g}{Dt}\left\{ \frac{\partial v_g}{\partial x} - \frac{\partial u_g}{\partial y} \right\} + \beta v_g + f_0 \left\{ \frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y} \right\} = 0. \tag{4.33}$$

(4.31) and (4.32) are now used to eliminate u_a, v_a and w_a to leave

$$\frac{D_g}{Dt}\left\{ \frac{\partial v_g}{\partial x} - \frac{\partial u_g}{\partial y} \right\} + \beta v_g + f_0 \frac{\partial}{\partial z} \left\{ \frac{D_g \tilde{\rho}}{Dt} / \frac{d\rho'_s}{dz} \right\} = 0 \tag{4.34}$$

Defining $\psi = \tilde{p}/\rho_0 f_0$, so that $u = -\psi_y$, $v = \psi_x$ and $\tilde{\rho} = -\rho_0 f_0 \psi_z / g$, and using (4.30) this reduces to

$$\frac{D_g}{Dt}\left\{ \psi_{xx} + \psi_{yy} + \left\{ \frac{f_0^2 \psi_z}{N^2} \right\}_z \right\} + \beta \psi_x = 0 \tag{4.35}$$

or

$$\frac{D_g q}{Dt} = 0 \quad \text{where} \quad q = \left\{ \psi_{xx} + \psi_{yy} + \left\{ \frac{f_0^2 \psi_z}{N^2} \right\}_z + \beta y \right\}. \tag{4.36}$$

Note $N^2 = -g\rho_0^{-1} d\rho'_s/dz$ and that, written in terms of ψ,

$$\frac{D_g}{Dt} = \frac{\partial}{\partial t} - \psi_y \frac{\partial}{\partial x} + \psi_x \frac{\partial}{\partial y}. \tag{4.37}$$

(4.36) is the *quasigeostrophic potential vorticity equation*. The quantity q is the *quasi-geostrophic potential vorticity*. Under the quasi-geostrophic approximation it is conserved following the (horizontal) geostrophic flow.

It is possible show that (4.35) or (4.36) are approximations to the statement of material conservation of Rossby-Ertel potential vorticity following the flow along ρ' surfaces (in a Boussinesq fluid) or θ surfaces (in a compressible fluid).

Note the physical interpretation of different contributions to the quasi-geostrophic potential vorticity q:

$$q = \underbrace{\psi_{xx} + \psi_{yy}}_{\text{relative vorticity}} + \underbrace{\left\{ \frac{f_0^2}{N^2} \psi_z \right\}_z}_{\text{stretching term}} + \underbrace{\beta y}_{\text{planetary vorticity}} \tag{4.38}$$

Note that the stretching term measures vertical gradients in density perturbations, hence the amount by which nearby density surfaces move apart or together.

Just as for shallow-water quasi-geostrophic flow, we have an explicit form of the inversion operator relating ψ to q.

$$\psi = \left\{ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial}{\partial z}\right) \right\}^{-1} [q - \beta y] \tag{4.39}$$

Note that this inversion operator requires boundary conditions on ψ or its derivatives.

(i). At rigid horizontal boundaries we typically require than the normal component of $\mathbf{u}$ is zero, equivalent to ψ being constant along the boundary.

(ii). At rigid vertical boundaries we again require the kinematic boundary condition to be satisfied, i.e. $Dz/Dt = w = Dh/Dt$ on $z = h$. Now using (4.31) to substitute for w and retaining only leading-order terms in Rossby number, this condition may be approximated as

$$\frac{D_g}{Dt}\psi_z = -\frac{N^2}{f_0}\frac{D_g}{Dt}h. \tag{4.40}$$

The boundary condition therefore takes the form of a prognostic equation for ψ_z. Note that the physical interpretation of this condition is as a statement of material rate of change of density or temperature (and material conservation in the case where there is no topography).

The density or temperature at horizontal boundaries therefore have similar status in the quasi-geostrophic equations to the quasi-geostrophic potential vorticity in the interior of the flow. Indeed there are some formulations in whic the boundary density or temperature is explicitly incorporated into the quasi-geostrophic potential vorticity, as delta functions localised at the boundaries. This is analogous to the way in which in electrostatics surface charge can either be separated from interior charge, with the surface charge implying a boundary condition for the normal component of the interior electric field, or incorporated within the interior charge distribution, in which case the boundary condition on the 'interior' electric field are that the normal component is zero at the boundary. For more discussion on this and other topics see the McIntyre article on 'Potential vorticity' in the Encyclopedia of Atmospheric Sciences (available from `http://www.atm.damtp.cam.ac.uk/mcintyre/papers/ENCYC/`.

The form of the operator acting on ψ to give q in (4.38) and hence the inversion operator that acts on q to give ψ suggests a natural ratio of horizontal length scale L to vertical length scale D given by

$$\frac{D}{L} \simeq \frac{f_0}{N} \qquad \text{Prandtl's ratio of scales} \tag{4.41}$$

In a problem where D is set by external factors, e.g. fluid depth, the natural horizontal length scale will be $L = ND/f_0$, which is the Rossby radius for this problem. (Note that with continuous stratification defined by buoyancy frequency N the Rossby radius depends on the height scale D.) On the other hand in a problem where L is set by external factors then the natural height scale will be $D = f_0 L/N$, which we might call the Rossby height.

The three-dimensional quasi-geostrophic equations resemble strongly the equations for two-dimensional vortex dynamics, though the difference is more marked than is the case for the shallow-water quasi-geostrophic equation because of the three-dimensional spatial variation. Note that the inversion operator implies that the response in ψ associated with a particular part of the q field is non-local not only in the horizontal (as we expect from two-dimensional vortex dynamics, where for example, the velocity field associated with a point vortex extends away from the position of the vortex) but also in the vertical. Thus the q field at a given level affects the flow field at other levels.

If N is constant in height then the operator appearing in (4.38) is isotropic in scaled co-ordinates x, y, Nz/f_0. However the evolution equations (4.35) or (4.36) are not isotropic since the flow only has components in the horizontal. We might therefore expect solutions of the quasi-geostrophic equations to have some tendency towards isotropy in the scaled co-ordinates just defined, but exact isotropy seems unlikely. Over the last decade or so, some numerical simulations of the quasi-geostrophic equations have suggested strong anisotropy (e.g. McWilliams, J. C., Weiss, J. B., Yavneh, I., 1994, Science, 264, 410–413), and others have not (e.g. Dritschel, D. G., Ambaum, M. H. P., 1997, Q. J. Roy. Meteorol. Soc., 123, 1097–1130).

4.9 A simple three-dimensional quasi-geostrophic flow

We consider the flow associated with a quasi-geostrophic 'point vortex', i.e. we assume $q = \delta(x, y, z)$, where $\delta(x, y, z)$ is the Dirac delta function and solve

$$\psi_{xx} + \psi_{yy} + \left\{ \frac{f_0^2}{N^2} \psi_z \right\}_z = \delta(x, y, z) \tag{4.42}$$

to give the corresponding $\psi(x, y, z)$. We assume that N is constant and that the βy term appearing in (4.38) may be neglected.

We first rescale z by defining $\overline{z} = Nz/f_0$. Then in Cartesians $(x, y, \overline{z})$ the operator on the left-hand side of (4.42) is the three-dimensional Laplacian and we deduce that a solution satisfying the boundary condition $\psi \to 0$ as $|x|$, $|y|$ and $|\overline{z}|$ tend to infinity is

$$\psi(x, y, z) = -\frac{1}{4\pi} \frac{1}{(x^2 + y^2 + \overline{z}^2)^{1/2}} = -\frac{1}{4\pi} \frac{1}{(x^2 + y^2 + N^2 z^2 / f_0^2)^{1/2}} \tag{4.43}$$

It follows that the horizontal velocity components (u, v) are given by

$$(u, v) = \frac{1}{4\pi} \frac{(-y, x)}{(x^2 + y^2 + N^2 z^2 / f_0^2)^{3/2}} \tag{4.44}$$

and the density perturbation is given by

$$\tilde{\rho} = -\frac{f\rho_0}{g} \psi_z = -\frac{f\rho_0}{4\pi g} \frac{z}{(x^2 + y^2 + N^2 z^2 / f_0^2)^{3/2}} \tag{4.45}$$

Away from the point vortex the contributions to the quasi-geostrophic potential vorticity from the relative vorticity and from the stretching term must cancel, but neither is exactly zero, allowing the circulation (both velocity and density anomalies) to extend away.

5 Rossby-wave propagation

5.1 Rossby waves on the β plane

We have already met Rossby-wave propagation in the context of topographic Rossby waves in the shallow water system. Rossby waves may also result from the variation in potential vorticity associated with the β-effect.

For example, consider the shallow-water quasigeostrophic equations on a β-plane, where a βy term is added to the expression for P_g on the right-hand side of (3.36). If we linearise (3.37) about a state of rest, with this modified expression for P_g, we obtain

$$\left(\psi_{xx} + \psi_{yy} - \frac{\psi}{L_R^2}\right)_t + \beta\psi_x = 0. \tag{5.1}$$

Now seeking plane wave solutions of the form $\psi = \mathrm{Re}(\hat{\psi}e^{ikx+ily-i\omega t})$ we deduce the dispersion relation

$$\omega = -\frac{\beta k}{k^2 + l^2 + L_R^{-2}}. \tag{5.2}$$

For $\beta > 0$ the phase speed is therefore always negative, i.e. westward.

The components of the group velocity $\mathbf{c}_g$ may be calculated as

$$\mathbf{c}_g = \left(\frac{\partial\omega}{\partial k}, \frac{\partial\omega}{\partial l}\right) = \left(\frac{\beta(k^2 - l^2 - L_R^{-2})}{(k^2 + l^2 + L_R^{-2})^2}, \frac{2\beta kl}{(k^2 + l^2 + L_R^{-2})^2}\right). \tag{5.3}$$

This shows that the group propagation is westward for waves with $k^2 < l^2 + L_R^{-2}$ and eastward for waves with $k^2 > l^2 + L_R^{-2}$. If we take $l = 0$ then long waves ($k < L_R^{-1}$) have group propagation to the west and short waves ($k > L_R^{-1}$) have group propagation to the east.

Note that for given l and L_R there is a maximum value of $|\omega|$, equal to $\beta/2(l^2 + L_R^{-2})^{1/2}$. For $|\omega|$ less than this maximum value there are two possible values of k, one (a short wave) corresponding to eastward group propagation and the other (a long wave) corresponding to westward group propagation. This has interesting implications for reflection from boundaries. A long wave (with potential vorticity dominated by the displacement to the free surface) propagating westward to encounter a boundary will be reflected as a short wave (with potential vorticity dominated by relative vorticity). This is part of the explanation for western boundary currents in the ocean.

The group propagation is northward if $kl > 0$ and southward if $kl < 0$. Note that $kl > 0$ implies that phase crests slope westward as y increases (i.e. they slope NW-SE). Similarly $kl < 0$ implies that phase crests slope NE-SW.

For future reference it is useful to note the correlation between u and v in a propagating Rossby wave. We consider $\overline{uv}$ where $\overline{(.)}$ means average in x and u and v are the velocity components associated with the wave. Then

$$\overline{uv} = \mathrm{Re}(ik\hat{\psi} \times il\hat{\psi}^*) = -kl|\hat{\psi}|^2. \tag{5.4}$$

Note that $\overline{uv} < 0$ for waves propagating northward and $\overline{uv} > 0$ for waves propagating southward. $\overline{uv}$ represents a *momentum flux*.

5.2 Vertical modes

Now return to the three-dimensional quasi-geostrophic equations. We consider an ocean with free surface at $z = 0$ in the undisturbed state, with a flat bottom at $z = -H$ and with buoyancy frequency $N(z)$.

Assume that in the disturbed state the height of the free surface is displaced to $z = \eta(x, y, t)$. If we assume that disturbances are small then we may estimate $p(x, y, 0, t) = p_{atm} + \rho_0 g\eta$ and hence $\rho_0 gw = D_g\tilde{p}(x, y, 0, t)/Dt$ at $z = 0$. Now using the expression for the vertical velocity under the quasi-geostrophic approximation, it follows that the boundary condition at $z = 0$ is

$$\frac{D_g\tilde{p}_z}{Dt} - \frac{1}{\rho_0}\frac{d\rho_0}{dz}\frac{D_g\tilde{p}}{Dt} = \frac{D_g\tilde{p}_z}{Dt} + \frac{N^2}{g}\frac{D_g\tilde{p}}{Dt} = 0. \tag{5.5}$$

At $z = -H$, the boundary condition is

$$\frac{D_g\tilde{p}_z}{Dt} = 0. \tag{5.6}$$

Now consider disturbances about a state of rest. The quasi-geostrophic potential vorticity equation is

$$\left(\psi_{xx} + \psi_{yy} + (\frac{f_0^2}{N^2}\psi_z)_z\right)_t + \beta\psi_x = 0 \tag{5.7}$$

and the linearised boundary conditions are $\psi_{zt} + (N^2/g)\psi_t = 0$ at $z = 0$ and $\psi_{zt} = 0$ at $z = -H$.

We now seek solutions of the form $\psi(x, y, z, t) = \phi(x, y, t)P(z)$, where

$$\frac{d}{dz}(\frac{1}{N^2}\frac{dP}{dz}) = -\frac{1}{gh}P \tag{5.8}$$

with h a suitable constant and with boundary conditions $P' + (N^2/g)P = 0$ at $z = 0$ and $P' = 0$ at $z = -H$. This is an eigenvalue equation for h, sometimes called *the vertical structure equation*, and we may expect a countable sequence of possible values $h_1 > h_2 > \cdots > 0$, with the maximum value h_1 corresponding to the simplest possible structure for $P(z)$.

Note furthermore that the height g/N^2 is typically large compared to the depth H (or the vertical length scale associated with variations in stratification) and therefore the boundary condition at $z = 0$ may be approximated by $P' = 0$. This is the so-called *rigid lid approximation*. Solving with this boundary condition gives P non-zero at $z = 0$ and the solution may therefore be used to give an good first estimate of the pressure variation at $z = 0$ and hence the variation in free-surface height.

If $N = N_0$ (constant) then the largest value h_1 is $N_0^2H^2/g\pi^2$, i.e. $(gh_1)^{1/2} = N_0H/\pi$. $P_1(z)$ for this case has a single zero in the interior of the layer. $P_1'(z)$, corresponding to the vertical displacement, has a single maximum in the interior of the layer. This corresponds to the *first baroclinic mode*. For realistic oceanic stratification, the first baroclinic mode is typically found to have $(gh_1)^{1/2} \simeq 3\text{ms}^{-1}$ and the second baroclinic mode $(gh_2)^{1/2} \simeq 1\text{ms}^{-1}$.

Given h_i and $P_i(z)$ the corresponding equation for $\phi_i(x, y, t)$, describing the horizontal structure of the ith mode will be

$$\left(\phi_{ixx} + \phi_{iyy} - \frac{f_0^2}{gh_i}\phi_i\right)_t + \beta\phi_{ix} = 0, \tag{5.9}$$

i.e. the quasi-geostrophic equation for a single layer of fluid of depth h_i. We have therefore reduced the three-dimensional problem to an equivalent single-layer problem, or a set of such problems, one for each mode, with the layer depths being determined as the eigenvalues of the vertical structure equation.

Note that for each vertical mode there is a corresponding Rossby radius of deformation, given by

$$L_{iR} = \frac{(gh_i)^{1/2}}{f_0}. \tag{5.10}$$

For the first baroclinic mode $L_{1R} \simeq 30\text{km}$. For the second $L_{2R} \simeq 10\text{km}$. (These are both estimates for midlatitudes.)

For scales larger than L_R then (5.3) implies that the phase and group velocities are westward and given by βL_R^2. (Note that the waves are non-dispersive in this limit.) This implies that at midlatitudes the phase/group speed for the first baroclinic mode Rossby wave is about 1.5 $\times 10^{-2}$ ms^{-1}. (This implies about 10 years to cross the Atlantic Ocean.)

At latitude λ the first baroclinic mode Rossby wave speed is $gh_1 \cos\lambda/2\Omega a \sin^2\lambda$. The speed of propagation therefore increases towards the equator. (This formula clearly breaks down as the equator is approach and the correct value applying close to the equator may be deduced from the dispersion relation for equatorial Rossby waves.)

This vertical-mode decomposition is most relevant to oceanic Rossby waves. Oceanic Rossby waves have now been clearly observed from satellite observations of sea-surface height, e.g. Chelton, D. B., Schlax, M. G., 1996, Science, 272, 234–238. However there is some disagreement between the wave speeds predicted by the simple theory above and those observed, e.g. Killworth, P. D., Chelton, D. B., DeSzoeke, R. A., 1997, J. Phys. Oceanogr., 27, 941-962. Oceanic Rossby waves are an important mechanism for propagation of information in the ocean (on time scales of years).

5.3 Topographically forced Rossby waves in a flow with vertical shear

Consider a basic flow $u_0(z)$ with associated potential vorticity distribution $(\beta - f_0^2 u_0''(z)/N^2)y$ in a channel with walls at $y = 0$, $y = L$. (We assume that N^2 is constant.) Note that basic flow is itself a solution of the equations of motion.

Now consider a topographic perturbation, so that lower boundary condition is

$$\frac{Dz}{Dt} = \frac{Dh}{Dt} \text{ at } z = 0 \tag{5.11}$$

If the topographic perturbation is small, this linearises to

$$w = u_0 \frac{\partial h}{\partial x} \text{ at } z = 0 \tag{5.12}$$

and in quasi-geostrophic theory becomes

$$-\frac{f_0}{N^2} u_0 \frac{\partial^2 \psi}{\partial x \partial z} = u_0 \frac{\partial h}{\partial x} \text{ at } z = 0 \tag{5.13}$$

In the interior we have the quasi-geostrophic PV equation, linearised about the basic state. We use the notation ψ' to denote the disturbance from the basic state, i.e. $\psi = \psi_0 + \psi'$, etc.

$$\left\{\frac{\partial}{\partial t} + u_0 \frac{\partial}{\partial x}\right\} \left\{\psi'_{xx} + \psi'_{yy} + \frac{f_0^2}{N^2}\psi'_{zz}\right\} + \left(\beta - \frac{f_0^2}{N^2}u_0''(z)\right)\psi'_x = 0 \tag{5.14}$$

and the side boundary conditions are $\partial\psi'/\partial x = 0$ at $y = 0$ and $y = L$.

For convenience we assume that the topographic perturbation h is sinusoidal in x and has a simple sinusoidal form in y so that we may write

$$h = \mathrm{Re}\left(\hat{h}e^{ikx} \sin\frac{\pi y}{L}\right). \tag{5.15}$$

The form of steady disturbances forced by the topography may be written in the form $\psi' = Re\left(\hat{\psi}(z)e^{ikx} \sin \pi y/L\right)$ where

$$\frac{f_0^2}{N^2}\frac{d^2\hat{\psi}}{dz^2} - \left(k^2 + \frac{\pi^2}{L^2}\right)\hat{\psi} + \frac{(\beta - f_0^2 u_0''(z)/N^2)}{u_0(z)}\hat{\psi} = 0, \tag{5.16}$$

i.e. $\hat{\psi}'' + m(z)^2\hat{\psi}$ or $\hat{\psi}'' - \mu(z)^2\hat{\psi} = 0$ where

$$m(z)^2 \text{ or } -\mu(z)^2 = \left\{\frac{\beta - f_0^2 u_0''(z)/N^2}{u_0(z)} - \left(k^2 + \frac{\pi^2}{L^2}\right)\right\}\frac{N^2}{f_0^2} \tag{5.17}$$

Solutions are wave-like in the vertical if $(m(z))^2 > 0$, i.e. if

$$0 < u_0(z) < \frac{\beta - f_0^2 u_0''(z)/N^2}{k^2 + \pi^2/L^2} = U_c, \tag{5.18}$$

otherwise they have exponential behaviour in the vertical.

Note that:

(i). Disturbances are trapped in the vertical if $u_0 < 0$ or $u_0 > U_c$;

(ii). Disturbances can propagate up into westerlies (eastward flow) providing that these are sufficiently weak;

(iii). U_c is a decreasing function of k. If $u_0(z)$ increases upwards, then the longest waves (with smallest k) will propagate through the greatest range of heights.

These results were first noted by Charney and Drazin in 1961. They are highly relevant to the circulation in the stratosphere, which may be disturbed from a symmetric state, where the flow is around latitude circles, by large-scale Rossby waves that are forced in the troposphere (by flow over topography and other processes) and, under suitable conditions, propagate up into the stratosphere.

Particular implications of the points above are as follows.

(i). appears to explain the disturbed circulation in the winter stratosphere – compared to the summer stratosphere;

(ii). may explain why the waves are so weak in the midwinter southern hemisphere stratosphere, where the winds are very strong;

(iii). explains the increased scale of the waves with height in the winter troposphere and stratosphere.

Note that in vertically trapped waves, with $\hat{\psi} \propto \exp(-\int^z \mu(z')dz')$, ψ' (pressure) and ψ'_z (density) are in phase. There is little generation of vorticity by tilting and the structure is referred to as *barotropic*.

We now consider the correlation between the north-south component of the wave velocity, $v' = \psi'_x$ and the wave density $\rho' = -f_0\psi'_z/g$. (We drop the $\tilde{\rho}$ notation.) This correlation represents a northward *density flux*, just as the correlation between u' and v' represents a northward momentum flux (strictly a northward flux of the x-component of momentum).

For vertically trapped waves the northward density flux is given by

$$\overline{v'\rho'} = \tfrac{1}{2}\rho_0 f g^{-1}\mathrm{Re}(ik\hat{\psi}\mu\hat{\psi}^*) = 0. \tag{5.19}$$

The northward density flux is zero.

Vertically propagating waves have $\hat{\psi} \propto \exp(\pm i\int^z m(z')dz')$. Which sign do we take in the exponential? This may be determined from the dispersion relation

$$\omega = ku_0 - \frac{\{\beta - f_0^2 u_0''(z)/N^2\}\, k}{k^2 + \pi^2/L^2 + m^2 f_0^2/N^2}, \tag{5.20}$$

which gives that the vertical component of the group velocity is

$$\frac{\partial\omega}{\partial m} = \frac{f_0^2}{N_0^2}\frac{2mk\,\{\beta - f^2 u_0''(z)/N^2\}}{k^2 + \pi^2/L^2 + m^2 f_0^2/N^2} \tag{5.21}$$

which is greater than 0 if $m > 0$ (under the assumption that $k > 0$. Given that the waves are being forced from below, we therefore chose the positive sign in the exponential so that the group propagation is upwards. Note that in a vertically propagating wave pattern ψ and ψ_z are $\pi/2$ out of phase. There is substantial generation of vorticity by tilting and the structure is referred to as *baroclinic*.

Note that in an upward propagating wave $\overline{v'\rho'} = \frac{1}{2}\rho_0 f g^{-1}\mathrm{Re}(ik\hat{\psi}\times im\hat{\psi}^*) = -\frac{1}{2}\rho_0 f g^{-1}km|\hat{\psi}|^2 < 0$. The northward density flux is negative, i.e. there is an southward flux of heavy (cold) fluid, or a northward flux of light (warm) fluid.

Important general points are that:

- vertically propagating Rossby waves are associated with a horizontal flux of heat or density.
- horizontally propagating Rossby waves are associated with a horizontal flux of momentum.

The above may be straightforwardly modified to include the decrease of density with height using the log-pressure coordinate (see §4.7) and noting that the corresponding form of the quasi-geostrophic potential vorticity is

$$q = \psi_{xx} + \psi_{yy} + e^{Z/H}(\frac{f_0^2}{N^2}e^{-Z/H}\psi_Z)_Z + \beta y. \tag{5.22}$$

The effect is to change the critical velocity U_c a little and to give an exponential increase in the amplitude of propagating waves with height.

5.4 Energy, wave energy and wave activity

We now turn to the general issue of quantifying Rossby wave propagation, without assuming, for example, any specific basic flow.

First we need to write down an energy conservation principle for the quasi-geostrophic equations. We form this by multiplying the quasi-geostrophic potential vorticity equation by ψ.

$$0 = \psi\frac{D_g q}{Dt} = \psi\{\frac{\partial q}{\partial t} + J(\psi, q)\} \tag{5.23}$$

$$= \psi\frac{\partial}{\partial t}\{\psi_{xx} + \psi_{yy} + (\frac{f_0^2}{N^2}\psi_z)_z\} + J(\psi, q\psi) \tag{5.24}$$

$$= (\psi\psi_{xt})_x + (\psi\psi_{yt})_y + (\psi\frac{f_0^2}{N^2}\psi_{zt})_z - \psi_x\psi_{xt} - \psi_y\psi_{yt} - \frac{f_0^2}{N^2}\psi_z\psi_{zt} + J(\psi, q\psi) \tag{5.25}$$

Hence

$$\frac{\partial}{\partial t}(\tfrac{1}{2}\psi_x^2 + \tfrac{1}{2}\psi_y^2 + \tfrac{1}{2}\frac{f_0^2}{N^2}\psi_z^2) + \frac{\partial}{\partial x}(-\psi\psi_{xt} + \psi_y q\psi) + \frac{\partial}{\partial y}(-\psi\psi_{yt} - \psi_x q\psi) + \frac{\partial}{\partial z}(-\frac{f_0^2}{N^2}\psi\psi_{zt}) = 0. \tag{5.26}$$

This is a conservation equation for the quantity whose density is $\frac{1}{2}\psi_x^2 + \frac{1}{2}\psi_y^2 + \frac{1}{2}f_0^2\psi_z^2/N^2$. (Note the isotropy of this expression in suitbaly scaled coordinates.) Since $\frac{1}{2}\rho_0(\psi_x^2 + \psi_y^2) = \frac{1}{2}\rho_0(u^2 + v^2)$ is the kinetic energy density, we interpret the equation as ($\rho_0^{-1}\times$) an energy conservation equation. The quantity $\frac{1}{2}\rho_0 f_0^2\psi_z^2/N^2$, which may be rewritten as $-\frac{1}{2}g\tilde{\rho}^2/(d\rho_0/dz)$ and related to the vertical displacements of density surfaces, we interpret as a potential energy density, strictly a density of *available potential energy.* We can derive this expression from first principles, will not do so here. For more details see §6.10 of Pedlosky, §5.3 of James or §7.8 of Gill.

We now turn to the problem of quantification of wave amplitudes. We consider small-amplitude disturbances superimposed on a mean state obtained by x-averaging the flow and write each quantity χ as $\chi = \chi' + \overline{\chi}$, where the $\overline{\chi}$ is the x-average and hence χ' is the disturbance part of χ.

We may define a wave energy density

$$\mathcal{E}_w = \rho_0\left(\tfrac{1}{2}(\psi_x')^2 + \tfrac{1}{2}(\psi_y')^2 + \tfrac{1}{2}\frac{f_0^2}{N^2}(\psi_z')^2\right) \tag{5.27}$$

However, for a general basic state (in particular where there is a non-trivial basic flow) this wave energy is not conserved, but satisfies an equation of the form

$$\frac{\partial \mathcal{E}_w}{\partial t} + \nabla \cdot \mathbf{F}_w = \mathcal{C} \tag{5.28}$$

where $\mathcal{C}$ cannot be put in the form of a divergence and may be written as

$$\mathcal{C} = -\overline{\psi}_{yy}\psi'_x\psi'_y - \overline{\psi}_{zy}\psi'_x\psi'_z = -\overline{u}_y u'v' - \overline{u}_z \frac{g}{f_0\rho_0}\rho'v' \tag{5.29}$$

Note for example, that if $\overline{u}_z > 0$ and the waves are propagating upwards, $\overline{\mathcal{C}} > 0$ and there will tend to be an increase in the globally integrated wave energy. $\mathcal{C}$ represents a conversion from mean-state energy to wave energy. Note that this conversion may be entirely reversible, e.g. waves may propagate through one region where they gain energy from the basic flow and then through another region where they lose energy to the basic flow.

Wave energy is an unsatisfactory diagnostic in many ways. For example it is difficult to distinguish between phenomena that involve wave propagation on a shear flow, (e.g. the Charney-Drazin problem with basic state velocity increasing upwards) and one that involves systematic growth of disturbances as a result of instability of the basic state (see later in course). Can we find an alternative wave quantity which is conserved in the sense that there are no conversion terms?

Consider the linearised quasi-geostrophic potential vorticity equation

$$\frac{\partial q'}{\partial t} + \overline{u}\frac{\partial q'}{\partial x} + v'\overline{q}_y = 0. \tag{5.30}$$

Note that

$$v'q' = \psi'_x\left(\psi'_{xx} + \psi'_{yy} + (\psi'_z\frac{f_0^2}{N^2})_z\right) \tag{5.31}$$

$$= \left(\tfrac{1}{2}\psi_x'^2\right)_x + \left(\psi'_x\psi'_y\right)_y - \left(\tfrac{1}{2}(\psi'_y)^2\right)_x + \left(\psi'_x\psi'_z\frac{f_0^2}{N^2}\right)_z - \left(\tfrac{1}{2}(\psi'_z)^2\frac{f_0^2}{N^2}\right)_x, \tag{5.32}$$

so multiplying by $q'/\overline{q}_y$ and neglecting time derivatives of the mean state, it follows that

$$\frac{\partial}{\partial t}\left(\frac{\frac{1}{2}q'^2}{\overline{q}_y}\right) + \frac{\partial}{\partial x}\left(\overline{u}\frac{\frac{1}{2}q'^2}{\overline{q}_y} + \tfrac{1}{2}(\psi'_x)^2 - \tfrac{1}{2}(\psi'_y)^2 - \tfrac{1}{2}(\psi'_z)^2\frac{f_0^2}{N^2}\right) + \frac{\partial}{\partial y}\left(\psi'_x\psi'_y\right) + \frac{\partial}{\partial z}\left(\psi'_x\psi'_z\frac{f_0^2}{N^2}\right) = 0, \tag{5.33}$$

i.e., $\partial\mathcal{A}/\partial t + \nabla\cdot\mathbf{F} = 0$, for suitably defined $\mathcal{A}$ and $\mathbf{F}$. This equation is called the Eliassen-Palm wave-activity relation.

Thus $A = \int q'^2/\overline{q}_y dxdydz$ is conserved, and the flux of A in the y-direction is $-u'v'$ and in the z-direction is $-(gf_0/N^2\rho_0)v'\rho'$. These expressions may be compared with earlier results on correlations between u' and v' and between ρ' and v' obtained by considering plane-wave solutions. Here no plane-wave assumption has been made. It may be shown that in the plane-wave limit $\mathbf{F}$ reduces to $\mathbf{c}_g\mathcal{A}$ (where $\mathbf{c}_g$ is the group velocity). The flux $\mathbf{F}$ is therefore said to satisfy the *group-velocity property.*

The Eliassen-Palm flux $\mathbf{F}$ has been calculated from observational atmospheric data and used as a quantitative diagnostic of wave propagation in the real atmosphere.

[Note that there is implicit non-uniqueness in the definition of the density $\mathcal{A}$ and the flux $\mathbf{F}$. The divergence of any flux $\mathbf{G}$ may be added to $\mathcal{A}$ and a corresponding correction $-\mathbf{G}_t$ added to the flux $\mathbf{F}$. Requiring that the group velocity property holds is one way of limiting (but not eliminating) the non-uniqueness.]

If we include non-conservative terms into the original equations (e.g to model the effects of dissipation) then extra terms would appear on the right-hand side of the quasi-geostrophic equation, and correspondingly on the right-hand side of the Eliassen-Palm relation, as a source/sink term.

What is A? Abstract general theories of dynamics tell us that conserved quantities are often associated symmetries of the system, e.g. energy conservation with time symmetry (i.e. time independence) and conservation of a particular component of momentum is associated with translation symmetry in that direction. Conservation of A arises from the invariance under translation in the x-direction, not of the flow itself, but of the waves with respect to the mean flow. Such a conserved quantity is often referred to as a 'pseudomomentum' or 'quasimomentum'.

Example:

Consider the 2-dimensional vorticity equation on a β-plane, with background linear shear flow $(u, v) = (\Lambda y, 0)$. We follow the evolution of disturbances by linearising about the background flow, leaving the following equation for the disturbance streamfunction ψ',

$$\left(\frac{\partial}{\partial t} + \Lambda y \frac{\partial}{\partial x}\right) \nabla^2 \psi' + \beta \frac{\partial \psi'}{\partial x} = 0. \tag{5.34}$$

This equation has solutions of the form

$$\psi' = \mathrm{Re}(\hat{\psi}(t) \exp\{ik(x - \Lambda y t) + ily\}), \tag{5.35}$$

where k and l are constants. It follows that $\hat{\psi}$ must satisfy an ordinary differential equation in time, given by

$$-\frac{d}{dt}[\{k^2 + (l - k\Lambda t)^2\}\hat{\psi}] + i\beta k \hat{\psi} = 0, \tag{5.36}$$

which has the solution

$$\hat{\psi} = \frac{\hat{q}_0}{k^2 + (l - k\Lambda t)^2} \exp\{\frac{i\beta}{k\Lambda} \tan^{-1}[\frac{k\Lambda t - l}{k}]\}. \tag{5.37}$$

Now consider the wave energy density

$$\tfrac{1}{2}\overline{u'^2 + v'^2} = \tfrac{1}{4}[k^2 + (l - k\Lambda t)^2]|\hat{\psi}|^2 = \frac{|\hat{q}_0|^2}{4[k^2 + (l - k\Lambda t)^2]}. \tag{5.38}$$

This ultimately decreases to zero as t increases. The wave energy is not constant in time and there is a conversion from wave energy to the energy in the mean state. This conversion is, in principle, entirely reversible. If the sign of Λ was reversed after a time

τ then the system would run 'backwards' and at time 2τ the value of the wave energy would be the same as that initially.

On the other hand the Eliassen-Palm wave activity density is given by

$$\tfrac{1}{2}\overline{q'^2}/\beta = \tfrac{1}{4}|\hat{q}_0|^2/\beta \tag{5.39}$$

which is constant in time, according to this solution.

(In fact the solution above implies systematically increasing gradients in the y-direction, hence dissipation will become important eventually and $\mathcal{A}$ will decrease. But this decrease is quite different from the decrease in wave energy, which is predicted to occur even in the absence of dissipation.)

6 Wave, mean-flow interaction

6.1 Evolution equations and eddy forcing terms

Returning to the division into 'mean' and 'eddy' parts (rather than 'basic state ' and 'disturbance' parts), we may apply the averaging operator to the Boussinesq β-plane primitive equations to give

$$\overline{u}_t + (\overline{uv})_y + (\overline{uw})_z - \overline{v}(f_0 + \beta y) = 0 \tag{6.1}$$

$$\overline{v}_t + \left(\overline{v^2}\right)_y + (\overline{wv})_z + (f_0 + \beta y)\overline{u} = -\frac{\overline{p}_y}{\rho_0} \tag{6.2}$$

$$-\overline{p}_z - \overline{\rho} g = 0 \tag{6.3}$$

$$\overline{v}_y + \overline{w}_z = 0 \tag{6.4}$$

$$\overline{\rho}_t + (\overline{\rho v})_y + (\overline{\rho w})_z = 0 \tag{6.5}$$

The definition of the averaging operator, together with (6.4), implies that

$$(\overline{uv})_y + (\overline{uw})_z = (\overline{u}\,\overline{v})_y + (\overline{u}\,\overline{v})_z + \left(\overline{u'v'}\right)_y + \left(\overline{u'w'}\right)_z = \overline{v}\,\overline{u}_y + \overline{w}\,\overline{u}_z + (\overline{u'v'})_y + (\overline{u'w'})_z \tag{6.6}$$

Note that to avoid confusion the primes are now used solely to denote disturbances from the x-average, not the pressure and density perturbations associated with the Boussinesq approximations. The primes are dropped from the latter quantities and ρ_0 is still the constant background density.

Now applying small Rossby number scaling as in §4, dividing horizontal velocities into geostrophic and ageostrophic parts and noting that $\overline{v}_g$ is zero, it follows that

$$\overline{u}_t - f_0 \overline{v}_a = -(\overline{u'v'})_y \tag{6.7}$$

$$f_0 \overline{u}_g = -\frac{\overline{p}_y}{\rho_0} \tag{6.8}$$

$$-\overline{p}_z - \overline{\rho} g = 0 \tag{6.9}$$

$$\overline{v}_{ay} + \overline{w}_{az} = 0 \tag{6.10}$$

$$\overline{\rho}_t + \overline{w}_a \frac{d\rho_s}{dz} = -(\overline{\rho' v'})_y. \tag{6.11}$$

We see that the above set of equations for mean quantities is forced by two terms involving eddy quantities, one involving the eddy momentum flux and the other the eddy density (or heat) flux. However the fact that both of these are non-zero does not necessarily mean that the mean flow will change (i.e. $\overline{u}_t \neq 0$ or $\overline{\rho}_t \neq 0$). The equations for $\overline{u}_t$ and $\overline{\rho}_t$ are linked by the quantities $\overline{v}_a$ and $\overline{w}_a$, describing the (ageostrophic) mean circulation in the (y, z) plane and these too should be regarded as part of the response. In principle it is possible for the response to appear *only* in $\overline{v}_a$ and $\overline{w}_a$. This will be the case if

$$\frac{\partial}{\partial y}\left\{ - \left(\overline{u'v'}\right)_y + \left(\frac{f_0 \overline{\rho' v'}}{d\rho_s/dz} \right)_z \right\} = 0 \tag{6.12}$$

Note that the quantity in brackets is exactly the zonal average of $\nabla \cdot \mathbf{F}$, where $\mathbf{F}$ is the Eliassen-Palm flux.

6.2 The transformed Eulerian mean equations

The effect of a particular combination of eddy momentum flux and eddy density flux is made explicit if we make the transformation

$$\overline{w}^*_a = \overline{w} + \frac{\left(\overline{\rho' v'}\right)_y}{d\rho_s/dz} \tag{6.13}$$

and define $\overline{v}^*_a$ such that

$$\overline{w}^*_{az} + \overline{v}^*_{ay} = 0. \tag{6.14}$$

Then

$$\overline{v}^*_a = \overline{v}_a - \frac{\partial}{\partial z}\left[\frac{\overline{\rho' v'}}{d\rho_s/dz} \right] \tag{6.15}$$

and the $\overline{u}$ and $\overline{\rho}$ equations become

$$\overline{u}_t - f_0 \overline{v}^*_a = - \left(\overline{u'v'}\right)_y + \left(\frac{f_0 \overline{\rho' v'}}{d\rho_s/dz} \right)_z = \nabla \cdot \mathbf{F} \tag{6.16}$$

$$\overline{\rho}_t + \overline{w}^*_a \frac{d\rho_s}{dz} = 0 \tag{6.17}$$

These, together with (6.8), (6.9) and (6.14) are the so-called transformed Eulerian mean equations.

The value of the transformation is that it combines the two eddy forcing terms into a single expression making it clear when the eddies will induce a change in the mean flow. Note that the quantities $\overline{u}_t$ and $\overline{\rho}_t$ do not change in going from the original set to the transformed set; what changes is the measure of the mean circulation in the (y, z)-plane.

Note that the horizontal density flux $\overline{\rho' v'}$ appears to play quite a different role in the two formalisms, though we know that its effect on the mean flow $\overline{u}$ is the same. In the

Eulerian-mean formalism $\overline{\rho' v'}$ appears as a forcing term in the density equation. Part of the response to this forcing will be an ageostrophic circulation $(\overline{v}_a, \overline{w}_a)$ which will lead to changes in $\overline{u}$ through the Coriolis force term in (6.7). On the other hand, in the transformed-Eulerian-mean formalism $\overline{\rho' v'}$ appears in (6.16) as if it were contributing to a vertical momentum flux.

In the transformed-Eulerian-mean formalism it is clear that, just as horizontally propagating Rossby waves transfer momentum in the horizontal, vertically propagating Rossby waves can be considered to transfer momentum in the vertical.

6.3 Non-acceleration conditions

The Eliassen-Palm flux (or more strictly its x-average) has thus appeared in two different equations, one for wave activity (5.33), which we write again as

$$\frac{\partial \mathcal{A}}{\partial t} + \nabla \cdot \mathbf{F} = \mathcal{D} \tag{6.18}$$

where we have now explicitly included the term $\mathcal{D}$ representing non-conservative effects (in-situ forcing or dissipation), and one for the mean flow. We note, in particular, that if the waves are steady and non-dissipative, (i.e. $\partial \mathcal{A}/\partial t = 0$ and $\mathcal{D} = 0$) then $\nabla \cdot \mathbf{F} = 0$ and hence there is no acceleration of the mean flow. Such a result is generally called a *non-acceleration theorem*.

The importance of non-acceleration theorems is that they focus attention on the properties that the waves must have if there is to be a mean flow acceleration. [In particular, many of the simplest models of waves assume steadiness and no dissipation – and are therefore useless from the point of view of considering interaction between waves and mean flow.]

Note that all the results derived so far on wave-activity conservation and its link to wave mean-flow interaction have been derived assuming small-amplitude waves. (Equations for wave quantities have been linearised.) It is possible to generalise some of the results to finite-amplitude disturbances, but the link between wave activity and mean-flow evolution is strongest in theories based on Lagrangian rather than Eulerian averages (and this introduces various technical difficulties). (See Andrews, Holton and Leovy §3.7 for more details on this.)

It may be shown that, under quasi-geostrophic scaling

$$\nabla \cdot \overline{\mathbf{F}} = \overline{v\,'q'}, \tag{6.19}$$

the quantity on the right-hand side being the northward flux of quasi-geostrophic potential vorticity. Indeed another statement of the mean flow equations would be via the quasi-geostrophic PV equation, and would take the form

$$\overline{q}_t + \frac{\partial}{\partial y}(\overline{v'q'}) = 0 \tag{6.20}$$

So we might just concentrate on the change in the mean potential vorticity brought about by the eddies and hence deduce the change in the mean velocity and density fields. However, one reason for not doing this is that the other part of the response, the mean circulation, is itself of interest (see later), although it is invisible in a view based solely on potential vorticity. Note that many of these results may be generalised to full primitive equation flow – again see Andrews, Holton and Leovy (1987).

6.4 Wave dissipation and breaking

In the presence of a steady dissipative wave pattern there may be a systematic steady force exerted on the flow. In the absence of overt dissipation, the forcing term appearing in the equations for the mean flow is simply equal to $-\partial\mathcal{A}/\partial t$. If we consider the propagation of a Rossby wave packet then as the wave arrives in a given region there will be a change in the mean flow. $-\partial\mathcal{A}/\partial t$ will be negative, so the effect will be as if a westward force were being exerted. However, when the wave leaves the region, $\partial A/\partial t$ will be positive, an eastward force will be exerted, and the net effect (at least according to our simplified set of equations) will be to leave zero net mean flow change.

In fact the only way in which there can be a net effect is if a wave arrives in a region, but never leaves it. Again, one example of this is the propagation of a wave packet on a shear flow $u = \Lambda y$ in the context of 2DVD on a β-plane. We use a WKB approach, i.e. assume that the y-variation in the basic state occurs on a scale that is much larger than the wavelength in the y-direction.

The dispersion relation is

$$\omega = k\Lambda y - \frac{\beta k}{(k^2 + l^2)} \tag{6.21}$$

and the ray-tracing equations (e.g. Lighthill) thus imply that

$$\frac{dk}{dt} = 0\ , \quad \frac{dl}{dt} = -k\Lambda\ , \quad \frac{dy}{dt} = \frac{2\beta kl}{(k^2 + l^2)^2}. \tag{6.22}$$

Thus a wave packet centred at $y = y_0$ at $t = 0$, with wavenumber k_0, l_0, has, after time t, $l = l_0 - k\Lambda t$. It follows, by integrating the group velocity, or directly from the dispersion relation using $d\omega/dt = 0$, that

$$y = y_0 + \frac{\beta}{\Lambda}\left\{\frac{1}{k^2 + (l_0 - \Lambda tk)^2} - \frac{1}{k^2 + l_0^2}\right\}$$

The packet thus stagnates at $y = y_0 - \frac{\beta}{\Lambda(k^2+l_0^2)} = y_*$.

In this case then, we might expect a permanent mean flow change near $y = y_*$, since, disregarding the effect of dissipation, A increases there but does not subsequently decrease.

Note that the y-wavenumber m becomes very large according to this solution, implying very small-scale structure in the (potential) vorticity field. This might be considered a reason for invoking some sort of dissipative mechanism (e.g. diffusion). Indeed the fact that the wave stagnates means that any dissipative mechanism, however weak when measured against the timescale for propagation, will eventually act to dissipate the wave. However the important point is that the force may be exerted on the mean flow long before any dissipation acts and the details of the dissipation may be almost completely irrelevant.

This is an example where the wave may be said to have 'broken' in the sense that the material contours that undulate during wave propagation become irreversibly deformed. For *Rossby waves* the relevant material contours are *potential vorticity* contours. If the waves 'break' then they can have a permanent effect on the mean flow, in the absence of or irrespective of dissipation, (since, in simple terms, they do not propagate anyway).

6.5 Wave mean-flow interaction in a Rossby wave critical layer

A useful example of wave dissipation and/or breaking and the resulting effect on the mean flow is provided by the case of the forced Rossby wave on a shear flow. This cannot be solved by WKBJ methods, but is analytically tractable in the small-amplitude case. The waves propagate towards the critical line (where phase speed = flow speed). In this location there is no possible steady linear non-dissipative balance in the equations. If dissipative processes are relatively strong then the waves dissipate in the region of the critical line. If dissipative processes are relatively weak then the waves break near the critical line. In either case there is a systematic force exerted on the mean flow by the waves in the neighbourhood of the critical line.

Consider steady disturbances forced at some location in $y > 0$ and propagating in the y-direction on a sheared basic flow $U_0(y)$ on a β-plane. Assume that there is vorticity dissipation at constant rate α (e.g. due to linear friction). Defining a basic state streamfunction by $\Psi_0'(y) = -U_0(y)$ and writing

$$\psi = \Psi_0(y) + \tilde{\psi}(x, y, t) \tag{6.23}$$

and

$$q = \beta y + \Psi_0''(y) + \nabla^2\tilde{\psi}(x, y, t) = q_0(y) + \tilde{q}(x, y, t) \tag{6.24}$$

it follows from the two-dimensional vorticity equation on a β-plane that

$$\frac{\partial \tilde{q}}{\partial t} + U_0(y)\frac{\partial \tilde{q}}{\partial x} + J(\tilde{\psi}, \tilde{q}) + B_0(y)\frac{\partial \tilde{\psi}}{\partial x} + \alpha\tilde{q} = 0, \tag{6.25}$$

where $B_0(y) = \beta - U_0''(y)$ and $J(.,.)$ is the Jacobian with respect to x and y.

We shall assume that $U_0''(y)$ is not too large, so that $B_0(y) > 0$. It follows that if $U_0(y) > 0$ in $y > 0$ and $U_0(y) < 0$ in $y < 0$ then the disturbances will take the form of propagating Rossby waves in $y > 0$ and will be spatially decaying in $y < 0$, so that $|\tilde{\psi}| \to 0$ as $y \to -\infty$. We shall therefore also assume that $U_0'(y) > 0$.

Assume that the flow is quasi-steady, the disturbance amplitude is small and the dissipation is weak, so that the dominant balance in (6.25) is between the second and fourth terms. This gives the familiar steady linear non-dissipative Rossby wave equation. Since

$$\tilde{q} \simeq -B_0(y)\tilde{\psi}/U_0(y) \tag{6.26}$$

it follows that the momentum flux $\overline{u'v'}$ satisfies

$$\frac{\partial}{\partial y}(\overline{u'v'}) = \overline{u'q'} \simeq \overline{\tilde{q}\frac{\partial\tilde{\psi}}{\partial x}} \simeq -\frac{B_0(y)}{U_0(y)}\overline{\tilde{\psi}\frac{\partial\tilde{\psi}}{\partial x}} \simeq 0. \tag{6.27}$$

If the balance (6.26) holds the momentum flux $\overline{u'v'}$ is therefore approximately constant and the forcing of the mean flow by the waves is approximately zero (as would have been expected from non-acceleration theorems).

However the balance (6.26) cannot hold everywhere since $U_0(y) = 0$ at $y = 0$. The steady, linear, non-dissipative equations have a singularity at $y = 0$ (and this location is called the *critical line*). It follows that one of the neglected processes in (6.25), i.e. time-dependence, nonlinearity or dissipation, must be of leading order importance near $y = 0$. The critical line singularity is resolved as a finite, but thin, *critical layer*.

Consider first the case where dissipation is the dominant process near $y = 0$. If the x-wavenumber of the disturbances is k, then the relative sizes of the advection and dissipation terms in this region are $kU_0'(0)y$ and α. It follows that dissipation must be included in a region of thickness $\delta_{\text{diss}} = \alpha/kU_0'(0)$. This is the width of the dissipative critical layer.

Consider now the steady linearised equation if dissipation is included. Writing $\tilde{\psi} = \text{Re}(\hat{\psi}(y)e^{ikx})$ we have that

$$(ikU_0(y) + \alpha)(\hat{\psi}_{yy} - k^2\hat{\psi}) + ikB_0(y)\hat{\psi} = 0. \tag{6.28}$$

In the critical layer region near $y = 0$ $|\hat{\psi}_{yy} - k^2\hat{\psi}| \sim k|B_0(0)||\hat{\psi}|/\alpha$, whereas elsewhere in the flow $|\hat{\psi}_{yy} - k^2\hat{\psi}| \sim |B_0||\hat{\psi}|/U_0$. The vorticity near $y = 0$ is therefore much larger, relative to $|\hat{\psi}|$, near to $y = 0$ than elsewhere provided that $\alpha/k|U_0| \ll 1$.

A self-consistent balance of terms in (6.28) is possible if it is $|\hat{\psi}_{yy}|$ that is large, rather than $|\hat{\psi}|$. It follows that the change in $|\hat{\psi}_y|$ moving across the critical layer is of order $kB_0(0)|\psi_h|/\alpha \times \alpha/kU_0'(0) = B_0(0)|\hat{\psi}|/U_0'(0)$, but this is no larger than the typical size of $|\hat{\psi}_y|$ elsewhere in the flow. Thus, at leading-order in $\alpha/k|U_0|$, $\hat{\psi}$ must be continuous across the critical layer.

Now consider the change in momentum flux across the critical layer,

$$\begin{aligned}
[\overline{u'v'}]_-^+ &= \int_-^+ \text{Re}\{ik\hat{\psi}^*(\hat{\psi}_{yy} - k^2\hat{\psi})\}dy = \int_-^+ \text{Re}\{\frac{k^2B_0(y)|\hat{\psi}|^2}{ikU_0(y) + \alpha}\}dy \\
&\simeq k^2B_0(0)|\hat{\psi}(0)|^2\int_-^+ \frac{\alpha}{k^2U_0(y)^2 + \alpha^2}dy \simeq k^2B_0(0)|\hat{\psi}(0)|^2\int_-^+ \frac{\alpha}{k^2U_0'(0)^2y^2 + \alpha^2}dy \\
&\simeq \frac{\pi B_0(0)k|\hat{\psi}(0)|^2}{U_0'(0)}.
\end{aligned} \tag{6.29}$$

Since the disturbances are non-propagating in $y < 0$ it follows that $\overline{u'v'} = 0$ in $y < 0$ and hence that $\overline{u'v'} = \pi B_0(0)k|\hat{\psi}(0)|^2/U_0'(0) > 0$ in $y > 0$, consistent with the fact that the disturbances are propagating in the negative y-direction in this region.

It follows there is a momentum flux divergence out of the critical layer and a negative force exerted on the mean flow in that region. [This negative force may be balanced in the steady state by the mean frictional force.] Note also that, viewing the momentum flux as a wave activity flux, it follows that the critical layer acts as an absorber of wave activity.

We obtain different results if we consider the case where nonlinear advection, rather than dissipation, is important at leading-order in the critical layer. This is a situation where the waves may be said to 'break', rather than dissipate.

Close to the critical line at $y = 0$ the streamfunction may be approximated by

$$\tilde{\psi}(x, y, t) \simeq -\tfrac{1}{2}U_0'(0)y^2 + |\hat{\psi}(0)| \cos kx. \tag{6.30}$$

The corresponding pattern of streamlines are know as Kelvin's cat's eye pattern. Near to $y = 0$ the dynamics is dominated by the advection of q around the streamlines. This advection is nonlinear in the sense that it is not captured by (6.25), which does take account of the fact that vorticity gradients in the y-direction change significantly. The width of the *nonlinear critical layer*, i.e. the region in which the changes in vorticity gradients are significant, is the width of the closed streamline region, $\delta_{\rm nl} \sim (|\hat{\psi}(0)|/|U_0'(0)|)^{1/2}$. Note that there is a nonlinear critical layer if $\delta_{\rm nl} \gg \delta_{\rm diss}$ and a dissipative critical layer if $\delta_{\rm diss} \gg \delta_{\rm nl}$.

As the q field evolves so does the Eliassen-Palm wave-activity in the critical layer. The net wave-activity flux into the critical layer, and hence the jump in momentum flux across the critical layer, evolve in a self-consistent manner, according to the x-average of (5.33) or (6.18) restricted to the two-dimensional case and integrated across the critical layer to give

$$[\overline{u'v'}]_-^+ = \frac{d}{dt}\int_-^+ \overline{\mathcal{A}}dy, \tag{6.31}$$

where $\overline{\mathcal{A}}$ is the x-average of the Eliassen-Palm wave-activity. Note that $\overline{\mathcal{A}}$ is related to the vorticity pertubation in the critical layer. In fact, (5.33) or (6.18) are not quite sufficient; since the flow in the critical layer is nonlinear it is necessary to use a finite-amplitude form of the wave-activity density. However the small-amplitude form of the flux is still valid outside the critical layer.

The evolution of the left-hand side of 6.31 may be followed in detail for a particular analytic solution (the Stewartson-Warn-Warn solution). (See pictures on handout.) For small times $\overline{\mathcal{A}}$ in the critical layer is increasing with time. There is a negative mean force on the critical layer, which acts as an absorber of waves. Later the rate of increase of $\overline{\mathcal{A}}$ slows down and $\overline{\mathcal{A}}$ reachs a maximum. At this instant there is no net force on the critical layer and net wave-activity flux into it, so it is said to act as a reflector. Subsequently the integrated $\overline{\mathcal{A}}$ starts to decrease. The right-hand side of (6.31) is thus negative, hence there is a positive mean force on the critical layer and it acts as a net emitter of wave-activity, i.e. an over-reflector. The Stewartson-Warn-Warn solution shows that the rate of change of the integrated $\overline{\mathcal{A}}$ oscillates in sign, along with the net force on the critical layer and the net wave-activity flux into or out of the critical layer. Integrating (6.31) with respect to time shows that if the integrated $\overline{\mathcal{A}}$ remains bounded then the time-averaged force on the critical layer, or the time-averaged wave-activity flux into it, must be zero, i.e. the critical layer must act as a reflector in the time average.

Irrespective of the complicated absorption-reflection behaviour, it is useful to note that at any instant the change in mean flow in the critical layer may be determined from the mean q distribution. As noted in §6.3, the effect of the waves on the mean flow may be quantified in terms of the potential vorticity flux due to the waves, giving an alternative, but entirely consistent, view of the mean-flow forcing than that obtained from consideration of the momentum (and density) flux. In the nonlinear critical layer there is advective rearrangement of the mean q field, through stirring by the breaking Rossby waves. This implies a negative change in the velocity $\overline{u}$ and is consistent with the negative mean force exerted on the fluid in the critical layer in the first part of its evolution.

If rearrangement is confined to a finite region and vorticity is conserved, we can show that $\int \Delta\overline{u}dy$ is bounded, thus the time-averaged force must tend to zero (averaged over long enough times). Hence the time average of $\overline{u'v'}$ in the region between the forcing and the critical layer must also tend to zero. In the time-average there can therefore be no flux of wave-activity into the critical layer and it must act as a reflector of waves.

The critical-layer problem illustrates some important general principles about wave-propagation and wave mean-flow interaction.

(i). There is 'long-range' transfer of momentum. Propagating waves transfer momentum from the region where they are generated to the region where they dissipate or break.

(ii). Dissipating or breaking waves change the potential vorticity distribution in the region where the dissipation or breaking occurs. This change is not usually consistent with local conservation of momentum in this region, but it is consistent with the long-range transfer of momentum by the waves into or out of the region.

7 Instability of geophysical flows

We have discussed the behaviour of forced disturbances, usually in the context of the linearised equations. A tacit assumption has been that there are no unforced solutions of the equations that grow without bound (thereby disrupting the forced solutions and perhaps the entire flow).

Given a steady solution of the equations of motion it is useful to consider the stability of such a solution, i.e. if a small disturbance is added, does it remain small? Often such considerations involve linearisation of the equations of motion, resulting in linearised equations for the evolution of disturbance. Normal modes may then be sought, i.e. disturbances that have constant spatial form, and whose time dependence may be expressed in terms of a constant frequency (perhaps complex). If the frequency has a non-zero imaginary part, with sign corresponding to temporal growth, then the flow is deduced to be unstable. We shall consider two important paradigms for instability in balanced flows.

7.1 Barotropic or shear instability

The simplest case is in 2-dimensional vortex dynamics. (Extension to the shallow-water quasigeostrophic system is straightforward). We consider a problem first considered by Rayleigh, the stability of a flow with piecewise constant vorticity.

The basic flow is taken to have velocity field $(u_0(y), 0)$, with

$$\begin{aligned} u_0 &= \tfrac{1}{2}\Lambda L \qquad y > \tfrac{1}{2}L \\ &= \Lambda y \qquad |y| < \tfrac{1}{2}L \\ &= -\tfrac{1}{2}\Lambda L \qquad y < -\tfrac{1}{2}L \end{aligned}$$

The basic state vorticity is therefore

$$\begin{aligned} -u_y &= 0 \qquad |y| > \tfrac{1}{2}L \\ &= -\Lambda \qquad |y| < \tfrac{1}{2}L \end{aligned}$$

and the basic state vorticity gradient is

$$-u_{yy} = \Lambda \left\{ \delta(y - \tfrac{1}{2}L) - \delta(y + \tfrac{1}{2}L) \right\}. \tag{7.1}$$

In physical terms, the vorticity field is piecewise-constant, with constant regions being separated by two interfaces, one whose undisturbed position is at $y = \frac{1}{2}L$ and the other whose undisturbed position is at $y = -\frac{1}{2}L$. Given material conservation of vorticity, the entire evolution of the flow may be followed by considering the displacement of these interfaces.

However, we choose to use the disturbance vorticity equation which, linearised about the basic state, is

$$\left\{ \frac{\partial}{\partial t} + u_0 \frac{\partial}{\partial x} \right\} \left(\psi'_{xx} + \psi'_{yy} \right) + \Lambda \left\{ \delta(y - \tfrac{1}{2}L) - \delta(y + \tfrac{1}{2}L) \right\} \psi'_x = 0. \tag{7.2}$$

We seek solutions of the form

$$\psi' = Re(\hat{\psi}(y) e^{ikx - ikct})$$

where c is a complex phase speed (and kc is a complex frequency). Then $\hat{\psi}$ satisfies the equation

$$(u_0 - c)(\hat{\psi}_{yy} - k^2 \hat{\psi}) + \Lambda \left\{ \delta(y - \tfrac{1}{2}L) - \delta(y + \tfrac{1}{2}L) \right\} \hat{\psi} = 0 \tag{7.3}$$

and with the boundary condition

$$\hat{\psi} \to 0 \quad \text{as} \quad y \to \pm\infty$$

defines an eigenvalue problem for c.

Since $\hat{\psi}_{yy} - k^2 \hat{\psi} = 0$ everywhere except $y = \pm\frac{1}{2}L$ it follows that

$$\hat{\psi} = Ae^{-ky} \qquad y > \tfrac{1}{2}L \tag{7.4}$$

$$= Be^{ky} + Ce^{-ky} \qquad |y| < \tfrac{1}{2}L \tag{7.5}$$

$$= De^{ky} \qquad y < -\tfrac{1}{2}L \tag{7.6}$$

with the constants A, B, C and D such that

$$\hat{\psi} \text{ is continuous at } y = \pm\tfrac{1}{2}L \tag{7.7}$$

$$(\tfrac{1}{2}\Lambda L - c) \left[\hat{\psi}_y \right]_{(L/2)^-}^{(L/2)^+} + \Lambda \hat{\psi}(\tfrac{1}{2}L) = 0 \tag{7.8}$$

$$-(\tfrac{1}{2}\Lambda L + c) \left[\hat{\psi}_y \right]_{(-L/2)^-}^{(-L/2)^+} - \Lambda \hat{\psi}(-\tfrac{1}{2}L) = 0. \tag{7.9}$$

These conditions give four equations for the unknowns A, B, C, D. The consistency condition for the four equations to have a non-trivial solution is the eigenvalue equation

$$c = \pm \frac{\Lambda}{2k} \left\{ (1 - kL)^2 - e^{-2kL} \right\}^{1/2}. \tag{7.10}$$

Since the term in braces $\simeq -k^2L^2$ as $kL \to 0$ it follows that $c \to \pm\frac{1}{2}i\Lambda L$ as $kL \to 0$, i.e. there is one mode decaying in time and one mode growing in time. On the other hand as $kL \to \infty$, $c \simeq \pm\frac{1}{2}\Lambda L$ and there are two separate 'neutral' modes each confined to one of the vorticity interfaces. Indeed if the one interface or the other is considered in isolation, the phase speeds of the two corresponding modes are

$$c = \pm(\tfrac{1}{2}\Lambda L - \frac{\Lambda}{2k}) \tag{7.11}$$

and these expressions are good approximations to the phase speeds when there are two interfaces, but kL is large.

As the wavenumber k decreases the two neutral modes, each existing predominantly on one interface, merge at some critical wavenumber $kL \simeq 1.279$. For smaller wavenumber one of the modes is growing, the other decaying, with maximum growth rate at $kL = 0.7968$.

This instability is generally referred to as shear instability. It arises from the horizontal shear in the basic flow and is relevant to laboratory-scale flows, as well as geophysical flows. Note that it follows from the solution for A, B, C and D, that for a growing mode the y component of the EP flux $-\overline{u'v'}$ is given by

$$-\overline{u'v'} = 0 \text{ for } |y| > \tfrac{1}{2}L \tag{7.12}$$

and

$$-\overline{u'v'} = -k^2\mathrm{Im}(C^*B) = \frac{|A|^2kc_ie^{-kL}}{4|\frac{1}{2}\Lambda L - c|^2} = \frac{|D|^2kc_ie^{-kL}}{4|\frac{1}{2}\Lambda L + c|^2} \text{ for } |y| < \tfrac{1}{2}L. \tag{7.13}$$

Note that $-\overline{u'v'} > 0$ takes the same sign as Λ in $|y| < \frac{1}{2}L$.

Indeed if we define the displacements of the vorticity interfaces

$$\eta'_+ = \mathrm{Re}\left(\hat{\eta}_+e^{ikx-ikct}\right) \tag{7.14}$$

$$\eta'_- = \mathrm{Re}\left(\hat{\eta}_-e^{ikx-ikct}\right) \tag{7.15}$$

then it follows that

$$-\overline{u'v'} = \tfrac{1}{4}kc_i\Lambda|\hat{\eta}_-|^2 = \tfrac{1}{4}kc_i\Lambda|\hat{\eta}_+|^2. \tag{7.16}$$

We can interpret this result in terms of wave activity. Recall that the x-averaged Eliassen Palm wave activity density is equal to $\frac{1}{2}\overline{q'^2}/q_{0y}$. If we use the result that the quasi-geostrophic PV is conserved following the fluid motion it follows that

$$q' + q_{0y}\eta' = 0 \tag{7.17}$$

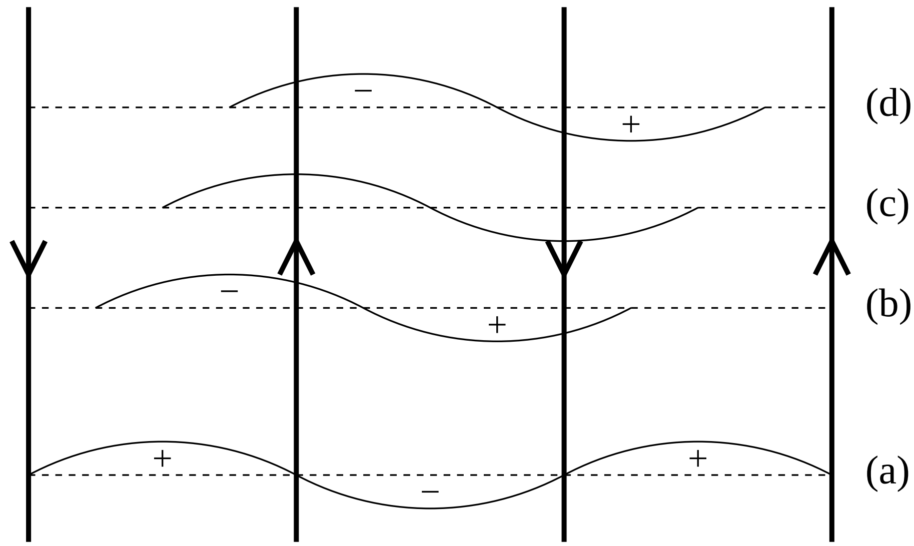

Figure 1: Schematic diagram showing effect of velocity field (shown by large arrows) associated with displacement pattern to vorticity interface at $y = -\frac{1}{2}L$ (a), on the interface at $y = \frac{1}{2}L$ where the phase displacement θ in the pattern on the latter is (b) $0 < \theta < \frac{1}{2}\pi$, (c) $\theta = \frac{1}{2}\pi$, (d) $\frac{1}{2}\pi < \theta < \pi$.

and hence that

$$\tfrac{1}{2}\overline{q'^2}/q_{0y} = \tfrac{1}{2}q_{0y}\overline{\eta'^2}. \tag{7.18}$$

If all the vorticity gradient is concentrated in a thin region, across which the jump in vorticity is Δq and across which η' is essentially constant, it follows that the integral of the wave activity density across that region is $\frac{1}{2}\Delta q\overline{\eta'^2}$.

Noting that $\Delta q = \Lambda$ for the vorticity interface at $y = \frac{1}{2}L$ and $\Delta q = -\Lambda$ for the vorticity interface at $y = -\frac{1}{2}L$, it follows that (7.16) may be interpreted as

$$-\overline{u'v'} = -\frac{\partial}{\partial t}\left\{\text{wave activity on interface at}\, y = -\tfrac{1}{2}L\right\} \tag{7.19}$$

$$= \frac{\partial}{\partial t}\left\{\text{wave activity at}\, y = \tfrac{1}{2}L\right\}. \tag{7.20}$$

The growing normal mode disturbance thus has increasing positive wave activity on one interface, increasing negative wave activity on the other, and there is a corresponding horizontal flux of wave activity between the two interfaces.

In the geophysical context (i.e. three-dimensional rotating stratified flow) there are similar examples of instability where, even though neither the basic state nor the disturbances may be depth independent, the important constrasts in the basic flow are in the horizontal and the dominant wave activity flux due to the disturbance is also horizontal. It is conventional to describe such instabilities as *barotropic* shear instability.

What causes the instability? Investigation of the wave energy shows that there is a conversion of kinetic energy from the mean state, but this doesn't really say why the instability occurs. (We can find flows in which there is plenty of kinetic energy in the basic state, but no instability.) The instability mechanism can be explained as the interaction of two waves, one on each interface. See Figure 1. Instability occurs when these two waves interact in such a way that the velocity field associated with one wave (i.e. induced by its vorticity pattern) causes the other to grow, and vice versa. In principle this is most efficient in configuration (c) shown in Figure 1, but it is also important that the velocity field keeps the displacement pattern fixed against the phase propagation due to advection by the shear flow and with the self-induced propagation on each interface. In case (b) the action on the pattern on the interface at $y = \frac{1}{2}L$ will be to shift it in the positive x-direction. This is needed when the self-induced propagation is relatively strong (when kL is small). In case (d) the action will be to shift the pattern in the negative x-direction. This is needed when the self-induced propagation is relatively weak (when kL is close to the maximum value for instability). The fact that the actual configuration of the displacements on each interface varies in this way as kL varies may be verified from the solution.

7.2 Baroclinic instability

We often consider that fluid-dynamical instabilities tend to give a gain in the energy of disturbances at the expense of the energy of the basic flow. One important instability mechanism in the atmosphere and the ocean involves growth at the expense of the potential energy in the basic state, specifically the potential energy that arises from horizontal density gradients which exist in the basic state when there is vertical shear.

A relevant model problem here is that considered by Eady. We consider a basic state on a f-plane, with flow in the x-direction $u_0 = \Lambda z$, so that $\psi_0 = -\Lambda zy$, with Λ a positive constant. The flow is taken to be bounded above and below by horizontal rigid boundaries at $z = 0$ and $z = H$, with constant buoyancy frequency N_0. It follows that the potential vorticity is constant in the basic state and therefore that there is no potential vorticity advection in the interior. What controls the flow evolution?

Recall that the boundary condition on a horizontal boundary in quasi-geostrophic flow takes the form

$$\psi_{zt} + J(\psi, \psi_z) = 0. \tag{7.21}$$

Linearising about the basic state, it follows that the upper and lower boundary conditions on the disturbance are

$$\psi'_{zt} + u_0\psi'_{xz} + v'\psi_{0yz} = 0. \tag{7.22}$$

so that

$$\psi'_{zt} - \Lambda\psi'_x = 0 \text{ on } z = 0 \tag{7.23}$$

$$\psi'_{zt} + \Lambda H\psi'_{zx} - \Lambda\psi'_x = 0 \text{ on } z = H \tag{7.24}$$

In the interior (since the basic-state potential vorticity is constant) we have

$$\psi'_{xx} + \psi'_{yy} + \frac{f_0^2}{N_0^2}\psi'_{zz} = 0 \tag{7.25}$$

As usual we seek solutions of the form

$$\psi' = \text{Re}\left(\hat{\psi}(z)e^{ikx+ily-ikct}\right) \tag{7.26}$$

Then the disturbance form of the quasigeostrophic potential vorticity equation (4.35) implies that

$$\hat{\psi}_{zz} - \kappa^2\hat{\psi} = 0 \tag{7.27}$$

where $\kappa^2 = N_0^2(l^2 + k^2)/f_0^2$ and hence that

$$\hat{\psi} = Ae^{\kappa z} + Be^{-\kappa z}. \tag{7.28}$$

Substituting into the boundary conditions it follows that

$$-c\,(\kappa A - \kappa B) - \Lambda(A + B) = 0 \tag{7.29}$$

$$(\Lambda H - c)(\kappa Ae^{\kappa H} - \kappa Be^{-\kappa H}) - \Lambda(Ae^{\kappa H} + Be^{-\kappa H}) = 0 \tag{7.30}$$

Combining these equations gives the solution of the eigenvalue problem

$$c = \tfrac{1}{2}\Lambda H \pm \tfrac{1}{2}H\Lambda\left(1 - \frac{4\coth\kappa H}{\kappa H} + \frac{4}{\kappa^2 H^2}\right)^{1/2}. \tag{7.31}$$

Note c is real for large values of κH, and has a non-zero imaginary part for small values of κH. The flow is therefore unstable for some k provided that l is small enough. Note also that, since c is a function of $\kappa H = N_0(k^2 + l^2)^{1/2}H/f_0$, and the growth rate is given by kc_i, it follows that for each value of κ the maximum growth rate occurs for $l = 0$.

The upward EP flux associated with a growing mode may be calculated as

$$\frac{f_0^2}{N_0^2}\overline{\psi'_x\psi'_z} = \tfrac{1}{2}\frac{f_0^2}{N_0^2}\text{Re}\left(ik\hat{\psi}\hat{\psi}_z^*\right) = \tfrac{1}{2}\text{Re}(ik\kappa(A^*B - AB^*)) \tag{7.32}$$

$$= k\kappa\text{Im}(AB^*) = k\kappa\text{Im}\left(AA^*\frac{c^*\kappa + \Lambda}{c^*\kappa - \Lambda}\right) = \frac{2\kappa^2 k|A|^2\Lambda\text{Im}(c)}{|c^*\kappa - \Lambda|^2} > 0 \tag{7.33}$$

So the EP flux is upward and positive for a growing mode.

Thus in a growing disturbance, on a basic state where there is a positive density gradient from equator to pole, there is a poleward flux of light fluid and an equatorward flux of heavy fluid. It follows that there is release of potential energy from the basic state and this is one criterion that is used to identify the instability as *baroclinic instability*. However, the release of energy from the basic state again does not seem to be a complete explanation for the instability. After all, we know that the presence of rotation tends to inhibit the conversion from potential energy to kinetic energy, so just because the potential energy is there does not mean that instability must result.

Returning to the dispersion relation, we note that in the limit $\kappa H \to \infty$, $c/\Lambda H \to 1/\kappa H$ or $c/\Lambda H \to 1 - 1/\kappa H$. In the first case it follows that $B \gg A$, and therefore that the wave is bottom trapped. In the second it follows that $A \sim B$ and hence, taking account that $\kappa H \gg 1$, that the wave is top trapped. We thus have a wave trapped at the bottom boundary, propagating in the positive x-direction against the flow, and another trapped at the top, propagating in the negative x-direction against the flow.

It is useful to consider the case where there is only a lower boundary. Then, putting $\psi' = Re(e^{ikx-ikct+ily}\hat{\psi})$, it again follows from the potential vorticity equation that $\hat{\psi}_{zz} - \kappa^2\hat{\psi} = 0$, where again κ has been defined by $\kappa = N_0(k^2 + l^2)^{1/2}/f_0$. Requiring $\hat{\psi}$ to be bounded as $z \to \infty$ gives that $\hat{\psi} = Ae^{-\kappa z}$. Substituting in the lower boundary condition, $kc\hat{\psi}_z + \Lambda\hat{\psi}k = 0$, it follows that the dispersion relation is

$$c = \frac{\Lambda}{\kappa}. \tag{7.34}$$

So we have an eastward travelling wave trapped at the lower boundary when there is positive vertical shear (or equivalently, a poleward increase in density).

The propagation mechanism for such a wave may be understood in terms of the circulation induced by a surface density change. It is appropriate to regard such a wave as a Rossby wave, but propagating on a surface density gradient, rather than on an interior potential vorticity gradient.

Returning to the case with two boundaries it can be argued that the instability mechanism results from the phase locking of two such waves (one trapped on the top boundary, the other on the bottom boundary) in a configuration such that the velocity field associated with one tends to increase the amplitude of the other, and vice versa.

Note that the ratio between the complex streamfunction amplitude at the top and bottom boundaries is given by

$$\frac{\hat{\psi}(H)}{\hat{\psi}(0)} = \cosh \kappa H - \frac{\Lambda}{\kappa c} \sinh \kappa H. \tag{7.35}$$

It may be shown by considering the argument of this expression that for growing waves the phase of the disturbance velocity on the top boundary leads that on the bottom boundary, by about $\pi/2$ in the case of the wave with the largest growth rate. The influence of the velocity field induced by the upper boundary temperature anomalies on the lower, and vice versa, is to shift the velocity field from being exactly $\pi/2$ out of phase with the temperature anomalies, as is the case with the single boundary, towards being in phase, allowing the whole pattern to grow.

7.3 Wave activity and its use in stability analysis

Consider a basic flow, perhaps with both and horizontal and vertical shear, again confined between rigid horizontal boundaries at $z = 0$ and $z = H$. Integrating the equation (5.33) for linearised EP wave activity over the whole fluid domain, we obtain

$$\frac{d}{dt}\int \tfrac{1}{2}\frac{q'^2}{q_{0y}}dxdydz + \int \left[\psi'_x\psi'_z\frac{f_0^2}{N^2}\right]_0^H dxdy = 0. \tag{7.36}$$

The first integral represents the wave activity in the interior, the second can be rewritten and interpreted as the wave-activity flux out of and into the horizontal boundaries. It has been assumed that the domain is periodic in the x-direction so that contributions from fluxes in the x-direction cancel, and that there are y boundaries on which $\psi'_x = 0$, so that there is no flux across these boundaries.

We may now write the above equation entirely as a rate of change. Multiplying the linearised boundary condition (7.22) by ψ'_z/ψ_{0zy}, we have

$$\frac{\partial}{\partial t}\left\{\tfrac{1}{2}\frac{\psi_z'^2}{\psi_{0zy}}\right\} + \frac{\partial}{\partial x}\left\{\tfrac{1}{2}u_0\frac{\psi_z'^2}{\psi_{0zy}}\right\} + \psi'_x\psi'_z = 0. \tag{7.37}$$

The last term may be used to substitute into the boundary term in (7.36), from which we obtain

$$\frac{d}{dt}\left\{\int \tfrac{1}{2}\frac{q'^2}{q_{0y}}dxdydz + \int \tfrac{1}{2}\frac{f_0^2}{N^2}\left.\frac{\psi_z'^2}{\psi_{0zy}}\right|_{z=0} dxdy - \int \tfrac{1}{2}\frac{f_0^2}{N^2}\left.\frac{\psi_z'^2}{\psi_{0zy}}\right|_{z=H} dxdy\right\} = 0. \tag{7.38}$$

Interpreting the first integral as the total wave activity in the interior, it follows that the second and third may be interpreted as the wave activity residing on the bottom and top boundaries respectively.

The relation just proved may be used to deduce that the flow is stable if

$$\left\{q_{0y}\ ,\quad \psi_{0zy}|_{z=0}\ ,\quad -\psi_{0zy}|_{z=H}\right\}\ \text{are all of the same sign} \tag{7.39}$$

since the quantity inside the time derivative is then of definite sign.

If we are considering normal mode disturbances, i.e. of fixed spatial form multiplied by a factor that is exponentially growing or decaying in time, $\psi' = \mathrm{Re}(\hat{\psi}(z)e^{ikx+ily-ikct})$ then substituting this form into the integrals in (7.38) gives that $c_i \times$ non-zero quantity $= 0$. So $c_i = 0$ and normal-mode instability is ruled out.

More generally, the positive (or negative) definite quantity in (6.6) may itself be used as a measure of the size of the disturbance, defining a 'norm' in the function-analytic sense, and hence the size of the disturbance is bounded.

Conversely the flow can be unstable only if

$$\left\{ q_{0y} \ , \quad \psi_{0zy}|_{z=0} \ , \quad -\psi_{0zy}|_{z=H} \right\} \text{ are not all of the same sign.} \tag{7.40}$$

[Note that $\psi_{0zy} = -U_{0z} = -g\rho_{0y}/\rho_0 f$, i.e. the stability depends on density or temperature gradients at the boundary.] This is a form of the Charney-Stern theorem, which is in turn a quasi-geostrophic generalization of the Rayleigh stability theorem for homogeneous shear flows. The Rayleigh theorem follows from the z-independent form of (7.38) and states that a two-dimensional flow is stable if the vorticity gradient q_{0y} is one-signed.

Note that the Charney-Stern and Rayleigh theorems give necessary conditions for instability, but not sufficient conditions. Showing that a particular flow is actually unstable almost always requires explicit calculation.

7.4 Further remarks on the Eady problem

The idea arises naturally from (7.38) that wave activity may reside at the lower or upper boundaries, provided that the quantity ψ_{0zy} is non-zero there, as well as in the interior. Indeed, the boundary trapped waves that appear in the Eady problem have all their wave-activity at the boundary, and none in the interior.

A growing Eady wave has negative wave activity increasing at the lower boundary and positive wave activity growing at the upper boundary. The requirement that the globally integrated wave activity is constant is therefore satisfied and there is flux of wave activity between lower boundary and the upper boundary.

The Eady problem was first studied as a paradigm for the waves that provide day-to-day and week-to-week variations in the atmospheric flow associated with weather systems. The steady longitudinally symmetric basic state in the Eady problem is not achievable in practice since it is unstable. The real atmosphere has asymmetries at its lower boundary which will perturb the flow from a symmetric state without any need for instability. However, it has been verified from numerical simulations in atmospheric models that even if such asymmetries are removed, the flow is not symmetric. The theoretically possible symmetric flow is indeed unstable.

We may insert realistic atmospheric values into the parameters of the Eady problem, setting H to be 10km, which is the approximate height of the tropopause (the notional boundary between the troposphere and the stratosphere). [The static stability is much larger in the stratosphere than in the troposphere, so the tropopause might in some ways act as a rigid lid for the tropospheric circulation.] It follows that obtain $NH/f_0 \simeq 1000$km, $f_0\Lambda/N \simeq 5 \times 10^{-5}\text{s}^{-1}$ (taking $\Lambda \simeq 5\text{ms}^{-1}\text{km}^{-1}$), giving maximum growth rates of about 1 per day and the wavelength of the fastest growing mode about 4000 km. This seems in reasonable accordance with observed weather disturbances. The instability mechanism in the Eady model (and its relatives) is generally regarded as one of the causes of such disturbances.

7.5 Some general remarks on instability in quasi-geostrophic dynamics

We are left with a general qualitative picture of instability in quasi-geostrophic systems that involves the interaction of two wave modes. These modes would otherwise be localised to a particular region of the flow. It is essential that these modes have opposite-signed wave-activity. In baroclinic instability the regions of localisation are vertically displaced, there is a vertical EP flux between them, and the growing mode is associated with a horizontal heat or density flux. Consideration of energy conversion shows that as the unstable waves grow the potential energy in the basic state is released. In two-dimensional shear instability (or barotropic instability) the EP flux is directed horizontally and the growing mode is therefore associated with a horizontal momentum flux. In this case, as the unstable waves grow the kinetic energy in the basic flow is released.

In a general atmospheric or oceanic flow both heat and momentum fluxes are likely to be important and the instability might be referred to as a mixed barotropic/baroclinic instability (or, more helpfully, as a quasi-geostrophic shear instability).

Such instabilities may involve

- (a) Opposite signed q_{0y} in horizontal – barotropic (e.g. Rayleigh shear instability)
- (b) opposite signed q_{0y} in vertical – internal baroclinic (possible examples relevant in the mesosphere [part of the atmosphere between 50km and 80km in height]))
- (c) $u_{0z} > 0$ at lower boundary, $u_{0z} > 0$ at upper boundary – baroclinic (e.g. Eady problem)
- (d) $u_{0z} < 0$ at upper boundary, $q_{0y} > 0$ in interior – baroclinic instability (e.g. westward flow in low-latitude part of oceanic wind gyres)
- (e) $u_{0z} > 0$ at lower boundary, $q_{0y} > 0$ in interior – baroclinic instability (e.g. Charney problem, another idealised model relevant to the atmosphere)

Note that all of these instabilities involve slow motion, essentially Rossby waves propagating on PV gradients or boundary density gradients. One reason for not accepting energy conversion as a characterisation of instability is that one can conceive of very different instabilities that lead to the same energy conversion. One can for example have instabilities which have very similar energy conversion properties to the above that involve coupling between (slow) Rossby-like waves and (fast) inertio-gravity-like waves.

7.6 Nonlinear behaviour

Linearised equations can only deal with cases where amplitude is small, in the sense that particle displacements are small compared to the length scale of the flow. If a flow is unstable according to linear theory then unbounded exponential growth is predicted. In reality the growth will be limited by nonlinear effects. There has been much work on weakly nonlinear stability theory, with parameters chosen such that weak nonlinear effects are sufficient to cause growth to cease. However, the physical mechanisms operating in such models are, with one or two exceptions, of little direct relevance to the behaviour of the real atmosphere or ocean. Insight into the strongly nonlinear behaviour relevant here has been obtained primarily from direct numerical simulation.

[Examples: Pictures of nonlinear contour-dynamics simulations of the Rayleigh shear layer and of nonlinear Eady model simulations, to give the idea of nonlinear saturation.]

One numerical simulation that has become a standard part of the subject is that of the life cycle of an unstable baroclinic wave growing on a jet similar to the observed atmospheric flow. The life cycle falls in to four stages:

- (i) The basic flow is initialised with a realistic jet structure and a small-amplitude disturbance is added, perhaps with the spatial form of an unstable normal mode, then grows exponentially in time.
- (ii) Nonlinear saturation occurs first at low levels.
- (iii) The remaining part of the disturbance, at mid and upper levels then propagates away as a Rossby wave 'packet', upward and then in latitude, primarily equatorward.
- (iv) The Rossby wave packet reaches regions at low latitudes where the basic flow speed is rather weak. Nonlinearity becomes important and the wave breaks.

These stages are nicely shown in cross-sections of the Eliassen-Palm flux (and indeed were identified using this diagnostic). The tendency for propagation of Rossby wave activity out of the centre of the jet and into low latitudes, is associated with a horizontal momentum flux directed towards the centre of the jet, i.e. up the gradient of momentum.

Observed (angular) momentum fluxes are also observed to be upgradient ($\overline{u'v'}$ has same sign as $\overline{u}_y$). If a standard flux- gradient relationship

$$\overline{u'v'} = -\tilde{\nu}\overline{u}_y, \tag{7.41}$$

where $\tilde{\nu}$ is an eddy diffusivity, were assumed this would require $\tilde{\nu} < 0$. Historically, in the context of a belief that eddies always stir quantities to give down-gradient transport, this phenomenon of 'negative diffusivity' was regarded as highly mysterious. We now understand that it arises as a natural result of Rossby-wave propagation. The surface density and the interior potential vorticity are indeed transported downgradient through local advective rearrangement (in stages (ii) and (iv) respectively) by the growing and saturating wave, but the momentum is not.

The fluxes of momentum, heat and potential vorticity associated with the baroclinic eddies are an important part of the atmospheric general circulation. In order to understand even the longitudinally averaged structure of the circulation it is necessary to take account of the systematic effects of the eddies (defined as the departures from longitudinal symmetry). The effect of the eddy fluxes on the averaged state may be deduced using the procedures outlined in §6. Either the Eulerian-mean or the transformed Eulerian-mean formulations may be used. It should be noticed that the eddy-flux of density enters the lower boundary condition on the transformed Eulerian-mean circulation $(\overline{v}^*, \overline{w}^*)$, whereas the boundary condition on the Eulerian-mean circulation is simply $\overline{w} = 0$. So the transformed Eulerian-mean view is perhaps not 'simpler' than the conventional view in this problem, but it can still be argued that it focusses attention on the important aspects of the eddy behaviour.

8 Mean meridional circulations

8.1 Introduction

We return to the subject of §6, and examine in more detail the mean response of the fluid to wave forcing, comparing the Eulerian-mean and transformed Eulerian-mean viewpoints. In order to solve the Eulerian-mean set of equations (6.7-6.11), or the transformed Eulerian-mean set, (6.16, 6.17, 6.14, 6.8, 6.9), it is useful to eliminate $\overline{u}$, $\overline{\rho}$ and $\overline{p}$ and write the ageostrophic circulation in terms of a stream function, so that

$$(\overline{v}_a, \overline{w}_a) = (\overline{\chi}_{az}, -\overline{\chi}_{ay}) \text{and} (\overline{v}^*_a, \overline{w}^*_a) = (\overline{\chi}^*_{az}, -\overline{\chi}^*_{ay}). \tag{8.1}$$

The ageostrophic circulation in the (y, z) plane is sometimes (particularly in the atmospheric context) called the *mean meridional circulation*.

It follows from the equations that

$$f_0^2 \overline{\chi}_{azz} + N^2 \overline{\chi}_{ayy} = -f_0 (\overline{u'v'})_{yz} - \frac{g}{\rho_0} (\overline{\rho' v'})_{yy} = f_0 \overline{F^{(y)}}_{yz} + \frac{N^2}{f_0} \overline{F^{(z)}}_{yy} \tag{8.2}$$

and

$$f_0^2 \overline{\chi}^*_{azz} + N^2 \overline{\chi}^*_{ayy} = -f_0 (\overline{u'v'})_{yz} + f_0^2 \left(\frac{\overline{\rho' v'}}{d\rho_s/dz} \right)_{zz} = f_0 (\nabla . \overline{\mathbf{F}})_z, \tag{8.3}$$

where $\overline{\mathbf{F}} = (0, \overline{F^{(y)}}, \overline{F^{(z)}})$ is the x-averaged EP flux.

These equations express the forcing of the mean meridional circulation by the eddy fluxes of momentum and density (or, equivalently, by the EP flux divergence).

8.2 Boundary conditions

The equations (8.2) and (8.3) require boundary conditions on $\overline{\chi}$ or $\overline{\chi}^*$. The side boundary condition is usually straightforward, but the bottom boundary condition needs to be derived with care, particularly when the bottom boundary is not flat. Consider for example the case where there is topographic forcing at the lower boundary, so that the full nonlinear boundary condition is

$$w = \frac{Dh}{Dt} \quad \text{at} \quad z = h, \tag{8.4}$$

where $h(x, y, t)$ is the topographic height.

If the topography is small amplitude a Taylor series expansion may be used to express the full boundary condition in terms of quantities at $z = 0$ and it follows that

$$w + hw_z = h_t + uh_x + vh_y + u_z hh_x + v_z hh_y + O(h^3), \tag{8.5}$$

where all quantities on the right-hand side are evaluated at $z = 0$.

The continuity equation may be used to rewrite the last term on the left-hand side as

$$\begin{aligned} w_z &= -hu_x - hv_y \\ &= -(hu)_x - (hv)_y + h_x u + h_y v \end{aligned} \tag{8.6}$$

and substituting into the previous equation it follows that

$$w = h_t + u_z hh_x + v_z hh_y + (hu)_x + (hv)_y + O(h^3). \tag{8.7}$$

Now assuming that the velocity in the absence of topography is purely in the x-direction, so that $v = O(h)$, and that $\overline{h} = 0$ for all t, it follows on taking x-averages that

$$\overline{w}(y, 0, t) = (\overline{h'v'})_y. \tag{8.8}$$

Thus the Eulerian mean velocity is not necessarily zero at $z = 0$. Note that the corresponding boundary condition on the transformed Eulerian mean circulation is that

$$\overline{w}^*(y, 0, t) = (\overline{h'v'})_y + \frac{(\overline{\rho'v'})_y}{d\rho_s/dz}. \tag{8.9}$$

If the waves are steady, and there is no density dissipation at the lower boundary, then the lower boundary condition may be written in the form

$$\overline{u}h'_x \frac{d\rho_s}{dz} = -\overline{u}\rho'_x - \psi'_x \overline{\rho}_y. \tag{8.10}$$

Multiplying by ψ' and averaging, it follows after some manipulation that $f_0 \overline{h'\psi'_x} = -\overline{F^{(z)}}$ and hence that the lower boundary conditions on $\overline{w}$ and $\overline{w}^*$ may be written as

$$\overline{w} = -f_0^{-1}(\overline{F^{(z)}})_y \quad \text{and} \quad \overline{w}^* = 0 \quad \text{on} \quad z = 0. \tag{8.11}$$

Note that under the above conditions it is $\overline{w}^*$ and not $\overline{w}$ that is zero at the lower boundary.

8.3 Example

We now consider a specific example in the small-Ro regime, a flow confined to a β-plane channel with rigid walls at $y = 0$ and $y = L$, with waves forced by topographic perturbations of the lower boundary, of the form $h = \mathrm{Re}(h_0 e^{ikx} \sin \pi y/L)$. The basic state flow is assumed to be in the x-direction, and the velocity a function of height $u_0(z)$. The buoyancy frequency is assumed constant and equal to N_0. The basic state quasi-geostrophic potential vorticity gradient at a height z is therefore $\tilde{\beta}(z) = \beta - (f_0^2 u_0'/N_0^2)'$. At leading-order in h the steady-state wave-field is described by the equation

$$u_0\{\psi'_{xx} + \psi'_{yy} + (\frac{f_0^2 \psi'_z}{N_0^2})_z\}_x + \tilde{\beta}\psi'_x = s' \tag{8.12}$$

where s' has been included as an extra term in the quasi-geostrophic potential vorticity equation resulting from dissipation acting on the waves. It follows from this equation (left as an exercise for the reader) that

$$\nabla.\overline{\mathbf{F}} = (\overline{\psi'_x\psi'_y})_y + (\frac{f_0^2\overline{\psi'_x\psi'_z}}{N^2})_z = -\frac{\overline{s'\psi'}}{u_0}. \tag{8.13}$$

For the particular case of interest we shall assume that the solution ψ' may be written in the form $\psi' = \text{Re}(\hat{\psi}(z)e^{ikx}\sin\pi y/L)$ and it immediately follows that $\overline{\psi'_x\psi'_y} = 0$, so that the EP flux is purely vertical.

We now examine the forcing of the mean meridional circulation by the waves. Note that the appropriate side wall boundary condition is that $\overline{\chi}_a = 0$ and $\overline{\chi}^*_a = 0$ on $y = 0$ and $y = L$. Consider first the case where there is no dissipative term acting on the waves and therefore $\overline{F^{(z)}} = F_0 \sin^2\pi y/L$, where F_0 is a constant. There is a non-zero forcing term in the equation (8.2) for $\overline{\chi}_a$ and a non-zero boundary condition at $z = 0$, and it follows that the solution is $\overline{\chi}_a = f_0^{-1}F_0\sin^2\pi y/L$. Substitution into the Eulerian-mean equations shows that $\overline{u}_t = 0$, $\overline{v}_a = 0$ and $\overline{\rho}_t = 0$. Thus, although there appears to be a non-zero response in the mean quatities, it is characterised entirely by the latitudinal derivative of the eddy density flux balancing vertical advection of density by $\overline{w}_a$. Nothing else happens. On the other hand according to the transformed Eulerian mean equations there is no forcing term in the equation for $\overline{\chi}^*_a$ and $\overline{\chi}^*_a$ is also required to be zero on the boundaries. It is self-evident from the start that nothing is happening and this is arguably the simpler view.

We now consider the case where $s' \neq 0$ and therefore $\overline{F^{(z)}}$ is not independent of z. To avoid the technicalities of solving (8.12) for the structure of the waves, we may choose s' such that $\overline{F^{(z)}}$ is a given function of z. For instance we might assume that the lower boundary is at $z = -D$ and that $\overline{F^{(z)}} = F_0\Theta(z)\sin^2\pi y/L$ where $\Theta(z) = 1$ for $-D < z < 0$ and $\Theta(z) = e^{-\lambda z}$ for $z > 0$. The equations (8.2) and (8.3) may then be solved using Fourier series methods, to determine $\overline{\chi}_a$ and $\overline{\chi}^*_a$.

We write

$$\overline{\chi}^*_a = F_0 f_0^{-1}\sum_{n=1}^{\infty}\chi_n(z)\sin\frac{n\pi y}{L} \tag{8.14}$$

substitute into (8.3) and then use the orthogonality properties to deduce that

$$\chi''_n(z) - \lambda_n^2\chi_n(z) = c_n\Theta''(z) \tag{8.15}$$

where $\lambda_n = nN_0\pi/Lf_0$ and

$$c_n = 8/[\pi n(4-n^2)] \ (n \text{ odd}) \text{ and } c_n = 0 \ (n \text{ even}).$$

The algebra is made easier if we consider the case $D \to \infty$. Solving the ordinary differential equation (8.15) for each n it follows that

$$\overline{\chi}^*_a = \sum_{n=1,n\text{ odd}}\frac{8F_0}{f_0\pi n(4-n^2)}\frac{\lambda}{2(\lambda+\lambda_n)}e^{\lambda_n z}\sin(\frac{n\pi y}{L}) \quad (z<0) \tag{8.16}$$

and

$$\overline{\chi}^*_a = \sum_{n=1,n\text{ odd}}\frac{8F_0}{f_0\pi n(4-n^2)}\left\{\frac{\lambda^2}{(\lambda^2-\lambda_n^2)}e^{-\lambda_n z} - \frac{\lambda}{2(\lambda-\lambda_n)}e^{-\lambda z}\right\}\sin(\frac{n\pi y}{L}) \quad (z>0). \tag{8.17}$$

8.4 Interpretation:

Note that the original problem had three vertical length scales, D, $f_0L/N_0 \sim \lambda_1^{-1}$ (the channel width scaled by Prandtl's ratio) and the decay scale λ^{-1} of $\overline{F^{(z)}}$. By allowing $D \to \infty$ we have implicity taken D to be much larger than the other two scales, so it may be ignored, and there are two important regimes.

(i). (*Deep forcing*) $\lambda \ll \lambda_n$ for all n, i.e. $\lambda f_0 L/N_0 \ll 1$. Here it may be seen from (8.16, 8.17) that $\overline{\chi}_a^* \sim (F_0/f_0)\lambda f_0 L/N \ll 1 \ll F_0/f_0$. It follows that the dominant balance in (8.3) is between the second term on the left-hand side and the right-hand-side and hence that

$$\overline{w}^* \sim \frac{f_0 L \lambda^2 F_0}{N^2} \quad \text{and} \quad \overline{v}^* \sim \frac{f_0 L^2 \lambda^3 F_0}{N^2}. \tag{8.18}$$

Note in particular that the ratio of $f\overline{v}^*$ to $\nabla.\mathbf{F}$ is $f_0^2\lambda^2L^2/N^2$ and therefore from (6.16) most of the response to $\nabla.\mathbf{F}$ appears as an acceleration.

(ii). (*Shallow forcing*) $\lambda \gtrsim \lambda_1$, i.e. $\lambda f_0 L/N \lesssim 1$. It follows from (8.16, 8.17) that $\overline{\chi}^* \sim F_0/f_0$. Here both terms on the left-hand side of (8.3) are equally important and

$$\overline{w}^* \sim \frac{\lambda F_0}{N} \quad \text{and} \quad \overline{v}^* \sim \frac{\lambda F_0}{f_0}. \tag{8.19}$$

The ratio of $f\overline{v}^*$ to $\nabla.\mathbf{F}$ is about 1. Thus as much of the response to $\nabla.\mathbf{F}$ appears in $\overline{v}^*$ and hence in $\overline{w}^*$ and in $\overline{\rho}_t$ as appears as an acceleration. Note also that the circulation decays away from $z = 0$ on a height scale $\lambda_1^{-1} \sim f_0L/N$.

The mean meridional circulation may be regarded as arising in order to maintain, under the effect of eddy forcing, the constraints of geostrophic and hydrostatic balance . Thus if a force is applied to a rotating system, the response cannot appear purely as an acceleration, but there must be an accompanying change in the density field. Broadly speaking, if a force is deep, in coordinates scaled by Prandtl's ratio, then most of the response will appear as acceleration, but if it is shallow, then most of the response will appear as a meridional circulation and hence a density change. Similarly, if an applied heating field is shallow, then most of the response appears as a change in temperature or density, but if it is deep, then most will appear as a meridional circulation, and hence as a change in velocity.

The mean meridional circulation plays an important role in many contexts of rotating, stratified flow, including laboratory-scale flows (the spin-down problem) and the atmospheric general circulation. With regard to the atmosphere, the particular problem studied here (sometimes called the Eliassen problem) is relevant to the response on relatively short timescales. Note for example that $\overline{u}_t$ and $\overline{\rho}_t$ are proportional to the instantaneous eddy forcing. If this forcing is persistent then the changes in $\overline{u}$ and $\overline{\rho}$ will become large and quasigeostrophic scaling may no longer be valid. In addition, we may have to take account of dissipative effects. For example in the stratosphere radiative transfer acts on time scales of a few days and its effects have to be included in the mean flow equations if the time scales of interest are longer than this.

8.5 Effect of thermal and mechanical dissipation

We drop all reference to x-averages and simply consider x-independent motion, with the effect of the eddies represented through a body force $\mathcal{F}$. We assume that the flow is subject to thermal (i.e. buoyancy) damping at rate α and mechanical damping at rate κ. It is convenient to consider the response to forcing at frequency ω and therefore to write $\mathcal{F} = \mathrm{Re}(\hat{\mathcal{F}}e^{i\omega t})$, $u = \mathrm{Re}(\hat{u}e^{i\omega t})$, etc. We also allow an x-independent buoyancy forcing $\mathcal{H}$ which in the atmospheric context might represent e.g. the effects of solar radiation.

The equations governing the system are

$$i\omega\hat{u} - f_0\hat{v} = -\kappa\hat{u} + \hat{\mathcal{F}}, \tag{8.20}$$

$$f_0\hat{u} = -\frac{\hat{p}_y}{\rho_0}, \tag{8.21}$$

$$\frac{g\hat{\rho}}{\rho_0} = -\frac{\hat{p}_z}{\rho_0}, \tag{8.22}$$

$$i\omega\hat{\rho} + \hat{w}\frac{d\rho_s}{dz} = \hat{\mathcal{H}} - \alpha\hat{\rho}, \tag{8.23}$$

and

$$\hat{v}_y + \hat{w}_z = 0. \tag{8.24}$$

The last equation implies that there exists a $\hat{\chi}$ such that $\hat{\chi}_z = \hat{v}$ and $\hat{\chi}_y = -\hat{w}$.

These equations may be combined into a single equation for $\hat{\chi}$ of the form

$$\hat{\chi}_{zz} + \frac{N^2}{f_0^2}\frac{i\omega+\kappa}{i\omega+\alpha}\hat{\chi}_{yy} = \frac{g\hat{\mathcal{H}}_y}{f_0^2\rho_0}\frac{i\omega+\kappa}{i\omega+\alpha} + \frac{\hat{\mathcal{F}}_z}{f_0}. \tag{8.25}$$

Comparing with (8.2) or (8.3) we note that including thermal and mechanical dissipation has altered the coefficients in the elliptic operator acting on χ. In particular the natural aspect ratio between the horizontal and vertical length scales, say $L^{(y)}$ and $L^{(z)}$, is now

$$\frac{L^{(y)}}{L^{(z)}} = \frac{N}{f_0}\left(\frac{i\omega+\kappa}{i\omega+\alpha}\right)^{1/2}. \tag{8.26}$$

This has important implications for the nature of the response to a force of given shape. For example, consider the steady case $\omega = 0$. Then if $\alpha/\kappa \gg 1$, i.e. thermal dissipation is much greater than mechanical dissipation, then $L^{(y)}/L^{(z)} = (N/f_0)(\kappa/\alpha)^{1/2} \ll (N/f_0)$,, i.e. a force that in the absence of dissipation might have been 'deep' (and therefore be balanced locally by an acceleration) might now be effectively 'shallow' (and be balanced by the Coriolis force due to a meridional circulation). The reason is that the temperature signal associated with the forcing must be maintained against thermal damping by a vertical velocity. Correspondingly the response to a given forcing must be taller, so that proportionally less of the total response appears as a temperature signal and more as a velocity signal.

Similarly if $\alpha/\kappa \ll 1$ then $L^{(y)}/L^{(z)} = (N/f_0)(\kappa/\alpha)^{1/2} \gg (N/f_0)$,, i.e. a force that in the absence of dissipation was shallow may now be deep (and will be balanced locally by mechanical friction rather than by the Coriolis force).

8.6 The seasonal cycle

We consider the meridional circulation forced by time-varying heating alone (representing of seasonally varying solar heating). A (very) simple model might be to choose

$$\frac{g\hat{\mathcal{H}}}{\rho_0} = \hat{J}_0 \cos(\frac{\pi y}{L}) \sin(\frac{\pi z}{D}) \tag{8.27}$$

in the region $0 < z < D$ and $0 < y < L$ (with $\chi = 0$ on the boundaries). It follows from (8.3) and the boundary conditions that

$$\hat{\chi} = \left\{ \frac{\pi^2}{D^2} + \frac{N^2}{f_0^2} \frac{i\omega + \kappa}{i\omega + \alpha} \frac{\pi^2}{L^2} \right\}^{-1} \frac{\pi \hat{J}_0}{L f_0^2} \frac{i\omega + \kappa}{i\omega + \alpha} \sin(\frac{\pi y}{L}) \sin(\frac{\pi z}{D}). \tag{8.28}$$

Comparing terms in (8.23) it may be seen that the ratio between the second term on the left-hand side, associated with vertical advection, and the first term on the right-hand side, associated with the thermal forcing, is $1/\{1 + [L^2 f_0^2 (i\omega + \alpha)/D^2 N^2 (i\omega + \kappa)]\}$.

In most of the atmosphere it reasonable to assume that thermal dissipation time scale is a few days and so much less than the seasonal time scale. Thus $\omega \ll \alpha$. If we also assume that $L f_0 \sim DN$, then the ratio above is small if $\kappa/\alpha \ll 1$. The dominant balance in (8.23) is then between the two terms on the right-hand side, i.e. most of the heating is balanced by thermal relaxation, and the vertical advection is weak.

This model gives useful insight into the meridional circulation in the middle atmosphere. (One might argue that the relevant value of κ/α is zero, or certainly small.) The seasonally varying thermal forcing alone will give rise only to a weak meridional circulation. In order to account for a meridional circulation of the observed magnitude (and equivalently, for the observed deviation from equilibrium between thermal forcing and thermal dissipation) it is necessary to take account of the mechanical forcing due to dissipating and breaking waves (large-scale Rossby waves and gravity waves).

This picture breaks down at low latitudes, since f_0 vanishes at the equator, and so, according to the above model, however small ω or κ relative to α, there will always be a region in which the meridional circulation plays a substantial role in (8.23).

8.7 A steady-state thermally driven circulation controlled by friction

We now investigate what happens at low latitudes in more detail, first assuming that mechanical friction is dominant and then, in §8.8, considering the possible effects of non-linear advection of momentum, using a variation on a model first studied by Held and Hou. The problem is motivated by the Hadley circulation in the troposphere, i.e. the rising motion in the tropics and sinking in the extratropics that is seen in the (Eulerian) longitudinal-mean velocity field.

We consider a case in which the temperature field is relaxed towards a state which has a maximum temperature at the equator. To make the mathematics as simple as possible we focus on low-latitude regions where the Coriolis parameter may be approximated as a linear function of latitude, i.e. $f_0 = \beta y$. (This is the so-called equatorial β-plane approximation.) The derivation of (8.25) from (8.20, 8.21, 8.22 8.23, 8.24) follows in the same way as before. We consider the steady-state problem, so that ω may be set equal to zero. Furthermore, we assume that $\hat{\mathcal{H}} = -\frac{1}{2}\alpha\rho_{b2}y^2$, where ρ_{b2} is a constant. With the temperature damping term, this gives the effect of relaxing the temperature distribution towards a profile with an equatorial maximum (or minimum).

Then (8.25) may be written in the form

$$\beta^2 y^2 \hat{\chi}_{zz} + N^2(\frac{\kappa}{\alpha})\hat{\chi}_{yy} = -\frac{g\rho_{b2}y}{\rho_0}\kappa. \tag{8.29}$$

We apply the boundary conditions $\hat{\chi} = 0$ on $z = 0$ and $z = D$ and seek solutions such that $\hat{\chi} \to 0$ as $|y| \to \infty$.

We seek a solution by expanding χ as a Fourier sine series in z, i.e.,

$$\hat{\chi} = \sum_{n=1}^{\infty} \hat{\chi}_n(y)\sin(\frac{n\pi z}{D}). \tag{8.30}$$

It follows that

$$-\beta^2 y^2 \frac{n^2\pi^2}{D^2}\hat{\chi}_n + N^2(\frac{\kappa}{\alpha})\hat{\chi}_{nyy} = -\frac{4}{n\pi}\frac{g\rho_{b2}y}{\rho_0}\kappa. \tag{8.31}$$

for n odd. For n even the forcing term on the right-hand side is zero and hence, recalling the boundary condition in y, it follows that $\chi_n(y) = 0$.

Concentrating on the case where κ is small (e.g. compared to α) an approximate solution to (8.29) might be based on neglecting the second term on the left-hand side, i.e.

$$\hat{\chi}_n \simeq \frac{4D^2}{n^3\pi^3}\frac{g\rho_b 2}{\rho_0}\frac{\kappa}{\beta^2 y}. \tag{8.32}$$

However, it is apparent that this approximation breaks down for sufficiently small y and indeed the two terms on the left-hand side become of comparable importance when

$$y \sim \left(\frac{ND}{n\pi\beta}\right)^{1/2}\left(\frac{\kappa}{\alpha}\right)^{1/4} = \left(\frac{1}{n\pi}\right)^{1/2}\delta_{\text{friction}}. \tag{8.33}$$

The first term in the product is the *equatorial Rossby radius* associated with height scale $D/(n\pi)$. When multiplied by the second factor, we have the scale of an equatorial boundary layer in which vertical advection plays a leading role in (8.23). The mechanical friction plays an important role in balancing the Coriolis torque associated with the corresponding meridional circulation and the thickness of this boundary layer therefore depends strongly on magnitude of the frictional relaxation.

This motivates a rescaling of the equations by defining a variable η, such that

$$y = (ND/\pi\beta)^{1/2}(\kappa/\alpha)^{1/4}\eta = \delta_{\text{friction}}\eta \tag{8.34}$$

and hence that

$$\hat{\chi}_{n\eta\eta} - n^2\eta^2\chi_n = -\frac{4}{n\pi}\frac{\kappa^{3/4}\alpha^{1/4}}{N^2}(ND/\pi\beta)^{3/2}\frac{g\rho_{b2}}{\rho_0}\eta. \tag{8.35}$$

This equation can be solved in terms of special functions if required, but the important result is that the size of each $\hat{\chi}_n$, and hence the size of the maximum value taken by $\hat{\chi}$ itself, is

$$\hat{\chi}_{\text{friction}} \sim \frac{\kappa^{3/4}\alpha^{1/4}}{N^2}(ND/\pi\beta)^{3/2}\frac{g\rho_{b2}}{\rho_0}. \tag{8.36}$$

We thus again have that the strength of the circulation cell (measured by the maximum or minimum value of the streamfunction) controlled by the strength of the mechanical friction, but the strength of the circulation (measured by the mass flux) is proportional to $\kappa^{3/4}$ and is therefore, for small κ, larger than the linear scaling with κ that would be suggested by (8.25) or (8.35). This 'stronger' circulation applies in a region of width proportional to $\kappa^{1/4}$, as predicted by (8.34)

One unsatisfactory aspect when applying this to the real atmosphere is the value of κ, or indeed whether a linear friction model is appropriate at all, is highly uncertain. It is therefore interesting to consider other possibilities for the low-latitude balance.

8.8 Steady-state thermally driven circulations - the frictionless limit

As κ becomes small the region of appreciable circulation becomes thinner. Note that the balance in the momentum equation is $-\beta y\hat{\chi}_z = \kappa u$, so that $u_{yy} = -\beta(y\hat{\chi}_{zyy} + 2\hat{\chi}_{zy})/\kappa$. Using (8.32) and (8.33), it follows that

$$u_{yy} \sim \beta\left(\frac{g\rho_{b2}}{\rho_0 N\beta}\right)\left(\frac{\alpha}{\kappa}\right)^{1/2}. \tag{8.37}$$

In writing down the form of the mean momentum equations (6.7) or (6.16) it has been assumed that the relative vorticity is much less than f, or equivalently near the equator, $u_{yy}/\beta \ll 1$. It is clear from the above estimate that this assumption breaks down if κ is small enough, specifically in this case if

$$\left(\frac{\kappa}{\alpha}\right)^{1/2}\left(\frac{\rho_0 N\beta}{g\rho_{b2}}\right) \ll 1. \tag{8.38}$$

It is then necessary to retain in the x-momentum equation the nonlinear terms associated with advection of (relative) momentum by the mean meridional circulation.

Similar conclusions are reached, though the scalings are different in detail, if the mechanical dissipation is provided not by linear friction, but by viscous diffusion. (It is vertical diffusion which is usually taken to be most important, since vertical length scales are so much smaller than horizontal length scales.) Since it is the viscous model that has been most studied historically, we shall concentrate on that from now on. To convert between the two cases of linear friction and viscous diffusion we might take $\kappa = \nu D^2$, where ν is the diffusivity and D is the height of the domain.

The linear solutions for the model introduced in §8.7 predict vanishing meridional circulation in the limit of weak friction. Is this also a possibility in a nonlinear solution? The entire flow would be in thermal equilibrium, with $\mathcal{H}$ balancing $\alpha\rho$, i.e. $\rho = \rho_b$. The velocity in the x-direction will then be determined, up to a function of latitude, through thermal wind balance. With diffusion, however weak, it is appropriate to apply a no-slip condition at the ground ($z = 0$), hence determining the velocity completely. At the upper boundary ($z = D$) some kind of zero-stress condition is more appropriate.

In practice, when the system is modelled numerically, this thermal equilibrium state is not observed. There are a number of arguments as to why not. For example, if the equilibrium temperature distribution has a maximum near the equator, then u_z must be positive there, and hence the total angular momentum (u plus the relevant contribution from the background rotation) must attain a maximum away from the lower boundary. This can be shown to be inconsistent with the assumption of downgradient diffusion of momentum. The thermal equilibrium state is thus ruled out.

It is therefore relevant to examine possibilities of flow configurations in which there is non-zero circulation. The form of the x-momentum equation (actually an angular momentum equation) is, including the nonlinear terms and in the limit of vanishing friction,

$$J(\chi, u - \tfrac{1}{2}\beta y^2) = 0. \tag{8.39}$$

The nonlinear form of the density equation is

$$J(\chi, \rho) = -\alpha(\rho - \rho_b). \tag{8.40}$$

The hydrostatic equation and the expression of geostrophic balance in the y-momentum equation remain as before. From (8.39) it follows that a meridional circulation is possible only if $u - \frac{1}{2}\beta y^2$ is constant. This motivates the following heuristic solution (a simplified form of that due to Held and Hou).

It is assumed that fluid rises at the equator $y = 0$ and moves poleward along the (stress-free) upper boundary, preserving its angular momentum, up to some latitude y_H say. Poleward of this latitude there is no circulation and the fluid is in thermal equilibrium. Thus if the value of the angular momentum is m_0, we have that

$$u(y, D) = \tfrac{1}{2}\beta y^2 + m_0.$$

Near the lower boundary viscous diffusion remains important, so that the no-slip condition still applies and $u(y, 0) = 0$.

The thermal wind equation therefore implies that

$$\beta y u(y, D) = \beta y(\tfrac{1}{2}\beta y^2 + m_0) = -\frac{g}{\rho_0}\int_{z=0}^{D} \rho_y(y, z')dz'. \tag{8.41}$$

Integrating the density equation over the region in which there is a circulation (and using the fact that the boundary is a streamline) it follows that

$$\int_{y=0}^{y_H}\int_{z=0}^{D} (\rho - \rho_b)dydz = 0. \tag{8.42}$$

It is convenient to define a function $R(y)$ by

$$R(y) = \int_{z=0}^{D} \frac{g\rho(y, z)}{\rho_0}dz. \tag{8.43}$$

A similar function $R_b(y)$ may be defined replacing ρ by ρ_b. Then (8.41) may be integrated in y to give

$$R(y) = R(0) - \tfrac{1}{2}m_0\beta y^2 - \tfrac{1}{8}\beta^2 y^4 \qquad (8.44)$$

and (8.42) may be written

$$\int_{y=0}^{y_H} R(y)dy = \int_{y=0}^{y_H} R_b(y)dy. \qquad (8.45)$$

In addition it is required that the density be continuous at $y = y_H$, so that

$$R(y_H) = R_b(y_H). \qquad (8.46)$$

Substitution of (8.44) into (8.45) and (8.46) gives two equations for three unknowns, $R(0)$, y_H and m_0. The last is conventionally specified by assuming that it is set by the position of the upwelling and, for the corresponding value of y is equal to $u - \frac{1}{2}\beta y^2$ at $z = 0$. Here the upwelling is at the equator, so $m_0 = 0$. This gives two equations for two unknowns, which may be solved to give

$$y_H = \left(\frac{5\rho_{b2}gD}{3\rho_0\beta^2}\right)^{1/2} = \left(\frac{5}{3}\right)^{1/2}\left(\frac{\rho_{b2}g}{\rho_0 N\beta}\right)^{1/2}\left(\frac{ND}{\beta}\right)^{1/2} = \left(\frac{5}{3}\right)^{1/2}\delta_{\text{nonlinear}}. \qquad (8.47)$$

This may be compared with the width of the frictional circulation (8.34)

It is also possible to estimate the maximum streamfunction (and hence the flux) in the meridional circulation. From the density equation (8.40) it follows that $wN^2/\alpha \sim g(\rho - \rho_b)/\rho_0 \sim R(0)/D$, and hence from (8.44) that $wN^2/\alpha \sim \beta^2 y_H^4/D$. Hence the maximum streamfunction is of order

$$\chi \sim wy_H \sim \frac{\alpha\beta^2 y_H^5}{N^2 D} \sim \alpha D\left(\frac{\rho_{b2}g}{\rho_0 N\beta}\right)^{5/2}\left(\frac{ND}{\beta}\right)^{1/2}, \qquad (8.48)$$

which should be compared with (8.36). Note that (8.36) may be written in the alternative form

$$\chi_{\text{friction}} \sim \kappa^{3/4}\alpha^{1/4}D\left(\frac{\rho_{b2}g}{\rho_0 N\beta}\right)\left(\frac{ND}{\beta}\right)^{1/2} = \left(\frac{\delta_{\text{friction}}}{\delta_{\text{nonlinear}}}\right)^3 \chi_{\text{nonlinear}}. \qquad (8.49)$$

We expect that the above nonlinear theory holds if $\delta_{\text{nonlinear}} \gg \delta_{\text{friction}}$ and the linear friction theory described in §8.7 holds if $\delta_{\text{nonlinear}} \ll \delta_{\text{friction}}$.

The above approach is certainly not mathematically rigorous, but it does seem to give useful insight into the behaviour that is actually seen in numerical solution when the friction is small. The Held and Hou model is attractive because it predicts a width and strength of the low-latitude circulation that is independent of friction. It seems a valuable simple model of the longitudinally averaged low-latitude circulation, not least because the predicted flow bears a strong qualitative resemblance to that seen at low latitudes in the real atmosphere and in realistic atmospheric models. Nonetheless (of course) there are important features that it does not capture, the most obvious being the strong longitudinal variation of the tropical atmosphere (particularly the so-called Walker circulation).