ATMOSPHERIC AND OCEANIC FLUID DYNAMICS
Fundamentals and Large-Scale Circulation

Geoffrey K. Vallis
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.11</td>
<td>* Acoustic-Gravity Waves in an Ideal Gas</td>
<td>100</td>
</tr>
<tr>
<td>2.12</td>
<td>The Ekman Layer</td>
<td>104</td>
</tr>
<tr>
<td>3</td>
<td>Shallow Water Systems and Isentropic Coordinates</td>
<td>123</td>
</tr>
<tr>
<td>3.1</td>
<td>Dynamics of a Single, Shallow Layer</td>
<td>123</td>
</tr>
<tr>
<td>3.2</td>
<td>Reduced Gravity Equations</td>
<td>129</td>
</tr>
<tr>
<td>3.3</td>
<td>Multi-Layer Shallow Water Equations</td>
<td>131</td>
</tr>
<tr>
<td>3.4</td>
<td>Geostrophic Balance and Thermal wind</td>
<td>134</td>
</tr>
<tr>
<td>3.5</td>
<td>Form Drag</td>
<td>135</td>
</tr>
<tr>
<td>3.6</td>
<td>Conservation Properties of Shallow Water Systems</td>
<td>136</td>
</tr>
<tr>
<td>3.7</td>
<td>Shallow Water Waves</td>
<td>140</td>
</tr>
<tr>
<td>3.8</td>
<td>Geostrophic Adjustment</td>
<td>144</td>
</tr>
<tr>
<td>3.9</td>
<td>Isentropic Coordinates</td>
<td>152</td>
</tr>
<tr>
<td>3.10</td>
<td>Available Potential Energy</td>
<td>155</td>
</tr>
<tr>
<td>4</td>
<td>Vorticity and Potential Vorticity</td>
<td>165</td>
</tr>
<tr>
<td>4.1</td>
<td>Vorticity and Circulation</td>
<td>165</td>
</tr>
<tr>
<td>4.2</td>
<td>The Vorticity Equation</td>
<td>167</td>
</tr>
<tr>
<td>4.3</td>
<td>Vorticity and Circulation Theorems</td>
<td>170</td>
</tr>
<tr>
<td>4.4</td>
<td>Vorticity Equation in a Rotating Frame</td>
<td>177</td>
</tr>
<tr>
<td>4.5</td>
<td>Potential Vorticity Conservation</td>
<td>180</td>
</tr>
<tr>
<td>4.6</td>
<td>* Potential Vorticity in the Shallow Water System</td>
<td>186</td>
</tr>
<tr>
<td>4.7</td>
<td>Potential Vorticity in Approximate, Stratified Models</td>
<td>188</td>
</tr>
<tr>
<td>4.8</td>
<td>The Impermeability of Isentropes to Potential Vorticity</td>
<td>191</td>
</tr>
<tr>
<td>5</td>
<td>Simplified Equations for Ocean and Atmosphere</td>
<td>199</td>
</tr>
<tr>
<td>5.1</td>
<td>Geostrophic Scaling</td>
<td>200</td>
</tr>
<tr>
<td>5.2</td>
<td>The Planetary Geostrophic Equations</td>
<td>204</td>
</tr>
<tr>
<td>5.3</td>
<td>The Shallow Water Quasi-Geostrophic Equations</td>
<td>209</td>
</tr>
<tr>
<td>5.4</td>
<td>The Continuously Stratified Quasi-Geostrophic System</td>
<td>217</td>
</tr>
<tr>
<td>5.5</td>
<td>* Quasi-geostrophy and Ertel Potential Vorticity</td>
<td>226</td>
</tr>
<tr>
<td>5.6</td>
<td>* Energetics of Quasi-Geostrophy</td>
<td>229</td>
</tr>
<tr>
<td>5.7</td>
<td>Rossby Waves</td>
<td>232</td>
</tr>
<tr>
<td>5.8</td>
<td>* Rossby Waves in Stratified Quasi-Geostrophic Flow</td>
<td>237</td>
</tr>
<tr>
<td>5.9</td>
<td>Appendix: Wave Kinematics, Group Velocity and Phase Speed</td>
<td>239</td>
</tr>
<tr>
<td>Part II</td>
<td>INSTABILITIES, WAVE–MEAN FLOW INTERACTION AND TURBULENCE</td>
<td>249</td>
</tr>
<tr>
<td>6</td>
<td>Barotropic and Baroclinic Instability</td>
<td>251</td>
</tr>
<tr>
<td>6.1</td>
<td>Kelvin-Helmholtz Instability</td>
<td>251</td>
</tr>
<tr>
<td>6.2</td>
<td>Instability of Parallel Shear Flow</td>
<td>253</td>
</tr>
<tr>
<td>6.3</td>
<td>Necessary Conditions for Instability</td>
<td>262</td>
</tr>
<tr>
<td>6.4</td>
<td>Baroclinic Instability</td>
<td>264</td>
</tr>
<tr>
<td>6.5</td>
<td>The Eady Problem</td>
<td>269</td>
</tr>
<tr>
<td>6.6</td>
<td>Two-Layer Baroclinic Instability</td>
<td>275</td>
</tr>
</tbody>
</table>
6.7 An Informal View of the Mechanism of Baroclinic Instability 281
6.8 * The Energetics of Linear Baroclinic Instability 286
6.9 * Beta, Shear and Stratification in a Continuous Model 288

7 Wave–Mean Flow Interaction 301
7.1 Quasi-geostrophic Preliminaries 302
7.2 The Eliassen-Palm Flux 304
7.3 * The Transformed Eulerian Mean 310
7.4 * The Nonacceleration Result 320
7.5 Influence of Eddies on the Mean Flow in the Eady Problem 325
7.6 * Necessary Conditions for Instability 330
7.7 * Necessary Conditions for Instability: Use of Pseudoenergy 334

8 Turbulence, Basic Theory 343
8.1 The Fundamental Problem of Turbulence 344
8.2 The Kolmogorov Theory 346
8.3 Two-Dimensional Turbulence 356
8.4 Predictability of Turbulence 367
8.5 * Spectra of Passive Tracers 372

9 Geostrophic Turbulence and Baroclinic Eddies 383
9.1 Effects of Differential Rotation 383
9.2 Stratified Geostrophic Turbulence 389
9.3 † A Scaling Theory for Geostrophic Turbulence 397
9.4 † Phenomenology of Baroclinic Eddies in the Atmosphere and Ocean 401

10 Turbulent Diffusion and Eddy Transport 411
10.1 Diffusive Transport 412
10.2 Turbulent Diffusion 413
10.3 Two-Particle Diffusivity 419
10.4 Mixing Length Theory 423
10.5 Homogenization of a Scalar that is Advected and Diffused 427
10.6 † Transport by Baroclinic Eddies 430
10.7 † Eddy Diffusion in the Atmosphere and Ocean 434
10.8 † Thickness Diffusion 444
10.9 † Eddy Transport and the Transformed Eulerian Mean 447

Part III LARGE-SCALE ATMOSPHERIC CIRCULATION 453

11 The Overturning Circulation: Hadley and Ferrel Cells 455
11.1 Basic Features of the Atmosphere 456
11.2 A Steady Model of the Hadley Cell 460
11.3 A Shallow Water Model of the Hadley Cell 471
11.4 † Asymmetry Around the Equator 472
11.5 Eddies, Viscosity and the Hadley Cell 476
11.6 The Hadley Cell: Summary and Numerical Solutions 480
11.7 The Ferrel Cell 483

12 Mid-Latitude Atmospheric Circulation 489
12.1 Surface Westerlies and the Maintenance of a Barotropic Jet 490
12.2 Layered Models of the Mid-latitude Circulation 501
12.3 † An Example of a Closed Model 517
12.4 * Eddy Fluxes and Necessary conditions for Instability 519
12.5 A Stratified Model and the Real Atmosphere 521
12.6 The Tropopause and the Stratification of the Atmosphere 528
12.7 † Baroclinic eddies and Potential Vorticity Mixing 533
12.8 † Extra-tropical Convection and the Ventilated Troposphere 538
Appendix: TEM & EP Flux for Primitive Eqs. in Spherical Coordinates 540

13 Zonally Asymmetries, Planetary Waves and Stratosphere 545
13.1 Forced and Stationary Rossby Waves 546
13.2 * Meridional Propagation and Dispersion 552
13.3 * Stratified Rossby Waves and their Vertical Propagation 556
13.4 * Effects of Thermal Forcing 562
13.5 Stratospheric Dynamics 568

Part IV LARGE-SCALE OCEANIC CIRCULATION 581

14 Wind-Driven Gyres 583
14.1 The Depth Integrated Wind-Driven Circulation 585
14.2 Using Viscosity Instead of Drag 593
14.3 Zonal Boundary Layers 597
14.4 The Nonlinear Problem 599
14.5 * Inertial Solutions 601
14.6 * Topographic Effects on Western Boundary Currents 608
14.7 * Vertical Structure of the Wind-driven Circulation 613
14.8 * A Model with Continuous Stratification 619

15 The Buoyancy Driven Circulation 629
15.1 A Brief Observational Overview 630
15.2 † Sideways Convection 631
15.3 The Maintenance of Sideways Convection 636
15.4 Simple Box Models 642
15.5 A Laboratory Model of the Abyssal Circulation 648
15.6 A Model for Oceanic Abyssal Flow 653
15.7 * A Shallow Water Model of the Abyssal Flow 658
15.8 Scaling for the Buoyancy-driven Circulation 661

16 Wind and Buoyancy Driven Circulation 667
16.1 The Main Thermocline: an Introduction 667
16.2 Scaling and Simple Dynamics of the Main Thermocline 670
16.3 The Internal Thermocline 674
An asterisk indicates more advanced, but usually uncontroversial, material that may be omitted on a first reading. A dagger indicates material that is still a topic of research or that is not settled. Caveat emptor.
NOTATION

The list below contains only the more important variables, or instances of non-obvious notation. Distinct meanings are separated with a semi-colon. Variables are normally set in italics, constants (e.g., π) in roman (i.e., upright), differential operators in roman, vectors in bold, and tensors in sans serif. Thus, vector variables are in bold italics, vector constants (e.g., unit vectors) in bold roman, and tensor variables are in slanting sans serif. Physical units are set in roman. A subscript denotes a derivative only if the subscript is a coordinate, such as x, y or z; a subscript 0 generally denotes a constant reference value (e.g., \( \rho_0 \)). The components of a vector are denoted by superscripts.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>( b )</td>
<td>Buoyancy, ( -g\delta \rho/\rho_0 ) or ( g\delta \theta/\tilde{\theta} ).</td>
</tr>
<tr>
<td>( c_\beta )</td>
<td>Group velocity, ((c^x_\beta, c^y_\beta, c^z_\beta)).</td>
</tr>
<tr>
<td>( c_p )</td>
<td>Phase speed; heat capacity at constant pressure.</td>
</tr>
<tr>
<td>( c_v )</td>
<td>Heat capacity constant volume.</td>
</tr>
<tr>
<td>( c_s )</td>
<td>Sound speed.</td>
</tr>
<tr>
<td>( f, f_0 )</td>
<td>Coriolis parameter, and its reference value.</td>
</tr>
<tr>
<td>( \mathbf{g}, g )</td>
<td>Vector acceleration due to gravity, magnitude of ( \mathbf{g} ).</td>
</tr>
<tr>
<td>( h )</td>
<td>Layer thickness (in shallow water equations).</td>
</tr>
<tr>
<td>( \mathbf{i}, \mathbf{j}, \mathbf{k} )</td>
<td>Unit vectors in ((x, y, z)) directions.</td>
</tr>
<tr>
<td>( i )</td>
<td>An integer index.</td>
</tr>
<tr>
<td>( i )</td>
<td>Square root of (-1).</td>
</tr>
<tr>
<td>( \mathbf{k} )</td>
<td>Wave vector, with components ((k, l, m)) or ((k^x, k^y, k^z)).</td>
</tr>
<tr>
<td>( k_d )</td>
<td>Wave number corresponding to deformation radius.</td>
</tr>
<tr>
<td>( L_d )</td>
<td>Deformation radius.</td>
</tr>
<tr>
<td>( L, H )</td>
<td>Horizontal length scale, vertical (height) scale.</td>
</tr>
<tr>
<td>( m )</td>
<td>Angular momentum about the earth's axis of rotation.</td>
</tr>
<tr>
<td>( M )</td>
<td>Montgomery function, ( M = c_pT + \Phi ).</td>
</tr>
<tr>
<td>( N )</td>
<td>Buoyancy, or Brunt-Väisälä, frequency.</td>
</tr>
<tr>
<td>( p )</td>
<td>Pressure.</td>
</tr>
<tr>
<td>( Pr )</td>
<td>Prandtl ratio, ( f_0/N ).</td>
</tr>
<tr>
<td>( q )</td>
<td>Quasi-geostrophic potential vorticity.</td>
</tr>
<tr>
<td>( Q )</td>
<td>Potential vorticity (in particular Ertel PV).</td>
</tr>
<tr>
<td>( \dot{Q} )</td>
<td>Rate of heating.</td>
</tr>
<tr>
<td>( Ra )</td>
<td>Rayleigh number.</td>
</tr>
<tr>
<td>( Re )</td>
<td>Real part of expression.</td>
</tr>
<tr>
<td>( Re )</td>
<td>Reynolds number, ( UL/\nu ).</td>
</tr>
<tr>
<td>( Ro )</td>
<td>Rossby number, ( U/\nu L ).</td>
</tr>
<tr>
<td>( S )</td>
<td>Salinity; source term on right-hand side of evolution equation.</td>
</tr>
<tr>
<td>( T )</td>
<td>Temperature.</td>
</tr>
<tr>
<td>( t )</td>
<td>Time.</td>
</tr>
<tr>
<td>( \mathbf{u} )</td>
<td>Two-dimensional, horizontal velocity, ((u, v)).</td>
</tr>
<tr>
<td>( \mathbf{v} )</td>
<td>Three-dimensional velocity, ((u, v, z)).</td>
</tr>
<tr>
<td>( x, y, z )</td>
<td>Cartesian coordinates, usually in zonal, meridional and vertical directions.</td>
</tr>
<tr>
<td>( Z )</td>
<td>Log-pressure, (-H \log p/\rho_R). We often use ( H = 7.5 \text{ km} ) and ( \rho_R = 10^5 \text{ Pa} ).</td>
</tr>
<tr>
<td>Variable</td>
<td>Description</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
</tr>
<tr>
<td>$A$</td>
<td>Wave activity.</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>Inverse density, or specific volume; aspect ratio.</td>
</tr>
<tr>
<td>$\beta$</td>
<td>Rate of change of $f$ with latitude, $\partial f / \partial y$.</td>
</tr>
<tr>
<td>$\beta_T$, $\beta_S$</td>
<td>Coefficient of expansion with respect to temperature, salinity.</td>
</tr>
<tr>
<td>$\epsilon$</td>
<td>Generic small parameter (epsilon).</td>
</tr>
<tr>
<td>$\varepsilon$</td>
<td>Cascade or dissipation rate of energy (varepsilon).</td>
</tr>
<tr>
<td>$\eta$</td>
<td>Specific entropy; perturbation height; enstrophy cascade or dissipation rate.</td>
</tr>
<tr>
<td>$F$</td>
<td>Eliassen Palm flux, $(F^y, F^z)$.</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>Vorticity gradient, $\beta - u_{yy}$; the ratio $c_p/c_v$.</td>
</tr>
<tr>
<td>$\Gamma$</td>
<td>Lapse rate.</td>
</tr>
<tr>
<td>$\kappa$</td>
<td>Diffusivity; the ratio $R/c_p$.</td>
</tr>
<tr>
<td>$K$</td>
<td>Kolmogorov or Kolomogorov-like constant.</td>
</tr>
<tr>
<td>$\Lambda$</td>
<td>Shear, e.g., $\partial U / \partial z$.</td>
</tr>
<tr>
<td>$\nu$</td>
<td>Kinematic viscosity.</td>
</tr>
<tr>
<td>$\psi$</td>
<td>Meridional component of velocity.</td>
</tr>
<tr>
<td>$p$</td>
<td>Pressure divided by density, $p/\rho$; passive tracer.</td>
</tr>
<tr>
<td>$\Phi$</td>
<td>Geopotential, usually $gz$.</td>
</tr>
<tr>
<td>$\Pi$</td>
<td>Exner function, $\Pi = c_p T/\theta = c_p (p/p_R) R/c_p$.</td>
</tr>
<tr>
<td>$\omega$</td>
<td>Vorticity.</td>
</tr>
<tr>
<td>$\Omega$, $\Omega^-$</td>
<td>Rotation rate of earth and associated vector.</td>
</tr>
<tr>
<td>$\psi$</td>
<td>Streamfunction.</td>
</tr>
<tr>
<td>$\rho$</td>
<td>Density.</td>
</tr>
<tr>
<td>$\rho_\theta$</td>
<td>Potential density.</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>Layer thickness, $\partial z / \partial \theta$; Prandtl number $\nu/\kappa$; measure of density, $\rho - 1000$.</td>
</tr>
<tr>
<td>$\tau$</td>
<td>Wind stress.</td>
</tr>
<tr>
<td>$\tau_z$</td>
<td>Zonal component or magnitude of wind stress; eddy turnover time.</td>
</tr>
<tr>
<td>$\theta$</td>
<td>Potential temperature.</td>
</tr>
<tr>
<td>$\theta, \lambda$</td>
<td>Latitude, longitude.</td>
</tr>
<tr>
<td>$\zeta$</td>
<td>Vertical component of vorticity.</td>
</tr>
<tr>
<td>$\left( \frac{\partial a}{\partial b} \right)_c$</td>
<td>Derivative of $a$ with respect to $b$ at constant $c$.</td>
</tr>
<tr>
<td>$\frac{\partial a}{\partial b} \bigg</td>
<td>_{a=c}$</td>
</tr>
<tr>
<td>$\nabla_a$</td>
<td>Gradient operator at constant value of coordinate $a$, e.g., $\nabla_z = i \partial_x + j \partial_y$.</td>
</tr>
<tr>
<td>$\nabla_a^\perp$</td>
<td>Perpendicular gradient, $\nabla_a^\perp \phi = k \times \nabla_a \phi$.</td>
</tr>
<tr>
<td>$\text{curl}_z$</td>
<td>Vertical component of $\nabla \times$ operator, $\text{curl}_z A = k \cdot \nabla \times A = \partial_x A_y - \partial_y A_x$.</td>
</tr>
<tr>
<td>$D \over Dt$</td>
<td>Material derivative (generic).</td>
</tr>
<tr>
<td>$D_3 \over Dt$, $D_2 \over Dt$, $D_1 \over Dt$</td>
<td>Material derivative in three dimensions and in two dimensions, for example $\partial / \partial t + \mathbf{v} \cdot \nabla$ and $\partial / \partial t + \mathbf{u} \cdot \nabla$ respectively.</td>
</tr>
<tr>
<td>$D_\theta \over Dt$</td>
<td>Material derivative using geostrophic velocity, for example $\partial / \partial t + \mathbf{u}_g \cdot \nabla$.</td>
</tr>
</tbody>
</table>
Part I

FUNDAMENTALS OF GEOPHYSICAL FLUID DYNAMICS
CHAPTER ONE

Equations of Motion

This chapter establishes the fundamental governing equations of motion for a fluid, with particular reference to the fluids of the earth's atmosphere and ocean. Our approach in many places is quite informal, and the interested reader may consult the references given for more detail.

1.1 TIME DERIVATIVES FOR FLUIDS

The equations of motion of fluid mechanics differ from those of rigid-body mechanics because fluids form a continuum, and because fluids flow and deform. Thus, even though both classical solid and fluid media are governed by the same relatively simple physical laws (Newton's laws and the laws of thermodynamics), the expression of these laws differs between the two. To determine the equations of motion for fluids we must clearly establish what the time derivative of some property of a fluid actually means, and that is the subject of this section.

1.1.1 Field and material viewpoints

In solid-body mechanics one is normally concerned with the position and momentum of identifiable objects — the angular velocity of a spinning top or the motions of the planets around the sun are two well-worn examples. The position and velocity of a particular object is then computed as a function of time by formulating equations of the form

\[ \frac{dx_i}{dt} = F(x_i, t) \]  

(1.1)

where \( x_i \) is the set of positions and velocities of all the interacting objects and the operator \( F \) on the right-hand side is formulated using Newton's laws of motion. For
example, two massive point objects interacting via their gravitational field obey

\[ \frac{dr_i}{dt} = v_i, \quad \frac{dv_i}{dt} = -\frac{Gm_j}{(r_i - r_j)^2} \hat{r}_{i,j}, \quad i = 1, 2; \quad j = 3 - i. \]  

(1.2)

We thereby predict the positions, \( r_i \) and velocities, \( v_i \) of the objects given their masses, \( m_i \) and the gravitational constant \( G \), and where \( \hat{r}_{i,j} \) is a unit vector directed from \( r_i \) to \( r_j \).

In fluid dynamics such a procedure would lead to an analysis of fluid motions in terms of the positions and momenta of particular fluid elements, each identified by some label, which might simply be their position at an initial time. We call this a material point of view, because we are concerned with identifiable pieces of material; it is also sometimes called a Lagrangian view, after J.-L. Lagrange. The procedure is perfectly acceptable in principle, and if followed would provide a complete description of the fluid dynamical system. However, from a practical point of view it is much more than we need, and it would be extremely complicated to implement. Instead, for most problems we would like to know what the values of velocity, density and so on are at fixed points in space as time passes. (A weather forecast we might care about tells us how warm it will be where we live, and if we are given that we don’t particularly care where a fluid parcel comes from.) Since the fluid is a continuum, this knowledge is equivalent to knowing how the fields of the dynamical variables evolve in space and time, and this is often known as the field or Eulerian viewpoint, after L. Euler. Thus, whereas in the material view we consider the time evolution of identifiable fluid elements, in the field view we consider the time evolution of the fluid field from a particular frame of reference. That is, we seek evolution equations of the form

\[ \frac{\partial}{\partial t} \varphi(x, y, z, t) = F, \]  

(1.3)

where the field \( \varphi(x, y, z, t) \) is a dynamical variable (e.g., velocity, density, temperature) which gives the value at any point in space-time, and \( F \) is some operator to be determined from Newton’s laws of motion and appropriate thermodynamic laws.

Although the field viewpoint will turn out to be the most practically useful, the material description is invaluable both in deriving the equations and in the subsequent insight it frequently provides. This is because the important quantities from a fundamental point of view are often those which are associated with a given fluid element: it is these which directly enter Newton’s laws of motion and the thermodynamic equations. It is thus important to have a relationship between the rate of change of quantities associated with a given fluid element and the local rate of change of a field. The material or advective derivative provides this relationship.

### 1.1.2 The material derivative of a fluid property

A fluid element is an infinitesimal, indivisible, piece of fluid — effectively a very small fluid parcel. The material derivative is the rate of change of a property (such as temperature, or momentum) of a particular fluid element or finite mass; that is to say, it is the total time derivative of a property of a piece of fluid. It is also known as the ‘substantive derivative’ (the derivative associated with a parcel of
For fluids, the ‘advective derivative’ (because the fluid property is being advected), the ‘convective derivative’ (convection is a slightly old-fashioned name for advection, still used in some fields), or the ‘Lagrangian derivative’.

Let us suppose that a fluid is characterized by a (given) velocity field \( \mathbf{v}(x,t) \), which determines its velocity throughout. Let us also suppose that it has another property \( \phi \), and let us seek an expression for the rate of change of \( \phi \) of a fluid element. Since \( \phi \) is changing in time and in space we use the chain rule:

\[
\delta \phi = \frac{\partial \phi}{\partial t} \delta t + \frac{\partial \phi}{\partial x} \delta x + \frac{\partial \phi}{\partial y} \delta y + \frac{\partial \phi}{\partial z} \delta z = \frac{\partial \phi}{\partial t} \delta t + \delta \mathbf{x} \cdot \nabla \phi. \quad (1.4)
\]

This is true in general for any \( \delta t, \delta x, \) etc. Thus the total time derivative is

\[
\frac{d\phi}{dt} = \frac{\partial \phi}{\partial t} + \mathbf{v} \cdot \nabla \phi. \quad (1.5)
\]

If this is to be a material derivative we must identify the time derivative in the second term on the right-hand side with the rate of change of position of a fluid element, namely its velocity. Hence, the material derivative of the property \( \phi \) is

\[
\frac{d\phi}{dt} = \frac{\partial \phi}{\partial t} + \mathbf{v} \cdot \nabla \phi. \quad (1.6)
\]

The right-hand side expresses the material derivative in terms of the local rate of change of \( \phi \) plus a contribution arising from the spatial variation of \( \phi \), experienced only as the fluid parcel moves. Because the material derivative is so common, and to distinguish it from other derivatives, we denote it by the operator \( \frac{D}{Dt} \). Thus, the material derivative of the field \( \phi \) is

\[
\frac{D\phi}{Dt} = \frac{\partial \phi}{\partial t} + (\mathbf{v} \cdot \nabla) \phi. \quad (1.7)
\]

The brackets in the last term of this equation are helpful in reminding us that \( (\mathbf{v} \cdot \nabla) \) is an operator acting on \( \phi \).

**Material derivative of vector field**

The material derivative may act on a vector field \( \mathbf{b} \), in which case

\[
\frac{D\mathbf{b}}{Dt} = \frac{\partial \mathbf{b}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{b}. \quad (1.8)
\]

In Cartesian coordinates this is

\[
\frac{D\mathbf{b}}{Dt} = \frac{\partial \mathbf{b}}{\partial t} + \mathbf{u} \frac{\partial \mathbf{b}}{\partial x} + \mathbf{v} \frac{\partial \mathbf{b}}{\partial y} + \mathbf{w} \frac{\partial \mathbf{b}}{\partial z},
\]

and for a particular component of \( \mathbf{b} \),

\[
\frac{Db_x}{Dt} = \frac{\partial b_x}{\partial t} + \mathbf{u} \frac{\partial b_x}{\partial x} + \mathbf{v} \frac{\partial b_x}{\partial y} + \mathbf{w} \frac{\partial b_x}{\partial z}, \quad (1.10)
\]

or, in Cartesian tensor notation,

\[
\frac{Db_i}{Dt} = \frac{\partial b_i}{\partial t} + \mathbf{v}_j \frac{\partial b_i}{\partial x_j} = \frac{\partial b_i}{\partial t} + \mathbf{v}_j \partial_j b_i. \quad (1.11)
\]
where the subscripts denote the Cartesian components and repeated indices are summed. In coordinate systems other than Cartesian the advective derivative of a vector is not simply the sum of the advective derivative of its components, because the coordinate vectors change direction with position; this will be important when we deal with spherical coordinates (and see problem 2.5). Finally, we note that the advective derivative of the position of a fluid element, \( \mathbf{r} \) say, is its velocity, and this may easily checked by explicitly evaluating \( \frac{D\mathbf{r}}{Dt} \).

### 1.1.3 Material derivative of a volume

The volume that a given, unchanging, mass of fluid occupies is deformed and advected by the fluid motion, and there is no particular reason why it should remain constant. Indeed, the volume will change as a result of the movement of each element of its bounding material surface, and will in general change if there is a non-zero normal component of the velocity at the fluid surface. That is, if the volume of some fluid is \( \int dV \), then

\[
\frac{D}{Dt} \int dV = \int \nabla \cdot \mathbf{v} \, dV. \tag{1.12}
\]

where the subscript \( V \) indicates that the integral is a definite integral over some finite volume \( V \), although the limits of the integral will be functions of time if the volume is changing. The integral on the right-hand side is over the closed surface, \( S \), bounding the volume. Although intuitively apparent (to some), this expression may be derived more formally using Leibnitz’s formula for the rate of change of an integral whose limits are changing (problem 1.2). Using the divergence theorem on the right-hand side, (1.12) becomes

\[
\frac{D}{Dt} \int dV = \int \nabla \cdot \mathbf{v} \, dV. \tag{1.13}
\]

The rate of change of the volume of an infinitesimal fluid element of volume \( \Delta V \) is obtained by taking the limit of this expression as the volume tends to zero, giving

\[
\lim_{\Delta V \to 0} \frac{1}{\Delta V} \frac{D\Delta V}{Dt} = \nabla \cdot \mathbf{v}. \tag{1.14}
\]

We will often write such expressions informally as

\[
\frac{D\Delta V}{Dt} = \Delta V \nabla \cdot \mathbf{v}, \tag{1.15}
\]

with the limit implied.

Consider now the material derivative of a fluid property, \( \xi \), multiplied by the volume of a fluid element, \( \Delta V \). This situation arises when \( \xi \) is the amount per unit volume of \( \xi \)-substance — it might, for example, be mass density or the amount of a dye per unit volume. Then we have

\[
\frac{D}{Dt} (\xi \Delta V) = \xi \frac{D\Delta V}{Dt} + \Delta V \frac{D\xi}{Dt}. \tag{1.16}
\]

Using (1.15) this becomes

\[
\frac{D}{Dt} (\xi \Delta V) = \Delta V \left( \nabla \cdot \mathbf{v} + \frac{D\xi}{Dt} \right), \tag{1.17}
\]
and the analogous result for a finite fluid volume is just

$$\frac{D}{Dt} \int_V \xi \, dV = \int_V \left( \xi \nabla \cdot \mathbf{v} + \frac{D\xi}{Dt} \right) \, dV.$$  \hfill (1.18)

This expression is to be contrasted with the Eulerian derivative for which the volume, and so the limits of integration, are fixed and we have

$$\frac{d}{dt} \int_V \xi \, dV = \int_V \frac{\partial \xi}{\partial t} \, dV.$$  \hfill (1.19)

Now consider the material derivative of a fluid property $\phi$ multiplied by the mass of a fluid element, $\rho \Delta V$. This arises when $\phi$ is the amount of $\phi$-substance per unit mass (note, for example, that the momentum of a fluid element is $\rho v \Delta V$). The material derivative of $\phi \rho \Delta V$ is given by

$$\frac{D}{Dt}(\phi \rho \Delta V) = \rho \Delta V \frac{D\phi}{Dt} + \phi \frac{D}{Dt}(\rho \Delta V)$$  \hfill (1.20)

But $\rho \Delta V$ is just the mass of the fluid element, and that is constant — it is how a fluid element is defined. Thus the second term on the right-hand side vanishes and

$$\frac{D}{Dt}(\phi \rho \Delta V) = \rho \Delta V \frac{D\phi}{Dt} \quad \text{and} \quad \frac{D}{Dt} \int_V \phi \rho \, dV = \int_V \rho \frac{D\phi}{Dt} \, dV,$$  \hfill (1.21a,b)

where (1.21b) applies to a finite volume. That expression may also be derived more formally using Leibnitz’s formula for the material derivative of an integral, and the result also holds when $\phi$ is a vector. The result is quite different from the corresponding Eulerian derivative, in which the volume is kept fixed; in that case we have:

$$\frac{d}{dt} \int_V \phi \rho \, dV = \int_V \frac{\partial}{\partial t} (\phi \rho) \, dV.$$  \hfill (1.22)

Various material and Eulerian derivatives are summarized in the shaded box on the following page.

### 1.2 THE MASS CONTINUITY EQUATION

In classical mechanics mass is absolutely conserved. However, in fluid mechanics fluid flows into and away from regions, and fluid density may change, and an equation that explicitly accounts for the flow of mass is one of the ‘equations of motion’ of the fluid.

#### 1.2.1 An Eulerian derivation

We will first derive the mass conservation equation from an Eulerian point of view; that is to say, our reference frame is fixed in space and the fluid flows through it.
Material and Eulerian Derivatives

The material derivative of a scalar ($\phi$) and a vector ($b$) field are given by:

$$ \frac{D\phi}{Dt} = \frac{\partial\phi}{\partial t} + \mathbf{v} \cdot \nabla \phi, \quad \frac{Db}{Dt} = \frac{\partial b}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{b}. $$ \hspace{1cm} (D.1)

Various material derivatives of integrals are:

$$ \frac{D}{Dt} \int_V \phi \, dV = \int_V \left( \frac{D\phi}{Dt} + \phi \nabla \cdot \mathbf{v} \right) \, dV = \int_V \left( \frac{\partial\phi}{\partial t} + \nabla \cdot (\phi \mathbf{v}) \right) \, dV, \quad \hspace{1cm} (D.2) $$

$$ \frac{D}{Dt} \int_V \, dV = \int_V \nabla \cdot \mathbf{v} \, dV, \quad \hspace{1cm} (D.3) $$

$$ \frac{D}{Dt} \int_V \rho \phi \, dV = \int_V \rho \frac{D\phi}{Dt} \, dV. \quad \hspace{1cm} (D.4) $$

These formulae also hold if $\phi$ is a vector. The Eulerian derivative of an integral is:

$$ \frac{d}{dt} \int_V \phi \, dV = \int_V \frac{\partial\phi}{\partial t} \, dV, \quad \hspace{1cm} (D.5) $$

so that

$$ \frac{d}{dt} \int_V \, dV = 0 \quad \text{and} \quad \frac{d}{dt} \int_V \rho \phi \, dV = \int_V \frac{\partial\rho\phi}{\partial t} \, dV. \quad \hspace{1cm} (D.6) $$

Fig. 1.1 Mass conservation in an Eulerian cuboid control volume.
**Cartesian derivation**

Consider an infinitesimal rectangular parallelepiped control volume $\Delta V = \Delta x \Delta y \Delta z$ that is fixed in space (Fig. 1.1). Fluid moves into or out of the volume through its surface, including through its faces in the $y$–$z$ plane of area $\Delta A = \Delta y \Delta z$ at coordinates $x$ and $x + \Delta x$. The accumulation of fluid within the control volume due to motion in the $x$-direction is evidently

$$\Delta y \Delta z \left[ (\rho u)(x) - (\rho u)(x + \Delta x) \right] = -\frac{\partial (\rho u)}{\partial x} \Delta x \Delta y \Delta z. \quad (1.23)$$

To this must be added the effects of motion in the $y$- and $z$-directions, namely

$$- \left[ \frac{\partial (\rho v)}{\partial y} + \frac{\partial (\rho w)}{\partial z} \right] \Delta x \Delta y \Delta z. \quad (1.24)$$

This net accumulation of fluid must be accompanied by a corresponding increase of fluid mass within the control volume. This is

$$\frac{\partial}{\partial t} (\text{Density} \times \text{Volume}) = \Delta x \Delta y \Delta z \frac{\partial \rho}{\partial t}, \quad (1.25)$$

because the volume is constant. Thus, because mass is conserved, (1.23), (1.24) and (1.25) give

$$\Delta x \Delta y \Delta z \left[ \frac{\partial \rho}{\partial t} + \frac{\partial (\rho u)}{\partial x} + \frac{\partial (\rho v)}{\partial y} + \frac{\partial (\rho w)}{\partial z} \right] = 0. \quad (1.26)$$

Because the control volume is arbitrary the quantity in square brackets must be zero and we have the **mass continuity equation**:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \quad (1.27)$$

**Vector derivation**

Consider an arbitrary control volume $V$ bounded by a surface $S$, fixed in space, with by convention the direction of $S$ being toward the outside of $V$, as in Fig. 1.2. The rate of fluid loss due to flow through the closed surface $S$ is then given by

$$\text{Fluid loss} = \int_S \rho \mathbf{v} \cdot d\mathbf{S} = \int_V \nabla \cdot (\rho \mathbf{v}) \, dV \quad (1.28)$$

using the divergence theorem. This must be balanced by a change in the mass $M$ of the fluid within the control volume, which, since its volume is fixed, implies a density change. That is

$$\text{Fluid loss} = -\frac{dM}{dt} = -\frac{d}{dt} \int_V \rho \, dV = -\int_V \frac{d\rho}{dt} \, dV. \quad (1.29)$$

Equating (1.28) and (1.29) yields

$$\int_V \left[ \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) \right] \, dV = 0 \quad (1.30)$$

Because the volume is arbitrary, the integrand must vanish and we recover (1.27).
1.2.2 Mass continuity via the material derivative

We now derive the mass continuity equation (1.27) from a material perspective. This is the most fundamental approach of all since the principle of mass conservation states simply that the mass of a given element of fluid is, by definition of the element, constant. Thus, consider a small mass of fluid of density $\rho$ and volume $\Delta V$. Then conservation of mass may be represented by

$$\frac{D}{Dt}(\rho \Delta V) = 0 \quad (1.31)$$

Both the density and the volume of the parcel may change, so

$$\Delta V \frac{D\rho}{Dt} + \rho \frac{D\Delta V}{Dt} = \Delta V \left( \frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} \right) = 0 \quad (1.32)$$

where the second expression follows using (1.15). Since the volume element is arbitrary, the term in brackets must vanish and

$$\frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0. \quad (1.33)$$

After expansion of the first term this becomes identical to (1.27). This result may be derived more formally by re-writing (1.31) as the integral expression

$$\frac{D}{Dt} \int_V \rho \, dV = 0. \quad (1.34)$$

Expanding the derivative using (1.18) gives

$$\frac{D}{Dt} \int_V \rho \, dV = \int_V \left( \frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} \right) \, dV = 0. \quad (1.35)$$

Because the volume over which the integral is taken is arbitrary the integrand itself must vanish and we recover (1.33). Summarizing, equivalent partial differential
1.3 The Momentum Equation

The momentum equation is a partial differential equation that describes how the velocity or momentum of a fluid responds to internal and imposed forces. We will derive it using material methods and informally deducing the terms representing the pressure, gravitational and viscous forces.

1.2.3 A general continuity equation

The derivation of continuity equation for a general scalar property of a fluid is similar to that for density, except that there may be an external source or sink, and potentially a means of transferring the property from one location to another than by fluid motion, for example by diffusion. If \( \xi \) is the amount of some property of the fluid per unit volume (which we will call the concentration of the property), and if the net effect per unit volume of all nonconservative processes is denoted by \( Q_v[\xi] \), then the continuity equation for concentration may be written:

\[
\frac{D}{Dt}(\xi \Delta V) = Q_v[\xi] \Delta V
\]

Expanding the left hand side and using (1.15) we obtain

\[
\frac{D\xi}{Dt} + \xi \nabla \cdot \mathbf{v} = Q_v[\xi]
\]

or equivalently

\[
\frac{\partial \xi}{\partial t} + \nabla \cdot (\xi \mathbf{v}) = Q_v[\xi].
\]

If we are interested in a tracer that is normally measured per unit mass of fluid (which is typical when considering thermodynamic quantities) then the conservation equation would be written

\[
\frac{D}{Dt}(\varphi \rho \Delta V) = Q_m[\varphi] \rho \Delta V,
\]

where \( \varphi \) is the tracer mixing ratio — that is, the amount of tracer per unit fluid mass — and \( Q_m[\varphi] \) represents nonconservative sources per unit mass. Then, since \( \rho \Delta V \) is constant we obtain

\[
\frac{D\varphi}{Dt} = Q_m[\varphi].
\]

The source term \( Q_m[\varphi] \) is evidently equal to the rate of change of \( \varphi \) of a fluid element. When this is so, it is common to write it simply as \( \dot{\varphi} \), so that

\[
\frac{D\varphi}{Dt} = \dot{\varphi}.
\]

A tracer obeying (1.42) with \( \dot{\varphi} = 0 \) is said to be materially conserved. If a tracer is materially conserved except for the effects of nonconservative sources then it is sometimes said to be ‘semi-materially conserved’.

1.3 THE MOMENTUM EQUATION

The momentum equation is a partial differential equation that describes how the velocity or momentum of a fluid responds to internal and imposed forces. We will derive it using material methods and informally deducing the terms representing the pressure, gravitational and viscous forces.
1.3.1 Advection

Let \( \mathbf{m}(x, y, z, t) \) be the momentum-density field (momentum per unit volume) of the fluid. Thus, \( \mathbf{m} = \rho \mathbf{v} \) and the total momentum of a volume of fluid is given by the volume integral \( \int_V \mathbf{m} \, dV \). Now, for a fluid the rate of change of a momentum of an identifiable fluid mass is given by the material derivative, and by Newton's second law this is equal to the force acting on it. Thus,

\[
\frac{D}{Dt} \int_V \rho \mathbf{v} \, dV = \int_V F \, dV \quad \text{(1.43)}
\]

Now, using (1.21b) (with \( \chi \) replaced by \( \mathbf{v} \)) to transform the left-hand side of (1.43), we obtain

\[
\int_V \left( \rho \frac{D\mathbf{v}}{Dt} - \mathbf{F} \right) \, dV = 0. \quad \text{(1.44)}
\]

Because the volume is arbitrary the integrand itself must vanish and we obtain

\[
\rho \frac{D\mathbf{v}}{Dt} = \mathbf{F}, \quad \text{or} \quad \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = \frac{\mathbf{F}}{\rho}, \quad \text{(1.45a,b)}
\]

having used (1.8) to expand the material derivative.

We have thus obtained an expression for how a fluid accelerates if subject to known forces. These forces are however not all external to the fluid itself; a stress arises from the direct contact between one fluid parcel and another, giving rise to pressure and viscous forces, sometimes referred to as contact forces. Because a complete treatment of these would be very lengthy, and is available elsewhere, we treat both of these very informally and intuitively.

1.3.2 The pressure force

Within or at the boundary of a fluid the pressure is the normal force per unit area due to the collective action of molecular motion. Thus

\[
dF_p = -p \, dS. \quad \text{(1.46)}
\]

where \( p \) is the pressure, \( F_p \) is the pressure force, and \( dS \) an infinitesimal surface element. If we grant ourselves this intuitive notion, it is a simple matter to assess the influence of pressure on a fluid, for the pressure force on a volume of fluid is the integral of the pressure over its boundary and so

\[
F_p = -\int_S p \, dS. \quad \text{(1.47)}
\]

The minus sign arises because the pressure force is directed inward, whereas \( S \) is a vector normal to the surface and directed outward. Applying a form of the divergence theorem to the right-hand side gives

\[
F_p = -\int_V \nabla p \, dV \quad \text{(1.48)}
\]

where the volume \( V \) is bounded by the surface \( S \). The pressure force per unit volume is therefore just \( -\nabla p \), and inserting this into (1.45b) we obtain

\[
\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{\nabla p}{\rho} + F', \quad \text{(1.49)}
\]

where \( F' \) includes only viscous and body forces.
### 1.3 The Momentum Equation

#### 1.3.3 Viscosity and diffusion

Viscosity, like pressure, is a force due to the internal motion of molecules. The effects of viscosity are apparent in many situations—the flow of treacle or volcanic lava are obvious examples. In other situations, for example large-scale flow the atmosphere, viscosity is to a first approximation negligible. However, for a constant density fluid viscosity is the only way that energy may be removed from the fluid, so that if energy is being added in some way viscosity must ultimately become important if the fluid is to reach an equilibrium where energy input equals energy dissipation. When tea is stirred in a cup, it is viscous effects that cause the fluid to eventually stop spinning after we have removed our spoon.

A number of textbooks show that, for most Newtonian fluids, the viscous force per unit volume is equal to \( \mu \nabla^2 \mathbf{v} \), where \( \mu \) is the viscosity. Although not exact, this is an extremely good approximation for most liquids and gases. With this term, the momentum equation becomes,

\[
\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{1}{\rho} \nabla p + \nu \nabla^2 \mathbf{v}
\]  

(1.50)

where \( \nu = \mu / \rho \) is the kinematic viscosity. For gases, dimensional arguments suggest that the magnitude of \( \nu \) should be given by

\[
\nu \sim \langle \text{mean free path} \times \text{mean molecular velocity} \rangle
\]  

(1.51)

which for a typical molecular velocity of 300 m s\(^{-1}\) and a mean free path of 7 \times \(10^{-8}\) m gives the not unreasonable estimate of \(2.1 \times 10^{-5}\) m\(^2\) s\(^{-1}\), within a factor of two of the experimental value (table 1.1). Interestingly, the kinematic viscosity is less for water and mercury than it is for air.

#### 1.3.4 Hydrostatic balance

The vertical component—meaning the component parallel to the gravitational force—of the momentum equation is

\[
\frac{Dw}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g
\]  

(1.52)

where \( w \) is the vertical component of the velocity. If the fluid is static the gravitational term is balanced by the pressure term and we have

\[
\frac{\partial p}{\partial z} = -\rho g,
\]  

(1.53)
and this relation is known as hydrostatic balance, or hydrostasy. It is clear in this case the pressure at a point is given by the weight of the fluid above it, provided \( p = 0 \) at the top of the fluid. It might also appear that (1.53) would be a good approximation to (1.52) provided vertical accelerations, \( \frac{Dw}{Dt} \), are small compared to gravity, which is nearly always the case in the atmosphere and ocean. While this statement is true if we need only a reasonable approximate value of the pressure at a point or in a column, the satisfaction of this condition is not sufficient to ensure that (1.53) provides an accurate enough pressure to determine the horizontal pressure gradients responsible for producing motion. We return to this point in section 2.7.

1.4 THE EQUATION OF STATE

In three dimensions the momentum and continuity equations provide four equations, but contain five unknowns — three components of velocity, density and pressure. Obviously other equations are needed, and an equation of state relates the various thermodynamic variables to each other. The conventional equation of state is an expression that relates temperature, pressure, composition (the mass fraction of the various constituents), and density, and we may write, rather generally,

\[
p = p(\rho, T, \mu_n),
\]

where \( \mu_n \) is mass fraction of the \( n \)th constituent. An equation of this form is not the most fundamental equation of state from a thermodynamic perspective, an issue we visit later, but it connects readily measurable quantities.

For an ideal gas the equation of state is

\[
p = \rho RT,
\]

where \( R \) is the gas constant for air and \( T \) is temperature. \( (R \) is related to the universal gas constant \( R_u \) by \( R = R_u/m \) where \( m \) is the mean molecular weight of the constituents of the gas. Also, \( R = nk \) where \( k \) is Boltzmann’s constant and \( n \) is the number of molecules per unit mass.) For dry air, \( R = 287 \, \text{J kg}^{-1} \, \text{K}^{-1}. \) Air has virtually constant composition except for variations in water vapour content. A measure of this is the water vapour mixing ratio, \( w = \rho_w/\rho_d \) where \( \rho_w \) and \( \rho_d \) are the densities of water vapour and dry air, respectively, and in the atmosphere \( w \) varies between 0 and 0.03. This variation makes the gas constant in the equation of state a weak function of water vapour content. A measure of this is the water vapour mixing ratio, \( w = \rho_w/\rho_d \) where \( \rho_w \) and \( \rho_d \) are the densities of water vapour and dry air, respectively, and in the atmosphere \( w \) varies between 0 and 0.03. This variation makes the gas constant in the equation of state a weak function of water vapour mixing ratio; that is, \( p = \rho R_{\text{eff}} T \) where \( R_{\text{eff}} = R_d(1 + w R_v/R_d)/(1 + w) \) where \( R_d \) and \( R_v \) are the gas constants of dry air and water vapour. Since \( w \sim 0.01 \) the variation of \( R_{\text{eff}} \) is quite small and is often ignored, especially in theoretical studies.\(^4\)

For a liquid such as seawater no expression like (1.55) is easily derivable, and semi-empirical equations are usually resorted to. For pure water in a laboratory setting a reasonable approximation of the equation of state is \( \rho = \rho_0[1 - \beta_T (T - T_0)] \), where \( \beta_T \) is a thermal expansion coefficient and \( S_0 \) and \( T_0 \) are constants. However, in the ocean the density is affected by pressure and dissolved salts: seawater is a solution of many ions in water — chloride (\( \approx 1.9% \) by weight) sodium (1%), sulfate (0.26%), magnesium (0.13%) and so on, with a total average concentration of about 35‰ (ppt, or parts per thousand). The ratio of the fractions of these salts is more-or-less constant throughout the ocean, and their total concentration may
be parameterized by a single measure, the salinity, \( S \). Given this, the density of seawater is a function of three variables — pressure, temperature, and salinity — and we may write

\[
\alpha = \alpha(T, S, p)
\]

where \( \alpha = 1/\rho \) is the specific volume. For small variations around a reference value we have

\[
d\alpha = \left( \frac{\partial \alpha}{\partial T} \right)_S dp + \left( \frac{\partial \alpha}{\partial S} \right)_T dS + \left( \frac{\partial \alpha}{\partial p} \right)_T dT = \alpha(\beta_T dT - \beta_S dS - \beta_p dp),
\]

where the second line serves to define the thermal expansion coefficient \( \beta_T \), the saline contraction coefficient \( \beta_S \), and the compressibility coefficient \( \beta_p \) (equal to \( \alpha \) divided by the bulk modulus). These are in general not constants, but for small variations around a reference state they may be treated as such and we have

\[
\alpha = \alpha_0 \left[ 1 + \beta_T (T - T_0) - \beta_S (S - S_0) - \beta_p (p - p_0) \right].
\]

Typical values of these parameters, with variations typically encountered through the ocean, are: \( \beta_T \approx 2 (\pm 1.5) \times 10^{-4} \text{K}^{-1} \) (values increase with both temperature and pressure), \( \beta_S \approx 7.6 (\pm 0.2) \times 10^{-4} \text{ppt}^{-1} \), \( \beta_p \approx 4.1 (\pm 0.5) \times 10^{-10} \text{Pa}^{-1} \). Since the variations around the mean density are small \( (1.58) \) implies that

\[
\rho = \rho_0 \left[ 1 - \beta_T (T - T_0) + \beta_S (S - S_0) + \beta_p (p - p_0) \right].
\]

A linear equation of state for seawater is emphatically not accurate enough for quantitative oceanography; the \( \beta \) parameters in \( (1.58) \) themselves vary with pressure, temperature and (more weakly) salinity so introducing nonlinearities to the equation. The most important of these are captured by an equation of state of the form

\[
\alpha = \alpha_0 \left[ 1 + \beta_T (T - T_0) + \beta_T^* \frac{(T - T_0)^2}{2} - \beta_S (S - S_0) - \beta_p (p - p_0) \right].
\]

The starred constants \( \beta_T^* \) and \( \gamma^* \) capture the leading nonlinearities: \( \gamma^* \) is the thermobaric parameter and \( \beta_T^* \) is the second thermal expansion coefficient. Even this expression has quantitative deficiencies and more complicated semi-empirical formulae are often used if high accuracy is needed.\(^5\) More discussion is to be found in section \( 1.8.2 \).

Clearly, the equation of state introduces, in general, a sixth unknown, temperature, and we will have to introduce another physical principle — the first law of thermodynamics or the principle of energy conservation — to obtain a complete set of equations. However, if the equation of state were such that it linked only density and pressure, without introducing another variable, then the equations would be complete; the simplest case of all is a constant density fluid for which the equation of state is just \( \rho = \text{constant} \). A fluid for which the density is a function of pressure alone is called a barotropic fluid; otherwise, it is a baroclinic fluid. (In this context, ‘barotropic’ is a shortening of ‘auto-barotropic’, which is the original phrase.) Equations of state of the form \( p = C \rho^y \), where \( y \) is a constant, are sometimes called ‘polytropic’.
1.5 THE THERMODYNAMIC EQUATION

1.5.1 A few fundamentals

A fundamental postulate of thermodynamics is that the internal energy of a system in equilibrium is a function of its extensive properties volume, entropy, and the mass of its various constituents. (Extensive means that their value depends on the amount of material present, as opposed to an intensive quantity such as temperature.) For our purposes it is more convenient to divide all of these by the mass of fluid present, so expressing the internal energy per unit mass, $I$, as a function of the specific volume (or inverse density) $\alpha = \rho^{-1}$, the specific entropy $\eta$, and the mass fractions of its various components, or its chemical composition, which we parameterize as its salinity $S$. Thus we have

$$I = I(\alpha, \eta, S),$$

or an equivalent equation for entropy,

$$\eta = \eta(I, \alpha, S).$$

Given the functional forms on the right-hand sides, either of these constitutes a complete description of the macroscopic state of a system in equilibrium, and we call them the fundamental equation of state. The first differential of (1.61a) gives, formally,

$$dI = \frac{\partial I}{\partial \alpha} d\alpha + \frac{\partial I}{\partial \eta} d\eta + \frac{\partial I}{\partial S} dS.$$  

We will now ascribe physical meaning to these differentials.

Conservation of energy states that the internal energy of a body may change because of work done by or on it, or because of a heat input, or because of a change in its chemical composition. We write this as

$$dI = dQ - dW + dC$$

where $dW$ is the work done by the body, $dQ$ is the heat input to the body, and $dC$ accounts for the change in internal energy caused by a change in its chemical composition (e.g., its salinity). This is the first law of thermodynamics. It is applicable to a definite fluid mass, so we can regard $dI$ as the change in internal energy per unit mass, and similarly for the other quantities. Let us consider the causes of variations in these quantities.

**Heat Input:** The heat input $dQ$ is not the differential of any quantity, and we cannot unambiguously define the heat content of a body as a function of its state. However, the second law of thermodynamics provides a relationship between the heat input and the change in the entropy of a body, namely that in an (infinitesimal) quasi-static or reversible process, with constant composition,

$$T d\eta = dQ,$$

where $\eta$ is the specific entropy of the body. The entropy is a function of the state of a body and is, by definition, an adiabatic invariant. Entropy itself is an extensive quantity, meaning that if we double the amount of material then we double the entropy. We will be dealing with the amount of a quantity per unit mass, so that $\eta$ is the specific entropy, although we will often refer to it just as the entropy.
Work done: The work done by a body is equal to the pressure times the change in its volume. Thus, per unit mass, we have
\[ dW = p \, d\alpha, \]  
(1.65)

where \( \alpha = 1/\rho \) is the specific volume of the fluid and \( p \) is the pressure.

Composition: The change in internal energy due to compositional changes is related to the change in salinity by
\[ dC = \mu \, dS, \]  
(1.66)

where \( \mu \) is the chemical potential of the solution. The salinity of a parcel of fluid is conserved unless there are explicit sources and sinks, such as precipitation and evaporation at the surface and molecular diffusion. When these effects do occur the internal energy of a fluid parcel changes by (1.66). However, these effects are usually small, and most important effect of salinity is that it changes the density of seawater. In the atmosphere the composition of a parcel of air primarily varies according to the amount of water vapour in it; however, the main importance of water vapour is that when condensation or evaporation occurs, heat is released (or required) which provides an entropy source in (1.64).

Collecting equations (1.63) – (1.66) together we have
\[ dI = T \, d\eta - p \, d\alpha + \mu \, dS. \]  
(1.67)

We refer to this (often with \( dS = 0 \)) as the fundamental thermodynamic relation. The fundamental equation of state, (1.61), describes the properties of a particular fluid, and the fundamental relation, (1.67), expresses the conservation of energy. Much of classical thermodynamics follows from these two expressions.

1.5.2 * More thermodynamic relations

From (1.67) it follows that
\[ T = \left( \frac{\partial I}{\partial \eta} \right)_{\alpha,S}, \quad p = -\left( \frac{\partial I}{\partial \alpha} \right)_{\eta,S}, \quad \mu = \left( \frac{\partial I}{\partial S} \right)_{\eta,\alpha}. \]  
(1.68a,b,c)

These may be regarded as the defining relations for these variables; it is because of the use of (1.67), and not just the formal expression (1.62), that the pressure and temperature defined this way are indeed related to the internal motion of motion of the molecules that constitute the fluid. Note that if we write
\[ d\eta = \frac{1}{T} \, dI + \frac{p}{T} \, d\alpha - \frac{\mu}{T} \, dS, \]  
(1.69)

it is also clear that
\[ p = T \left( \frac{\partial \eta}{\partial \alpha} \right)_{I,S}, \quad T^{-1} = \left( \frac{\partial \eta}{\partial I} \right)_{\alpha,S}, \quad \mu = -T \left( \frac{\partial \eta}{\partial S} \right)_{I,\alpha}. \]  
(1.70a,b,c)
In the following derivations, we will unless noted suppose that the composition of a fluid parcel is fixed, and drop the suffix S on partial derivatives unless ambiguity might arise.

Because the right-hand side of (1.67) is equal to an exact differential, the second derivatives are independent of the order of differentiation. That is,

\[
\frac{\partial^2 I}{\partial \eta \partial \alpha} = \frac{\partial^2 I}{\partial \alpha \partial \eta} \tag{1.71}
\]

and therefore, using (1.68)

\[
\left( \frac{\partial T}{\partial \alpha} \right)_\eta = - \left( \frac{\partial p}{\partial \eta} \right)_\alpha \tag{1.72}
\]

This is one of the Maxwell relations, which are a collection of four similar relations which follow directly from the fundamental thermodynamic relation (1.67) and simple relations between second derivatives. A couple of others will be useful.

Define the enthalpy of a fluid by

\[
h \equiv I + p\alpha \tag{1.73}
\]

then, for a parcel of constant composition, (1.67) becomes

\[
dh = T \, d\eta + \alpha \, dp \tag{1.74}
\]

But \( h \) is a function only of \( \eta \) and \( p \) so that in general

\[
dh = \left( \frac{\partial h}{\partial \eta} \right)_p \, d\eta + \left( \frac{\partial h}{\partial p} \right)_\eta \, dp \tag{1.75}
\]

Comparing the last two equations we have

\[
T = \left( \frac{\partial h}{\partial \eta} \right)_p \quad \text{and} \quad \alpha = \left( \frac{\partial h}{\partial p} \right)_\eta \tag{1.76}
\]

Noting that

\[
\frac{\partial^2 h}{\partial \eta \partial p} = \frac{\partial^2 h}{\partial p \partial \eta} \tag{1.77}
\]

we evidently must have

\[
\left( \frac{\partial T}{\partial p} \right)_\eta = \left( \frac{\partial \alpha}{\partial \eta} \right)_p \tag{1.78}
\]

and this is our second Maxwell relation.

To obtain the third, we write

\[
dI = T \, d\eta - p \, d\alpha = d(T \eta) - \eta \, dT - d(p \alpha) + \alpha \, dp \tag{1.79}
\]

or

\[
dG = -\eta \, dT + \alpha \, dp \tag{1.80}
\]

where \( G \equiv I - T \eta + p \alpha \) is called the ‘Gibbs free energy’. Now, formally, we have

\[
dG = \left( \frac{\partial G}{\partial T} \right)_p \, dT + \left( \frac{\partial G}{\partial p} \right)_T \, dp \tag{1.81}
\]
Comparing the last two equations we see that \(\eta = -(\partial G/\partial T)_p\) and \(\alpha = (\partial G/\partial p)_T\). Furthermore, because

\[
\frac{\partial^2 G}{\partial p \partial T} = \frac{\partial^2 G}{\partial T \partial p}
\]  

(1.82)

we have our third Maxwell equation,

\[
\left(\frac{\partial \eta}{\partial p}\right)_T = -\left(\frac{\partial \alpha}{\partial T}\right)_p.
\]  

(1.83)

The fourth Maxwell equation, whose derivation is left to the reader, is

\[
\left(\frac{\partial \eta}{\partial \alpha}\right)_T = \left(\frac{\partial p}{\partial T}\right)_\alpha,
\]  

(1.84)

and all four Maxwell equations are summarized in the box on the next page. All of them follow from the fundamental thermodynamic relation, (1.67), which is the real silver hammer of thermodynamics.

**Equation of state revisited**

The fundamental equation of state (1.61) gives complete information about a fluid in thermodynamic equilibrium, and given this we can obtain expressions for the temperature, pressure and chemical potential using (1.68). These are also equations of state; however, each of them contains less information than the fundamental equation because a derivative has been taken, although all three together provide the same information. Equivalent to the fundamental equation of state are, using (1.74), an expression for the enthalpy as a function of pressure, entropy and composition, or, using (1.80) the Gibbs function as a function of pressure, temperature and composition. (Of these, the Gibbs function is often the most practically useful because the pressure, temperature and composition may all be measured in the laboratory.) The conventional equation of state, (1.54), is obtained by eliminating entropy between (1.68a) and (1.68b). Given the fundamental equation of state, the thermodynamic state of a body is fully specified by a knowledge of any two of \(p, \rho, T, \eta\) and \(I\), plus its composition.

One simple fundamental equation of state is to take the internal energy to be a function of density and not entropy; that is, \(I = I(\alpha)\). Bodies with such a property are called homentropic. Using (1.68), temperature and chemical potential have no role in the fluid dynamics and the density is a function of pressure alone — the defining property of a barotropic fluid. Neither water nor air are, in general, homentropic but under some circumstances the flow may be adiabatic and \(p = p(\rho)\) (e.g., problem 1.10).

In an ideal gas the molecules do not interact except by elastic collisions, and the volume of the molecules is presumed negligible compared to the total volume they occupy. The internal energy of the gas then depends only on temperature, and not on the density. A simple ideal gas is an ideal gas for which the heat capacity is constant, so that

\[
I = cT,
\]  

(1.85)

where \(c\) is a constant. Using this and the conventional ideal gas equation, \(p = \rho RT\), where \(R\) is also constant, we can infer the fundamental equation of state; however,
**Maxwell’s Relations**

The four Maxwell equations are:

\[
\begin{align*}
\frac{\partial T}{\partial \alpha} &= -\frac{\partial p}{\partial \eta}, \\
\frac{\partial T}{\partial p} &= \frac{\partial \alpha}{\partial \eta}, \\
\frac{\partial \eta}{\partial p} &= -\frac{\partial \alpha}{\partial T}, \\
\frac{\partial \eta}{\partial \alpha} &= \frac{\partial p}{\partial T}.
\end{align*}
\]  

(M.1)

These imply:

\[
\frac{\partial (T, \eta)}{\partial (p, \alpha)} = \frac{\partial T}{\partial p} \frac{\partial \eta}{\partial \alpha} - \frac{\partial T}{\partial \alpha} \frac{\partial \eta}{\partial p} = 0. 
\]  

(M.2)

we will defer that until we discuss potential temperature in section 1.5.4. A general ideal gas also obeys \( p = \rho RT \), but it has heat capacities that may be a function of temperature (but only of temperature — see problem 1.12).

**Internal energy and specific heats**

We can obtain some useful relations between the internal energy and specific heat capacities, and some useful estimates of their values, by some simple manipulations of the fundamental thermodynamic relation. Assuming that the composition of the fluid is constant (1.67) is

\[
T \, d\eta = dI + p \, d\alpha,
\]  

so that

\[
T \, d\eta = \left( \frac{\partial I}{\partial T} \right)_\alpha \, dT + \left[ \left( \frac{\partial I}{\partial \alpha} \right)_T + p \right] \, d\alpha.
\]  

From this, we see that the heat capacity at constant volume (or constant \( \alpha \)) \( c_v \) is given by

\[
c_v \equiv T \left( \frac{\partial \eta}{\partial T} \right)_\alpha = \left( \frac{\partial I}{\partial T} \right)_\alpha.
\]  

(1.88)

Thus, \( c \) in (1.85) is equal to \( c_v \).

Similarly, using (1.74) we have

\[
T \, d\eta = dh - \alpha \, dp = \left( \frac{\partial h}{\partial T} \right)_p \, dT + \left[ \left( \frac{\partial h}{\partial p} \right) - \alpha \right] \, dp.
\]  

(1.89)

The heat capacity at constant pressure, \( c_p \), is then given by

\[
c_p \equiv T \left( \frac{\partial \eta}{\partial T} \right)_p = \left( \frac{\partial h}{\partial T} \right)_p.
\]  

(1.90)

For later use, we define the ratios \( \gamma \equiv c_p/c_v \) and \( \kappa \equiv R/c_p \).

For an ideal gas \( h = I + RT = T(c_v + R) \). But \( c_p = (\partial h/\partial t)_p \), and hence \( c_p = c_v + R \), and \( (\gamma - 1)/\gamma = \kappa \). Statistical mechanics tells us that for a simple ideal
gas the internal energy is equal to $kT/2$ per molecule, or $RT/2$ per unit mass, for each excited degree of freedom, where $k$ is the Boltzmann constant and $R$ the gas constant. The diatomic molecules $N_2$ and $O_2$ that comprise most of our atmosphere have two rotational and three translational degrees of freedom, so that $I \approx 5RT/2$, and so $c_v \approx 5R/2$ and $c_p \approx 7R/2$, both being constants. These are in fact very good approximations to the measured values for the earth's atmosphere, and give $c_p \approx 10^3$ J kg$^{-1}$ K$^{-1}$. The internal energy is simply $c_v T$ and the enthalpy is $c_p T$. For a liquid, especially one containing dissolved salts such as seawater, no such simple relations are possible: the heat capacities are functions of the state of the fluid, and the internal energy is a function of pressure (or density) as well as temperature.

### 1.5.3 Thermodynamic equations for fluids

The thermodynamic relations — for example (1.67) — apply to identifiable bodies or systems; thus, the heat input affects the fluid parcel to which it is applied, and we can apply the material derivative to the above thermodynamic relations to obtain equations of motion for a moving fluid. But in doing so we make two assumptions:

(i) That locally the fluid is in thermodynamic equilibrium. This means that, although the thermodynamic quantities like temperature, pressure and density vary in space and time, locally they are related by the thermodynamic relations such as the equation of state and Maxwell's relations.

(ii) That macroscopic fluid motions are reversible and so not entropy producing. Thus, the diabatic term $dQ$ represents the entropy sources — such effects as viscous dissipation of energy, radiation, and conduction — whereas the macroscopic fluid motion itself is adiabatic.

The first point requires that the temperature variation on the macroscopic scales must be slow enough that there can exist a volume that is small compared to the scale of macroscopic variations, so that temperature is effectively constant within it, but that is also sufficiently large to contain enough molecules so that macroscopic variables such as temperature have a proper meaning. Accepting these assumptions, the expression

$$T \, d\eta = dQ \quad (1.91)$$

implies that we may write

$$T \frac{D\eta}{Dt} = \dot{Q}, \quad (1.92)$$

where $\dot{Q}$ is the rate of heat input per unit mass. Eq. (1.92) is a thermodynamic equation of motion of the fluid.

For seawater a full specification of its thermodynamic state requires a knowledge of the salinity $S$, and this is determined by the conservation equation

$$\frac{DS}{Dt} = \dot{S}, \quad (1.93)$$

where $\dot{S}$ represents effects of evaporation and precipitation at the ocean surface, and molecular diffusion. Somewhat analogously, for atmosphere the thermodynamics involve water vapour whose evolution is given by the conservation of water vapour mixing ratio

$$\frac{Dw}{Dt} = \dot{w}, \quad (1.94)$$
where $\psi$ represents the effects of condensation and evaporation. Salt has an important effect on the density of seawater, whereas the effect of water vapour on the density of air is slight.

Equation (1.92) is not a useful equation unless the entropy can be related to the other fluid variables, temperature, pressure and density. This can be done using the equation of state and the thermodynamic relations we have derived, and is the subject of the following sections. An ideal gas is the simplest case with which to start.

1.5.4 Thermodynamic equation for an ideal gas

For a fluid parcel of constant composition the fundamental thermodynamic relation is

$$dQ = dI + p \, d\alpha \quad (1.95)$$

For an ideal gas the internal energy is a function of temperature only and $dI = c_v \, dT$ (also see problems 1.12 and 1.14), so that

$$dQ = c_v \, dT + p \, d\alpha \quad \text{or} \quad dQ = c_p \, dT - \alpha \, dp, \quad (1.96a,b)$$

where the second expression is derived using $\alpha = RT/p$ and and $c_p - c_v = R$. Forming the material derivative of (1.95) gives the general thermodynamic equation

$$\frac{DI}{Dt} + p \frac{D\alpha}{Dt} = \dot{Q}. \quad (1.97)$$

Similarly, for an ideal gas (1.96a,b) respectively give

$$c_v \frac{DT}{Dt} + p \frac{D\alpha}{Dt} = \dot{Q}, \quad \text{or} \quad c_p \frac{DT}{Dt} - \frac{RT}{p} \frac{Dp}{Dt} = \dot{Q}. \quad (1.98a,b)$$

Although (1.98) are equations in the state variables $p$, $T$ and/or $\alpha$, time derivatives act on two variables and this is not convenient for many purposes. Using the mass continuity equation, (1.98b) may be written

$$c_v \frac{DT}{Dt} + p \alpha \nabla \cdot \mathbf{v} = \dot{Q}. \quad (1.99)$$

Alternatively, using the ideal gas equation we may eliminate $T$ in favor of $p$ and $\alpha$, giving the equivalent equation

$$\frac{Dp}{Dt} + y \alpha \nabla \cdot \mathbf{v} = \dot{Q} \frac{\rho R}{c_v}. \quad (1.100)$$

Potential temperature and potential density

When a fluid parcel changes pressure adiabatically, it will expand or contract and, using (1.96b), its temperature change is determined by

$$c_p \, dT = \alpha \, dp. \quad (1.101)$$

As this temperature change is not caused by diabatic effects (e.g., heating), it is useful to define a temperature-like quantity that changes only if diabatic effects are
present. To this end, we define the potential temperature, $\theta$, to be the temperature that a fluid would have if moved adiabatically to some reference pressure (often taken to be the 1000 hPa, which is close to the pressure at the earth’s surface). Thus, in adiabatic flow the potential temperature of a fluid parcel is conserved, essentially by definition, and

$$\frac{D\theta}{Dt} = 0.$$  \hfill (1.102)

For this equation to be useful we must be able to relate $\theta$ to the other thermodynamic variables. For an ideal gas we use (1.96b) and the equation of state to write the thermodynamic equation as

$$d\eta = c_p d\ln T - R d\ln p.$$  \hfill (1.103)

The definition of potential temperature then implies that

$$c_p d\ln \theta = c_p d\ln T - R d\ln p,$$  \hfill (1.104)

and this is satisfied by

$$\theta = T \left( \frac{p_R}{p} \right)^\kappa,$$  \hfill (1.105)

where $p_R$ is a reference pressure and $\kappa = R/c_p$. Note that

$$d\eta = c_p \frac{d\theta}{\theta}$$  \hfill (1.106)

and, if $c_p$ is constant,

$$\eta = c_p \ln \theta.$$  \hfill (1.107)

Equation (1.106) is in fact a general expression for potential temperature of a fluid parcel of constant composition (see section 1.8.1), but (1.107) applies only if $c_p$ is constant, as it is, to a good approximation, in the earth’s atmosphere. Using (1.104), the thermodynamic equation in the presence of heating is then

$$c_p \frac{D\theta}{Dt} = \frac{\theta}{T} \dot{Q},$$  \hfill (1.108)

with $\theta$ given by (1.105). Equations (1.99), (1.100) and (1.108) are all equivalent forms of the thermodynamic equation for an ideal gas.

The potential density, $\rho_\theta$, is the density that a fluid parcel would have if moved adiabatically and at constant composition to a reference pressure, $p_R$. If the equation of state is written as $\rho = f(p, T)$ then the potential density is just

$$\rho_\theta = f(p_R, \theta).$$  \hfill (1.109)

For an ideal gas we therefore have

$$\rho_\theta = \frac{p_R}{R \theta^\gamma},$$  \hfill (1.110)

that is, potential density is proportional to the inverse of potential temperature. We may also write (1.110) as

$$\rho_\theta = \rho \left( \frac{p_R}{p} \right)^{1/\gamma}.$$  \hfill (1.111)
Finally, for later use we note that for small variations around a reference state manipulation of the ideal gas equation gives

\[
\frac{\delta \theta}{\theta} = \frac{\delta T}{T} - \kappa \frac{\delta p}{p} = \frac{1}{\gamma} \frac{\delta p}{p} - \frac{\delta \rho}{\rho}.
\]

(1.112)

* Potential temperature and the fundamental equation of state

Eq. (1.107) is closely related to the fundamental equation of state: using \( I = c_v T \), (1.105), and the equation of state \( p = \rho RT \), we can express the entropy explicitly in terms of the density and the internal energy, to wit

\[
\eta = c_v \ln I - R \ln \rho + \text{constant}.
\]

(1.113)

This is the fundamental equation of state for a simple ideal gas. If we were to begin with this, we could straightforwardly derive all the thermodynamic quantities of interest for a simple ideal gas: for example, using (1.70a) we immediately recover \( P = \rho RT \), and from (1.70b) we obtain \( I = c_v T \). Indeed, (1.113) could be used to define a simple ideal gas, but such an a priori definition may seem a little unmotivated. Of course the heat capacities must still be determined by experiment or by a kinetic theory — they are not given by the thermodynamics, and (1.113) holds only if they are constant.

1.5.5 * Thermodynamic equation for liquids

For a liquid such as seawater no simple exact equation of state exists. Thus, although (1.108) holds at constant salinity for a liquid by virtue of the definition of potential temperature, an accurate expression then relating potential temperature to the other thermodynamic variables is nonlinear, complicated and, to most eyes, uninformative. Yet for both theoretical and modelling work a thermodynamic equation is needed to represent energy conservation, and an equation of state needed to close the system, and one of two approaches is thus generally taken: For most theoretical work and for idealized models a simple analytic but approximate equation of state is used, but in situations where more accuracy is called for, such as quantitative modelling or observational work, an accurate but complicated semi-empirical equation of state is used. This section outlines how relatively simple thermodynamic equations may be derived that are adequate in many circumstances, and that illustrate the principles used in deriving more complicated equations.

Thermodynamic equation using pressure and density

If we regard \( \eta \) as a function of pressure and density (and salinity if appropriate) we obtain

\[
T d\eta = T \left( \frac{\partial \eta}{\partial \rho} \right)_{p,S} d\rho + T \left( \frac{\partial \eta}{\partial p} \right)_{\rho,S} d\rho + T \left( \frac{\partial \eta}{\partial S} \right)_{\rho,p} dS
\]

\[
= T \left( \frac{\partial \eta}{\partial \rho} \right)_{p,S} d\rho - T \left( \frac{\partial \eta}{\partial \rho} \right)_{p,S} \left( \frac{\partial p}{\partial \rho} \right)_{\rho,S} d\rho + T \left( \frac{\partial \eta}{\partial S} \right)_{\rho,p} dS.
\]

(1.114)
1.5 The Thermodynamic Equation

From this, and using (1.92) and (1.93), we obtain for a moving fluid

\[ T \left( \frac{\partial \eta}{\partial \rho} \right)_{p,s} \frac{D\rho}{Dt} - T \left( \frac{\partial \eta}{\partial \rho} \right)_{p,S} \left( \frac{\partial \rho}{\partial p} \right)_{\eta,S} \frac{Dp}{Dt} = \dot{Q} - T \left( \frac{\partial \eta}{\partial S} \right)_{p,p} \dot{S}. \]  

(1.115)

But \((\partial p/\partial \rho)_{\eta,S} = c_s^2\) where \(c_s\) is the speed of sound (see section 1.6). This is a measurable quantity in a fluid, and often nearly constant, and so useful to keep in an equation. Then the thermodynamic equation may be written in the form

\[ \frac{D\rho}{Dt} - \frac{1}{c_s^2} \frac{Dp}{Dt} = Q[\rho], \]  

(1.116)

where \(Q[\rho] = (\partial \rho/\partial \eta)_{p,S} \dot{Q}/T - (\partial \rho/\partial S)_{p,p} \dot{S}\) appropriately represents the effects of all diabatic and salinity source terms. This form of the thermodynamic equation is valid for both liquids and gases.

**Approximations using pressure and density**

The speed of sound in a fluid is related to its compressibility — the less compressible the fluid, the greater the sound speed. In a fluid it is often sufficiently high that the second term in (1.116) can be neglected, and the thermodynamic equation takes the simple form:

\[ \frac{D\rho}{Dt} = Q[\rho]. \]  

(1.117)

This equation is a very good approximation for many laboratory fluids. Note that this equation is a thermodynamic equation, arising from the principle of conservation of energy for a liquid. It is a very different equation from the mass conservation equation, which for compressible fluids is also an evolution equation for density.

In the ocean the enormous pressures resulting from columns of seawater kilometers deep mean that although the the second term in (1.116) may be small, it is not negligible, and a better approximation results if we suppose that the pressure is given by the weight of the fluid above it — the hydrostatic approximation. In this case \(d\rho = -\rho g dz\) and (1.116) becomes

\[ \frac{D\rho}{Dt} + \frac{\rho g}{c_s^2} \frac{Dz}{Dt} = Q[\rho]. \]  

(1.118)

In the second term the height field varies much more than the density field, so a good approximation is to replace \(\rho\) by a constant, \(\rho_0\), in this term only. Taking the speed of sound also to be constant gives

\[ \frac{D}{Dt} \left[ \rho + \frac{\rho_0 z}{H_\rho} \right] = Q[\rho] \]  

(1.119)

where

\[ H_\rho = c_s^2/g \]  

(1.120)

is the density scale height of the ocean. In water, \(c_s \approx 1500\) m s\(^{-1}\) so that \(H_\rho \approx 200\) km. The quantity in square brackets in (1.119) is (in this approximation) the potential density, this being the density that a parcel would have if moved adiabatically and with constant composition to the reference height \(z = 0\). The density
scale height as defined here is due to the mean compressibility (i.e., the change in density with pressure) of seawater and, because sound speed varies only slightly in the ocean, this is nearly a constant. The adiabatic lapse rate of density is the rate at which the density of a parcel changes when undergoing an adiabatic displacement. From (1.119) it is approximately

\[- \left( \frac{\partial \rho}{\partial z} \right)_\eta \approx \frac{\rho_0 \theta}{c_s^2} \approx 5 \text{ (kg m}^{-3}\text{)/km} \]  

(1.121)

so that if a parcel is moved adiabatically from the surface to the deep ocean (5 km depth, say) its density its density will increase by about 25 kg m\(^{-3}\), a fractional density increase of about 1/40 or 2.5%.

**Thermodynamic equation using pressure and temperature**

Taking entropy to be a function of pressure and temperature (and salinity if appropriate) we have

\[ T d\eta = T \left( \frac{\partial \eta}{\partial T} \right)_{p,S} dT + T \left( \frac{\partial \eta}{\partial p} \right)_{T,S} dp + T \left( \frac{\partial \eta}{\partial S} \right)_{T,p} dS \]

\[ = c_p dT + T \left( \frac{\partial \eta}{\partial p} \right)_{T,S} dp + T \left( \frac{\partial \eta}{\partial S} \right)_{T,p} dS. \]

(1.122)

For a moving fluid, and using (1.92) and (1.93), this implies,

\[ \frac{DT}{Dt} + \frac{T}{c_p \rho^2} \left( \frac{\partial \rho}{\partial T} \right)_p \frac{dp}{dt} = Q[T]. \]

(1.123)

where \( Q[T] = \dot{Q}/c_p - T c_p^{-1} \dot{S}(\partial \eta/\partial S) \) includes the effects of the entropy and saline source terms. Now substitute the Maxwell relation (1.83) in the form

\[ \left( \frac{\partial \eta}{\partial p} \right)_T = \frac{1}{\rho^2} \left( \frac{\partial \rho}{\partial T} \right)_p \]

(1.124)

to give

\[ \frac{DT}{Dt} + \frac{T}{c_p \rho^2} \left( \frac{\partial \rho}{\partial T} \right)_p \frac{dp}{dt} = Q[T], \]

(1.125a)

or, equivalently,

\[ \frac{DT}{Dt} - \frac{T}{c_p} \left( \frac{\partial \rho}{\partial T} \right)_p \frac{dp}{dt} = \dot{Q}[T]. \]

(1.125b)

The density and temperature are related through a measurable coefficient of thermal expansion \( \beta_T \) where

\[ \left( \frac{\partial \rho}{\partial T} \right)_p = -\beta_T \rho \]

(1.126)

Equation (1.125) then becomes

\[ \frac{DT}{Dt} - \frac{\beta_T T}{c_p \rho} \frac{dp}{dt} = \dot{Q}[T]. \]

(1.127)

This is form of the thermodynamic equation is valid for both liquids and gases, and in an ideal gas \( \beta_T = 1/T \).
Approximations using pressure and temperature

Liquids are characterized by a small coefficient of thermal expansion, and it is sometimes acceptable in laboratory fluids to neglect the second term on the left-hand side of (1.127). We then obtain an equation analogous to (1.117), namely

$$\frac{DT}{Dt} = Q[T].$$

This approximation relies on the smallness of the coefficient of thermal expansion. A better approximation is to again suppose that the pressure in (1.127) varies according only to the weight of the fluid above it. Then $dp = -\rho g dz$ and (1.127) becomes

$$\frac{1}{T} \frac{DT}{Dt} + \frac{\beta_T g}{c_p} \frac{Dz}{Dt} = \frac{Q[T]}{T}. \quad (1.129)$$

For small variations of $T$, and if $\beta_T$ is nearly constant, this simplifies to

$$\frac{D}{Dt} \left( T + \frac{T_0 z}{H_T} \right) = Q[T] \quad (1.130)$$

where

$$H_T = \frac{c_p}{(\beta_T g)} \quad (1.131)$$

is the temperature scale height of the fluid. The quantity $T + T_0 z / H_T$ is (in this approximation) the potential temperature, $\theta$, so called because it is the temperature that a fluid at a depth $z$ would have if moved adiabatically to a reference depth, here taken as $z = 0$ — the temperature changing because of the work done by or on the fluid parcel as it expands or is compressed. That is,

$$\theta \approx T + \frac{\beta_T g T_0}{c_p} z \quad (1.132)$$

In seawater, however, the expansion coefficient $\beta_T$ and $c_p$ are functions of pressure and (1.132) is not good enough for quantitative calculations. With the approximate values for the ocean of $\beta_T \approx 2 \times 10^{-4} \text{K}^{-1}$ and $c_p \approx 4 \times 10^3 \text{Jkg}^{-1} \text{K}^{-1}$ we obtain $H_T \approx 2000 \text{km}$.

The adiabatic lapse rate is rate at which the temperature of a parcel changes in the vertical when undergoing an adiabatic displacement. From (1.129) it is

$$\Gamma_{ad} = -\left( \frac{\partial T}{\partial z} \right)_\eta = \frac{T g \beta_T}{c_p}. \quad (1.133)$$

In general it is a function of temperature, salinity and pressure, but it is a calculable quantity if $\beta_T$ is known. With the oceanic values above, it is approximately $0.15 \text{Kkm}^{-1}$. Again this is not accurate enough for quantitative oceanography because the expansion coefficient is a function of pressure. Nor is it a good measure of stability, because of the effects of salt.

It is interesting that the scale heights given by (1.120) and (1.131) differ so much. The first is due to the compressibility of seawater (and so related to $c_s^2$, or $\beta_p$ in (1.59)) whereas the second is due to the change of density with temperature ($\beta_T$ in (1.59)), and is the distance over which the the difference between temperature and potential temperature changes by an amount equal to the temperature
Forms of the Thermodynamic Equation

**General form**
For a parcel of constant composition the thermodynamic equation is

\[ T \frac{D \eta}{D t} = \dot{Q} \quad \text{or} \quad c_p \frac{D \ln \theta}{D t} = \frac{1}{T} \dot{Q} \]  

(T.1)

where \( \eta \) is the entropy, \( \theta \) is the potential temperature, \( c_p \ln \theta = \eta \) and \( \dot{Q} \) is the heating rate. Applying the first law of thermodynamics \( T d\eta = dI + p \, d\alpha \) gives:

\[ \frac{D I}{D t} + p \frac{D \alpha}{D t} = \dot{Q} \quad \text{or} \quad \frac{D I}{D t} + R \, T \, \nabla \cdot \mathbf{v} = \dot{Q} \]  

(T.2)

where \( I \) is the internal energy.

**Ideal gas**
For an ideal gas \( dI = c_v \, dT \), and the (adiabatic) thermodynamic equation may be written in the following equivalent, exact, forms:

\[ c_p \frac{D T}{D t} - \alpha \frac{D p}{D t} = 0, \quad \frac{D p}{D t} + \gamma p \nabla \cdot \mathbf{v} = 0, \]

\[ c_v \frac{D T}{D t} + p \alpha \nabla \cdot \mathbf{v} = 0, \quad \frac{D \theta}{D t} = 0, \]  

(T.3)

where \( \theta = T(p_{R}/p)^{\kappa} \). The two expressions on the second line are usually the most useful in modelling and theoretical work.

**Liquids**
For liquids we may usefully write the (adiabatic) thermodynamic equation as a conservation equation for potential temperature \( \theta \) or potential density \( \rho_{pot} \) and represent these in terms of other variables. For example:

\[ \frac{D \theta}{D t} = 0, \quad \theta \approx \begin{cases} T \\ T + (\beta_T g z/c_p) \end{cases} \]  

(approximately)

(with some thermal expansion),

(T.4a)

\[ \frac{D \rho_{pot}}{D t} = 0, \quad \rho_{pot} \approx \begin{cases} \rho \\ \rho + (\rho_0 g z/c_s^2) \end{cases} \]  

(very approximately)

(with some compression).  

(T.4b)

Unlike (T.3), these are not equivalent forms. More accurate semi-empirical expressions that may also include saline effects are often used for quantitative applications.
itself (i.e., by about 273 K). The two heights differ so much because the value of thermal expansion coefficient is not directly tied to the compressibility — for example, fresh water at 4° C has a zero thermal expansion, and so would have an infinite temperature scale height, but its compressibility differs little from water at 20° C.

In the atmosphere the ideal gas relationship gives $\beta_T = 1/T$ and so

$$\Gamma_{ad} = \frac{\beta}{c_p}$$ (1.134)

which is approximately $10 \text{ K km}^{-1}$. The only approximation involved in deriving this is the use of the hydrostatic relationship.

**Thermodynamic equation using density and temperature**

Taking entropy to be a function of density and temperature (and salinity if appropriate) we have

$$T d\eta = T \left( \frac{\partial \eta}{\partial T} \right)_{T,S} dT + T \left( \frac{\partial \eta}{\partial \alpha} \right)_{T,S} dp + T \left( \frac{\partial \eta}{\partial S} \right)_{T,\alpha} dS$$

$$= c_v dT + T \left( \frac{\partial \eta}{\partial \alpha} \right)_{T,S} d\alpha + T \left( \frac{\partial \eta}{\partial S} \right)_{T,\alpha} dS.$$ (1.135)

For a moving fluid this implies,

$$\frac{dT}{Dt} + \frac{T}{c_v} \left( \frac{\partial \eta}{\partial \alpha} \right)_{T,S} \frac{D\alpha}{Dt} = \frac{\dot{Q}}{c_v}.$$ (1.136)

If density is nearly constant, as in many liquids, then the second term on the left-hand side of (1.136) is small, and also $c_p \approx c_v$.

The thermodynamic equations for a fluid are summarized on page 28, and the complete equations of motion for a fluid are summarized on page 30. Also, note that for ideal gas (1.116) and (1.127) are exactly equivalent to (1.99) or (1.100) (problem 1.11).

### 1.6 SOUND WAVES

*Full of sound and fury, signifying nothing.*


We now consider, rather briefly, one of the most common phenomena in fluid dynamics yet one which is relatively unimportant for geophysical fluid dynamics — sound waves. Sound itself is not a meteorologically or oceanographically important phenomenon, except in a few special cases, for the pressure disturbance produced by sound waves is a tiny fraction of the ambient pressure and too small to be of importance for the circulation. For example, the ambient surface pressure in the atmosphere is about $10^5 \text{ Pa}$ and variations due to large-scale weather phenomena are about $10^3 \text{ Pa}$, often larger, whereas sound waves of 70 dB (i.e., a loud conversation) produce pressure variations of about 0.06 Pa. (1 dB = $20 \log_{10} (p/p_s)$ where $p_s = 2 \times 10^{-5} \text{ Pa}$.)
The Equations of Motion of a Fluid

For dry air, or for a salt-free liquid, the complete set of equations of motion may be written as follows:

The mass continuity equation:

\[
\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0. \tag{EOM.1}
\]

If density is constant this reduces to \( \nabla \cdot \mathbf{v} = 0 \).

The momentum equation:

\[
\frac{D \mathbf{v}}{Dt} = -\nabla p + \frac{\nu}{\rho} \nabla^2 \mathbf{v} + \mathbf{F}, \tag{EOM.2}
\]

where \( \mathbf{F} \) represents the effects of body forces such as gravity and \( \nu \) is the kinematic viscosity. If density is constant, or pressure is given as a function of density alone (e.g., \( p = C \rho^\gamma \)), then (EOM.1) and (EOM.2) form a complete system.

The thermodynamic equation:

\[
\frac{D \theta}{Dt} = \frac{1}{c_p} \left( \frac{\theta}{T} \right) \dot{Q}. \tag{EOM.3}
\]

where \( \dot{Q} \) represents external heating and diffusion, the latter being \( \kappa \nabla^2 \theta \) where \( \kappa \) is the diffusivity.

The equation of state:

\[
\rho = g(\theta, p) \tag{EOM.4}
\]

where \( g \) is a given function. For example, for an ideal gas, \( \rho = p^\kappa / (R \theta p^{\kappa-1}) \).

The equations describing fluid motion are called the Euler equations if the viscous term is omitted, and the Navier-Stokes equations if viscosity is included. Sometimes the Euler equations are taken to mean the momentum and mass conservation equations for an inviscid fluid of constant density.

The smallness of the disturbance produced by sound waves justifies a linearization of the equations of motion about a spatially uniform basic state that is a time-independent solution to the equations of motion. Thus, we write \( \mathbf{v} = \mathbf{v}_0 + \mathbf{v}' \), \( \rho = \rho_0 + \rho' \) (where a subscript 0 denotes a basic state and a prime denotes a perturbation) and so on, substitute in the equations of motion, and neglect terms involving products of primed quantities. By choice of our reference frame we will simplify matters further by setting \( \mathbf{v}_0 = 0 \). The linearized momentum and mass conservation equations are then

\[
\rho_0 \frac{\partial \mathbf{v}'}{\partial t} = -\nabla p', \tag{1.137a}
\]
\( \frac{\partial \rho'}{\partial t} = -\rho_0 \nabla \cdot \mathbf{v}' \). \hfill (1.137b)

(Note that these linear equations do not in themselves determine the magnitude of the disturbance.) Now, sound waves are largely adiabatic. Thus,

\[ \frac{dp}{dt} = \left( \frac{\partial p}{\partial \rho} \right)_\eta \frac{d\rho}{dt}, \hfill (1.138) \]

where \( \left( \frac{\partial p}{\partial \rho} \right)_\eta \) is the derivative at constant entropy, whose particular form is given by the equation of state for the fluid at hand. Then, from \((1.137a) - (1.138)\) we obtain a single equation for pressure,

\[ \frac{\partial^2 p'}{\partial t^2} = c_s^2 \nabla^2 p', \hfill (1.139) \]

where \( c_s^2 = \left( \frac{\partial p}{\partial \rho} \right)_\eta \). Eq. \((1.139)\) is the classical wave equation; solutions propagate at a speed \( c_s \) which may be identified as the speed of sound. For adiabatic flow in an ideal gas, manipulation of the equation of state leads to \( p = C\rho^\gamma \), where \( \gamma = c_p/c_v \), whence \( c_s^2 = \gamma p/\rho = \gamma RT \). Values of \( \gamma \) typically range from 5/3 for a monatomic gas to 7/5 for a diatomic gas and so for air, which is almost entirely diatomic, we find \( c_s \approx 350 \text{ m s}^{-1} \) at 300 K. In seawater no such theoretical approximation is easily available, but measurements show that \( c_s \approx 1500 \text{ m s}^{-1} \).

1.7 COMPRESSIBLE AND INCOMPRESSIBLE FLOW

Although there are probably no fluids of truly constant density, in many cases the density of a fluid will vary so little that it is a very good approximation to consider the density effectively constant in the mass conservation equation. The fluid is then said to be incompressible. For example, in the earth’s oceans the density varies by less that 5%, even though the pressure at the ocean bottom, a few kilometers below the surface, is several hundred times the atmospheric pressure at the surface. Let us first consider how the mass conservation equation simplifies when density is strictly constant, and then consider conditions under which treating density as constant is a good approximation.

1.7.1 Constant density fluids

If a fluid is strictly of constant density then the mass continuity equation, namely

\[ \frac{D\rho}{Dt} + \rho \nabla \cdot \mathbf{v} = 0, \hfill (1.140) \]

simplifies easily by neglecting all derivatives of density yielding

\[ \nabla \cdot \mathbf{v} = 0. \hfill (1.141) \]

The **prognostic** equation \((1.140)\) has become a **diagnostic** equation \((1.141)\), or a constraint to be satisfied by the velocity at each instant of the fluid motion. A consequence of this equation is that the volume of each material fluid element
remains constant. To see this recall the expression for the conservation of mass in the form
\[ \frac{D}{Dt}(\rho \Delta V) = 0. \]  (1.142)
If \( \rho \) is constant this reduces to an expression for volume conservation, \( D\Delta V/Dt = 0 \), whence (1.141) is recovered because \( D\Delta V/Dt = \Delta V \nabla \cdot \mathbf{v} \).

### 1.7.2 Incompressible flows

An incompressible fluid is one in which the density of a given fluid element does not change. Thus, in the mass continuity equation, (1.140), the material derivative of density is neglected and we recover (1.141). In reality no fluid is truly incompressible and for (1.141) to approximately hold we just require that
\[ |D\rho/Dt| \ll |\rho \nabla \cdot \mathbf{v}|, \]  (1.143)
and we delineate some conditions under which this inequality holds below. Our working definition of incompressibility, then, is that in an incompressible fluid density changes (from whatever cause) are so small that they have a negligible affect on the mass balance, allowing (1.140) to be replaced by (1.141). We do not need to assume that the densities of differing fluid elements are similar to each other, but in the ocean (and in most liquids) it is in fact the case that variations in density, \( \delta \rho \), are everywhere small compared to the mean density, \( \rho_0 \). That is, a sufficient condition for incompressibility is that
\[ \frac{\delta \rho}{\rho_0} \ll 1. \]  (1.144)
The atmosphere is not incompressible and (1.141) does not in general hold there. Note also that the fact that \( \nabla \cdot \mathbf{v} = 0 \) does not imply that we may independently use \( D\rho/Dt = 0 \). Indeed for a liquid with equation of state \( \rho = \rho_0 (1 - \beta_T (T - T_0)) \) and thermodynamic equation \( c_p DT/Dt = \dot{Q} \) we obtain
\[ \frac{D\rho}{Dt} = -\frac{\beta_T \rho_0}{c_p} \dot{Q}. \]  (1.145)
Furthermore, incompressibility does not necessarily imply the neglect of density variations in the momentum equation — it is only in the mass continuity equation that density variations are neglected.

**Some conditions for incompressibility**

The conditions under which incompressibility is a good approximation to the full mass continuity equation depend not only on the physical nature of the fluid but also on the flow itself. The condition that density is largely unaffected by pressure gives one necessary condition for the legitimate use of (1.141), as follows. First assume adiabatic flow, and omit the gravitational term. Then
\[ \frac{dp}{dt} = \left( \frac{\partial p}{\partial \rho} \right)_{\text{h}} \frac{d\rho}{dt} = c_s^2 \frac{d\rho}{dt}, \]  (1.146)
so that the density and pressure variations of a fluid parcel are related by
\[ \delta p \sim c_s^2 \delta \rho. \]  (1.147)
From the momentum equation we estimate

\[ \frac{U^2}{L} \sim \frac{1}{L} \frac{\delta p}{\rho_0}, \tag{1.148} \]

where \( U \) and \( L \) are typical velocities and lengths and where \( \rho_0 \) is a representative value of the density. Using (1.147) and (1.148) gives \( U^2 \sim c_s^2 \delta \rho/\rho_0 \). The incompressibility condition (1.144) then becomes

\[ \frac{U^2}{c_s^2} \ll 1. \tag{1.149} \]

That is, for a flow to be incompressible the fluid velocities must be less than the speed of sound; that is, the Mach number, \( M \equiv U/c_s \), must be small.

In the earth’s atmosphere it is apparent that density changes significantly with height. Assuming hydrostatic balance and an ideal gas, then

\[ \frac{1}{\rho} \frac{\partial p}{\partial z} = -g, \tag{1.150} \]

and if (for simplicity) we assume that atmosphere is isothermal then

\[ \frac{\partial p}{\partial z} = \left( \frac{\partial p}{\partial \rho} \right)_T \frac{\partial \rho}{\partial z} = RT_0 \frac{\partial \rho}{\partial z}. \tag{1.151} \]

Using (1.150) and (1.151) gives

\[ \rho = \rho_0 \exp(-z/H_\rho), \tag{1.152} \]

where \( H_\rho = RT_0/g \) is the (density) scale height of the atmosphere. It is easy to see that density changes are negligible only if we concern ourselves with motion less than the scale height, so this is another necessary condition for incompressibility.

In the atmosphere, although the Mach number is small for most flows, vertical displacements often exceed the scale height and in those cases the flow cannot be considered incompressible. In the ocean density changes from all causes are small and in most circumstances the ocean may be considered to contain an incompressible fluid. We return to this in the next chapter when we consider the Boussinesq equations.

### 1.8 * More Thermodynamics of Liquids

#### 1.8.1 Potential temperature, potential density and entropy

For an ideal gas we were able to derive a single prognostic equation for a single variable, potential temperature. As potential temperature is in turn simply related to the temperature and pressure, this is a useful prognostic equation. Can we achieve something similar with a more general equation of state, with non-constant coefficients of expansion?
Potential temperature

The potential temperature is defined as the temperature that a parcel would have if moved adiabatically to a given reference pressure $p_R$, often taken as $10^5$ Pa (or 1000 hPa, or 1000 mb, approximately the pressure at the sea-surface). Thus it may be calculated, at least in principle, through an integral of the form

$$\theta(S, T, p; p_R) = T + \int_{p}^{p_R} \Gamma'_a(S, T, p) \, dp$$

(1.153)

where $\Gamma'_a = (\partial T/\partial p)_{\eta}$. The potential temperature of a fluid parcel is directly related to its entropy, provided its composition does not change. We already demonstrated this for an ideal gas, and to see it explicitly in the general case let us first write the equation of state in the form

$$\eta = \eta(S, T, p).$$

(1.154)

Now, by definition of potential temperature we have

$$\eta = \eta(S, \theta; p_R) \quad \text{and} \quad \theta = \theta(\eta, S; p_R).$$

(1.155)

For a parcel of constant salinity, changes in entropy are caused only by changes in potential temperature so that

$$d\eta = \frac{\partial \eta(S, \theta; p_R)}{\partial \theta} \, d\theta.$$  

(1.156)

Now, if we express entropy as a function of temperature and pressure then

$$T \, d\eta = T \left( \frac{\partial \eta}{\partial T} \right)_p \, dT + T \left( \frac{\partial \eta}{\partial p} \right)_T \, dp$$

$$= c_p \, dT - T \left( \frac{\partial \alpha}{\partial T} \right)_p \, dp.$$  

(1.157)
using one of the Maxwell relations. Suppose a fluid parcel moves adiabatically, then $d\eta = 0$ and, by (1.156), $d\theta = 0$. That is, the potential temperature at each point along its trajectory is constant and $\theta = \theta(\eta)$. How do we evaluate this function? Simply note that the temperature at the reference pressure, $p_R$, is the potential temperature, so that directly from (1.157)

$$d\eta = c_p(p_R, \theta) \frac{d\theta}{\theta},$$

(1.158)

and $d\eta/d\theta = c_p(p_R, \theta)/\theta$. If $c_p$ is constant this integrates to

$$\eta = c_p \ln \theta + \text{constant},$$

(1.159)

as for a simple ideal gas (1.107).

Since potential temperature is conserved in adiabatic motion, the thermodynamic equation can be written

$$c_p \frac{D\theta}{Dt} = \frac{\theta}{T} \dot{Q},$$

(1.160)

where the right-hand side represents heating. (If salinity is changing, then the right-hand side should also include any saline source terms and saline diffusion. However, such terms usually have a very small effect.) This equation is equivalent to (1.116) or (1.127), although it is only useful if $\theta$ can be simply related to the other state variables. In practice empirical relationships have been derived that express potential temperature in terms of the local temperature, pressure and salinity, and density in terms of potential temperature, salinity and pressure (see section 1.8.2 for more discussion).

**Potential density**

Potential density, $\rho_\theta$, is defined as the density that a parcel would have if moved adiabatically and with fixed composition to a given reference pressure $p_R$ often, but not always, taken as $10^5 \text{Pa}$, or 1 bar. If the equation of state is of the form $\rho = \rho(S, T, p)$ then by definition we have

$$\rho_\theta = \rho(S, \theta; p_R).$$

(1.161)

For a parcel moving adiabatically (i.e., fixed salinity and entropy or potential temperature) its potential density is therefore conserved. For an ideal gas (1.161) gives $\rho_\theta = p_R/(R\theta)$ [as in (1.110)] and potential density provides no more information than potential temperature. However, in the oceans potential density accounts for the effect of salinity on density and so is a much better measure of the static stability of a column of water than density itself.

From (1.119) an approximate expression for the potential density in the ocean is

$$\rho_\theta = \left( \rho + \frac{\rho_0 dz}{c_T^2} \right).$$

(1.162)

Although this may suffice for theoretical or some modelling work, the vertical gradient of potential temperature in the ocean is often close to zero and a still more accurate, generally semi-empirical, expression is needed to determine stability properties.
Because density is so nearly constant in the ocean, it is common to subtract the amount \(1000 \text{ kg m}^{-3}\) before quoting its value, and depending on whether this value refers to \textit{in situ} density or the potential density the results are called \(\sigma_T\) (‘sigma-tee’) or \(\sigma_\theta\) (‘sigma-theta’) respectively. Thus,

\[
\sigma_T = \rho(p, T, S) - 1000, \quad \sigma_\theta = \rho(p_R, \theta, S) - 1000.
\]  

(1.163a,b)

If the potential density is referenced to a particular level, this is denoted by a subscript on the \(\sigma\). Thus, \(\sigma_2\) is the potential density referenced to 200 bars of pressure, or about 2 kilometers depth.

1.8.2 * More About Seawater

We now consider, rather didactically, some of the properties of the equation of state for seawater, noting in particular those nonlinearities that, although small, give it somewhat peculiar properties. We use a prototypical equation of state, (1.60) that, although not highly accurate except for small variations around a reference state, does capture the essential nonlinearities. That equation of state may be written as:

\[
\alpha = \alpha_0 \left[ 1 + \beta_T (1 + y^* p)(T - T_0) + \frac{\beta_T^*}{2} (T - T_0)^2 - \beta_S (S - S_0) - \beta_p (p - p_0) \right],
\]  

(1.164)

where we may take \(p_0 = 0\) and \(\beta_p = \alpha_0/c_s^2\), where \(c_s^2\) is a reference sound speed. The starred parameters are associated with the nonlinear terms: \(\beta_T^*\) is the second expansion coefficient and \(y^*\) is the ‘thermobaric parameter’, which determines the extent to which the thermal expansion of water depends on pressure. An equation of this form is useful because its coefficients can, in principle, be measured in the field or in the laboratory, and approximate values are given in table [1.2]. However, it may not be the most useful form for modelling or observational work, because \(T\) is not materially conserved. Let us use this equation to deduce various thermodynamic quantities of interest, and also transform it to a more useful form for modelling.

\textbf{Potential temperature of seawater}

It would be useful to express (1.164) in terms of materially conserved variables, and so in terms of potential temperature rather than temperature. Now, by definition the potential temperature is obtained by integrating the adiabatic lapse rate from the in situ pressure to the reference pressure (zero); that is

\[
\theta - T = \int_z^{z(p=0)} \left( \frac{\partial T}{\partial z} \right)_\eta \, dz = \int_p^0 \left( \frac{\partial T}{\partial p} \right)_\eta \, dp
\]  

(1.165)

Using (1.157), the adiabatic lapse rate is

\[
\left( \frac{\partial T}{\partial p} \right)_\eta = \frac{T}{c_p} \left( \frac{\partial \alpha}{\partial T} \right)_{p,S} = \frac{T}{c_p} \alpha_0 [\beta_T (1 + y^* p) + \beta_T^* (T - T_0)].
\]  

(1.166)

Now, \(c_p\) satisfies \(c_p = T(\partial \eta/\partial T)_p\), so that, using the Maxwell relation (1.83),

\[
\left( \frac{\partial c_p}{\partial p} \right)_{T,S} = T \left( \frac{\partial}{\partial T} \left( \frac{\partial \eta}{\partial p} \right)_T \right)_p = T \frac{\partial^2 \alpha}{\partial T^2}.
\]  

(1.167)
### Table 1.2 Various thermodynamic and equation-of-state parameters for seawater.

Specifically, these parameters may be used in the approximate equations of state (1.60) and (1.173).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_0$</td>
<td>Reference Density</td>
<td>$1.027 \times 10^3 \text{kg m}^{-3}$</td>
</tr>
<tr>
<td>$\alpha_0$</td>
<td>Reference Specific Volume</td>
<td>$9.738 \times 10^{-4} \text{m}^3 \text{kg}^{-1}$</td>
</tr>
<tr>
<td>$T_0$</td>
<td>Reference temperature</td>
<td>283 K</td>
</tr>
<tr>
<td>$S_0$</td>
<td>Reference salinity</td>
<td>35 psu $\approx 35%$</td>
</tr>
<tr>
<td>$c_s$</td>
<td>Reference sound speed</td>
<td>1490 m s$^{-1}$</td>
</tr>
<tr>
<td>$\beta_T$</td>
<td>First thermal expansion coefficient</td>
<td>$1.67 \times 10^{-4} \text{K}^{-1}$</td>
</tr>
<tr>
<td>$\beta_T^*$</td>
<td>Second thermal expansion coefficient</td>
<td>$1.00 \times 10^{-5} \text{K}^{-2}$</td>
</tr>
<tr>
<td>$\beta_S$</td>
<td>Haline contraction coefficient</td>
<td>$0.78 \times 10^{-3} \text{psu}^{-1}$</td>
</tr>
<tr>
<td>$\rho_p$</td>
<td>Inverse bulk modulus ($\approx \alpha_0/c_s^2$)</td>
<td>$4.39 \times 10^{-10} \text{m}^2 \text{kg}^{-1}$</td>
</tr>
<tr>
<td>$\gamma^*$</td>
<td>Thermobaric parameter ($\approx \gamma^*$)</td>
<td>$1.1 \times 10^{-8} \text{Pa}^{-1}$</td>
</tr>
<tr>
<td>$c_{p0}$</td>
<td>Specific heat capacity at constant pressure</td>
<td>3986 J kg$^{-1}$ K$^{-1}$</td>
</tr>
</tbody>
</table>

### Fig. 1.4

Examples of variation of potential temperature of seawater with pressure, temperature and salinity. Left panel: the sloping lines show potential temperature as a function of pressure at fixed salinity ($S = 35$ psu) and temperature (13.36°C). The solid line is computed using an accurate, empirical equation of state, the almost-coincident dashed line uses the simpler expression (1.172b) and the dotted line (labelled L) uses the linear expression (1.172c). The near vertical solid line, labelled S, shows the variation of potential temperature with salinity at fixed temperature and pressure. Right panel: Contours of the difference between temperature and potential temperature, $(T - \theta)$ in the pressure-temperature plane, for $S = 35$ psu. The solid lines use an accurate empirical formula, and the dashed lines use (1.172). The simpler equation can be improved locally, but not globally, by tuning the coefficients. (100 bars of pressure ($10^7 \text{ Pa}$ or 10 MPa) is approximately 1 km depth.)
Thus, for our equation of state, we have
\[
\left( \frac{\partial c_p}{\partial p} \right)_{T,S} = -T \alpha_0 \beta^*_T, \tag{1.168}
\]
and therefore
\[
c_p = c_{p0}(T,S) - pT \alpha_0 \beta^*_T. \tag{1.169}
\]
The first term cannot be determined solely from the conventional equation of state; in fact for seawater specific heat varies very little with temperature (of order one part in a thousand for a 10 K temperature variation). It varies more with salinity, changing by about \(-5 \text{ J kg}^{-1} \text{ K}^{-1}\) per part-per-thousand change in salinity. Thus we take
\[
c_{p0}(T,S) = c_{p1} + c_{p2}(S - S_0), \tag{1.170}
\]
where \(c_{p1}\) and \(c_{p2}\) are constants that may be experimentally determined.

Using (1.169) and (1.166) in (1.165) gives,
\[
\theta = T \exp \left\{ -\frac{c_p}{c_{p0}} \left[ 1 + \frac{1}{2} \frac{\gamma^* p + \beta^*_T (T - T_0)}{\beta_T} \right] \right\}. \tag{1.171}
\]

This equation is a relationship between \(T, \theta\) and \(p\) analogous to (1.105) for an ideal gas. The exponent itself is small, the second and third terms in square brackets are small compared to unity, and the deviations of both \(T\) and \(\theta\) from \(T_0\) are also presumed small. Taking advantage of all of this enables the expression to be rewritten, with increasing levels of approximation, as
\[
T' \approx \frac{T_0 \alpha_0 \beta_T}{c_{p0}} p \left( 1 + \frac{1}{2} \gamma^* p + T_0 \frac{\alpha_0 \beta^*_T}{c_{p0}} p \right) + \theta' \left( 1 + T_0 \frac{\alpha_0 \beta^*_T}{c_{p0}} p \right), \tag{1.172a}
\]
\[
\approx \frac{T_0 \alpha_0 \beta_T}{c_{p0}} p \left( 1 + \frac{1}{2} \gamma^* p \right) + \theta' \left( 1 + T_0 \frac{\alpha_0 \beta^*_T}{c_{p0}} p \right), \tag{1.172b}
\]
\[
\approx \frac{T_0 \alpha_0 \beta_T}{c_{p0}} p + \theta', \tag{1.172c}
\]
where \(T' = T - T_0\) and \(\theta' = \theta - T_0\). The last of the three, (1.172c), holds for a linear equation of state, and is useful for calculating approximate differences between temperature and potential temperature; making use of the hydrostatic approximation reveals that it is essentially the same as (1.132). Note that the potential temperature is related to temperature via the thermal expansion coefficient and not, as one might naively have expected, the compressibility coefficient. Plots of the difference between temperature and potential temperature, that also give both a sense of the accuracy of these simpler formula, is given in Fig. 1.4.

Using (1.172c) in the equation of state (1.164) we find that, approximately,
\[
\alpha = \alpha_0 \left[ 1 - \frac{\alpha_0 p}{c_s^2} + \beta_T (1 + y'^* p) \theta' + \frac{1}{2} \beta^*_T \theta'^2 - \beta_S (S - S_0) \right], \tag{1.173}
\]
where \(y'^* = y^* + T_0 \beta^*_T \alpha_0 / c_{p0} \approx y^*\) and \(c_s^{-2} = \alpha_0^{-2} - \beta_S^{-2} T_0 / c_p \approx \alpha_0^{-2}\) is a reference value of the speed of sound (\(y^*\) and \(y'^*\) differ by a few percent, and \(c_s^{-2}\) and \(c_s'^{-2}\) differ by only a few parts in a thousand). Given (1.173), it is in principle straightforward,
1.9 The Energy Budget

The total energy of a fluid includes the kinetic, potential and internal energies. Both fluid flow and pressure forces will in general move energy from place to place, but
we nevertheless expect energy to be conserved in an enclosed volume. Let us now consider what form energy conservation takes in a fluid.

### 1.9.1 Constant density fluid

For a constant density fluid the momentum equation and the mass continuity equation \( \nabla \cdot \mathbf{v} = 0 \), are sufficient to completely determine the evolution of a system. The momentum equation is

\[
\frac{D\mathbf{v}}{Dt} = -\nabla (\phi + \Phi) + \nu \nabla^2 \mathbf{v},
\]

(1.175)

where \( \phi = p/\rho_0 \) and \( \Phi \) is the potential for any conservative force (e.g., \( g z \) for a uniform gravitational field). We can rewrite the advective term on the left-hand side using the identity,

\[
(\mathbf{v} \cdot \nabla) \mathbf{v} = -\mathbf{v} \times \mathbf{\omega} + \nabla (\frac{\mathbf{v}^2}{2}),
\]

(1.176)

where \( \mathbf{\omega} = \nabla \times \mathbf{v} \) is the vorticity, discussed more in later chapters. Then, omitting viscosity, we have

\[
\frac{\partial \mathbf{v}}{\partial t} + \mathbf{\omega} \times \mathbf{v} = -\nabla \mathbf{B},
\]

(1.177)

where \( \mathbf{B} = (\phi + \Phi + \frac{\mathbf{v}^2}{2}) \) is the Bernoulli function for constant density flow. Consider for a moment steady flows \( \frac{\partial \mathbf{v}}{\partial t} = 0 \). Streamlines are, by definition, parallel to \( \mathbf{v} \) everywhere, and the vector \( \mathbf{v} \times \mathbf{\omega} \) is everywhere orthogonal to the streamlines, so that taking the dot product of the steady version of (1.177) with \( \mathbf{v} \) gives \( \mathbf{v} \cdot \nabla \mathbf{B} = 0 \). That is, for steady flows the Bernoulli function is constant along a streamline, and \( \frac{DB}{Dt} = 0 \).

Reverting back to the time-varying case, take the dot product with \( \mathbf{v} \) and include the density to yield

\[
\frac{1}{2} \frac{\partial \rho_0 \mathbf{v}^2}{\partial t} + \rho_0 \mathbf{v} \cdot (\mathbf{\omega} \times \mathbf{v}) = -\rho_0 \mathbf{v} \cdot \nabla \mathbf{B}
\]

(1.178)

The second term on the left-hand side vanishes identically. Defining the kinetic energy density \( K \), or energy per unit volume, by \( K = \rho_0 \frac{\mathbf{v}^2}{2} \), (1.178) becomes an expression for the rate of change of \( K \),

\[
\frac{\partial K}{\partial t} + \nabla \cdot (\rho_0 \mathbf{v} \mathbf{B}) = 0.
\]

(1.179)

Because \( \Phi \) is time-independent this may be written

\[
\frac{\partial}{\partial t} \left[ \rho_0 \left( \frac{1}{2} \mathbf{v}^2 + \Phi \right) \right] + \nabla \cdot \left[ \rho_0 \mathbf{v} \left( \frac{1}{2} \mathbf{v}^2 + \phi \right) \right] = 0.
\]

(1.180)

or

\[
\frac{\partial E}{\partial t} + \nabla \cdot [\mathbf{v} (E + p)] = 0.
\]

(1.181)

where \( E = K + \rho_0 \phi \) is the total energy density (i.e, the total energy per unit volume). This has the form of a general conservation equation in which a local change in a quantity is balanced by the divergence of its flux. However, the energy flux, \( \mathbf{v} (\rho_0 \frac{\mathbf{v}^2}{2} + \rho_0 \Phi + \rho_0 \phi) \), is not simply the velocity times the energy density \( \rho_0 (\mathbf{v}^2/2 + \Phi) \).
1.9 The Energy Budget

\( \Phi \); there is an additional term, \( \nu p \), that represents the energy transfer occurring when work is done by the fluid against the pressure force.

Now consider a volume through which there is no mass flux, for example a domain bounded by rigid walls. The rate of change of energy within that volume is then given by the integral of (1.178)

\[
\frac{d\hat{K}}{dt} = \frac{d}{dt} \int_V K \, dV = - \int_V \nabla \cdot (\rho_0 \nu B) \, dV = - \int_S \rho_0 B \nu \cdot dS = 0,
\]

using the divergence theorem, and where \( \hat{K} \) is the total kinetic energy. Thus, the total kinetic energy within the volume is conserved. Note that for a constant density fluid the gravitational potential energy, \( \hat{P} \), is given by

\[
\hat{P} = \int_V \rho_0 g z \, dV,
\]

which is a constant, not affected by a re-arrangement of the fluid. Thus, in a constant density fluid there is no exchange between kinetic energy and potential energy and the kinetic energy itself is conserved.

1.9.2 Variable density fluids

We start with the momentum equation

\[
\rho \frac{D\nu}{Dt} = -\nabla p - \rho \nabla \Phi,
\]

and take its dot product with \( \nu \) to obtain an equation for the evolution of kinetic energy,

\[
\frac{1}{2} \rho \frac{D\nu^2}{Dt} = -\nu \cdot \nabla p - \rho \nu \cdot \nabla \Phi
= -\nabla \cdot (p \nu) + p \nabla \cdot \nu - \rho \nu \cdot \nabla \Phi.
\]

From (1.86) the internal energy equation for adiabatic flow is

\[
\rho \frac{DI}{Dt} = -\rho p \frac{D\alpha}{Dt} = -p \nabla \cdot \nu
\]

where the second equality follows by use of the continuity equation. Finally, and somewhat trivially, the potential energy density obeys

\[
\rho \frac{D\Phi}{Dt} = \rho \nu \cdot \nabla \Phi.
\]

Adding (1.185), (1.186) and (1.187) we obtain

\[
\rho \frac{D}{Dt} \left( \frac{1}{2} \nu^2 + I + \Phi \right) = -\nabla \cdot (p \nu),
\]

which, on expanding the material derivative and using the mass conservation equation, becomes

\[
\frac{\partial}{\partial t} \left[ \rho \left( \frac{1}{2} \nu^2 + I + \Phi \right) \right] + \nabla \cdot \left[ \rho \nu \left( \frac{1}{2} \nu^2 + I + \Phi + p/\rho \right) \right] = 0.
\]
This may be written

\[ \frac{\partial E}{\partial t} + \nabla \cdot [v(E + p)] = 0, \]  

(1.190)

where \( E = \rho(v^2/2 + I + \Phi) \) is the total energy per unit volume, or the total energy density, of the fluid. This is the energy equation for an unforced, inviscid and adiabatic, compressible fluid. Just as for the constant density case, the energy flux contains the term \( p v \) that represents the work done against pressure and, again, the second term vanishes when integrated over a closed domain with rigid boundaries, implying that the total energy is conserved. However, now there can now be an exchange of energy between kinetic, potential and internal components. The quantity \( \sigma = I + p \alpha + \Phi = h + \Phi \) is sometimes referred to as the static energy, or the dry static energy. However, it is not a component of the globally conserved total energy; the conserved energy contains only the quantity \( I + \Phi \) plus the kinetic energy, and it is it is only the flux of static energy that affects the energetics. For an ideal gas we have \( \sigma = c_v T + RT + \Phi = c_p T + \Phi \), and if the potential is caused by a uniform gravitational field then \( \sigma = c_p T + gz \).

**Bernoulli’s theorem**

For steady flow \( \partial / \partial t = 0 \) and \( \nabla \cdot \rho v = 0 \) so that (1.189) may be written \( v \cdot \nabla B = 0 \) where \( B \) is the Bernoulli function given by

\[ B = \left( \frac{1}{2} v^2 + I + \Phi + p/\rho \right) = \left( \frac{1}{2} v^2 + h + \Phi \right). \]  

(1.191)

Thus, for steady flow only,

\[ \frac{DB}{Dt} = v \cdot \nabla B = 0, \]  

(1.192)

and the Bernoulli function is a constant along streamline. For an ideal gas in a constant gravitation field \( B = v^2/2 + c_p T + gz \).

For adiabatic flow we also have \( D\theta/Dt = 0 \). Thus, steady flow is both along surfaces of constant \( \theta \) and along surfaces of constant \( B \), and the vector

\[ l = \nabla \theta \times \nabla B \]  

(1.193)

is parallel to streamlines. A related result for unsteady flow is given in section 4.8.

1.9.3 Viscous Effects

We might expect that viscosity will always act to reduce the kinetic energy of a flow, and we will demonstrate this for a constant density fluid. Retaining the viscous term in (1.175), the energy equation becomes

\[ \frac{dE}{dt} = \frac{d}{dt} \int_V E \, dv = \mu \int_V v \cdot \nabla^2 v \, dv. \]  

(1.194)

The right hand side is negative definite. To see this we use the vector identity

\[ \nabla \times (\nabla \times v) = \nabla (\nabla \cdot v) - \nabla^2 v, \]  

(1.195)
and because $\nabla \cdot \mathbf{v} = 0$ we have $\nabla^2 \mathbf{v} = -\nabla \times \mathbf{\omega}$, where $\mathbf{\omega} \equiv \nabla \times \mathbf{v}$. Thus,

$$\frac{d \hat{E}}{dt} = -\mu \int_{V} \mathbf{v} \cdot (\nabla \times \mathbf{\omega}) \, dV = -\mu \int_{V} \mathbf{\omega} \cdot (\nabla \times \mathbf{v}) \, dV = -\mu \int_{V} \mathbf{\omega}^2 \, dV, \quad (1.196)$$

after integrating by parts, providing $\mathbf{v} \times \mathbf{\omega}$ vanishes at the boundary. Thus, viscosity acts to extract kinetic energy from the flow. The loss of kinetic energy reappears as an irreversible warming of the fluid (Joule heating'), and the total energy of the fluid is conserved, but this effect plays no role in a constant density fluid. The effect is normally locally small, at least in the earth’s ocean and atmosphere, although it is sometimes included in comprehensive General Circulation Models.

1.10 AN INTRODUCTION TO NON-DIMENSIONALIZATION AND SCALING

The units we use to measure length, velocity and so on are irrelevant to the dynamics, and not necessarily the most appropriate units for a given problem. Rather, it is convenient to express the equations of motion, so far as is possible, in so-called ‘nondimensional’ variables, by which we mean expressing every variable (such as velocity) as the ratio of its value to some reference value. For velocity the reference could, for example, be the speed of of light — but this would not be very helpful for fluid dynamical problems in the earth’s atmosphere or ocean! Rather, we should choose the reference value as a natural one for a given flow, in order that, so far as possible, the nondimensional variables are order-unity quantities, and doing this is called scaling the equations. Evidently, there is no reference velocity that is universally appropriate, and much of the art of fluid dynamics lies in choosing sensible scaling factors for the problem at hand. Non-dimensionalization plays an important role in fluid dynamics, and we introduce it here with a simple example.

1.10.1 The Reynolds number

Consider the constant-density momentum equation in Cartesian coordinates. If a typical velocity is $U$, a typical length is $L$, a typical timescale is $T$, and a typical value of the pressure deviation is $\Phi$, then the approximate sizes of the various terms in the momentum equation are given by

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\nabla \phi + \nu \nabla^2 \mathbf{v} \quad (1.197a)$$

$$\frac{U}{T} \sim \frac{U^2}{L} \sim \frac{\Phi}{L} \sim \frac{U}{L^2}. \quad (1.197b)$$

The ratio of the inertial terms to the viscous terms is $(U^2/L)/(\nu U/L^2) = UL/\nu$, and this is the Reynolds number. More formally, we can nondimensionalize the momentum equation by writing

$$\hat{\mathbf{v}} = \frac{\mathbf{v}}{U}, \quad \hat{x} = \frac{x}{L}, \quad \hat{t} = \frac{t}{T}, \quad \hat{\phi} = \frac{\phi}{\Phi}, \quad (1.198)$$

where the terms with hats on are nondimensional values of the variables and the capitalized quantities are known as scaling values, and these are the approximate magnitudes of the variables. We choose the nondimensionalization so that the
nondimensional variables are of order unity. Thus, for example, we choose $U$ so that $u = \mathcal{O}(U)$ where this should be taken to mean that the magnitude of the variable $u$ is approximately $U$, or that $u \sim U$, and we say that ‘$u$ scales like $U’$. [This $\mathcal{O}(\cdot)$ notation differs from the conventional mathematical meaning of ‘order’, in which $a = \mathcal{O}(\epsilon^\alpha)$ represents a limit in which $a/\epsilon^\alpha \to$ constant as $\epsilon \to 0$.] Thus, if there are well-defined length and velocity scales in the problem, and we choose these scales to perform the nondimensionalization, then the nondimensional variables are of order unity. That is, $\hat{u} = \mathcal{O}(1)$, and similarly for the other variables.

Because there are no external forces in this problem, appropriate scaling values for time and pressure are

$$T = \frac{L}{U}, \quad \Phi = U^2. \quad (1.199)$$

Substituting (1.198) and (1.199) into the momentum equation we obtain

$$\frac{U^2}{L} \left[ \frac{\partial \hat{\Phi}}{\partial t} + (\hat{v} \cdot \nabla) \hat{v} \right] = -\frac{U^2}{L} \nabla \hat{\Phi} + \frac{\nu U}{L^2} \nabla^2 \hat{v}, \quad (1.200)$$

where we use the convention that when $\nabla$ operates on a nondimensional variable it is a nondimensional operator. Eq. (1.200) simplifies to

$$\frac{\partial \hat{v}}{\partial t} + (\hat{v} \cdot \nabla) \hat{v} = -\nabla \hat{\Phi} + \frac{1}{Re} \nabla^2 \hat{v}, \quad (1.201)$$

where

$$Re \equiv \frac{UL}{\nu} \quad (1.202)$$

is, again, the Reynolds number. If we have chosen our length and velocity scales sensibly — that is, if we have scaled them properly — each variable in (1.201) is order unity, with the viscous term being multiplied by the parameter $1/Re$. There are two important conclusions:

(i) The ratio of the importance of the inertial terms to the viscous terms is given by the Reynolds number, defined by (1.202). In the absence of other forces, such as those due to gravity and rotation, the Reynolds number is the only non-dimensional parameter explicitly appearing in the momentum equation. Hence its value, along with the boundary conditions, controls the behaviour of the system.

(ii) More generally, by scaling the equations of motion appropriately the parameters determining the behaviour of the system become explicit. Scaling the equations is intelligent nondimensionalization.

Notes

1 Parts of the first few chapters, and many of the problems, draw on notes prepared over the years for a graduate class at Princeton University taught by Steve Garner, Isaac Held, Yoshio Kurihara, Paul Kushner and myself.

2 Joseph-Louis Lagrange (1736–1813) was a Franco-Italian, born and raised in Turin who then lived and worked mainly in Germany and France. He made notable contributions in analysis, number theory and mechanics and was recognized as one of the
greatest mathematicians of the 18th century. He laid the foundations of the calculus of variations (to wit, the 'Lagrange multiplier') and first formulated the principle of least action, and his monumental treatise *Mécanique Analytique* (1788) provides a unified analytic framework (it contains no diagrams, a feature virtually emulated in Whitaker’s *Analytical Dynamics*, 1927) for all Newtonian mechanics.

Leonard Euler (1707–1783), a Swiss mathematician who lived and worked for extended periods in Berlin and St. Petersburg, made important contributions in many areas of mathematics and mechanics, including the analytical treatment of algebra, the theory of equations, calculus, number theory and classical mechanics. He was the first to establish the form of the equations of motion of fluid mechanics, writing down both the field description of fluids and what we now call the material or advective derivative.

Truesdell (1954) points out that 'Eulerian' and 'Lagrangian' coordinates, especially the latter, are inappropriate eponyms. The so-called Eulerian description was introduced by d’Alembert in 1749 and generalized by Euler in 1752, and the so-called Lagrangian description was introduced by Euler in 1759. The modern confusion evidently stems from a monograph by Dirichlet in 1860 that credits Euler in 1757 and Lagrange in 1788 for the respective methods.

3 For example Batchelor (1967).

4 \( R_d \) and \( R_v \) are related by the molecular weights of water and dry air, \( M_v \) and \( M_d \), so that \( \alpha = R_d / R_v = M_v / M_d = 0.62 \). Rather than allow the gas constant to vary, meteorologists sometimes incorporate the variation of humidity into the definition of temperature, so that instead of \( p = \rho R_{eff} T \) we use \( p = \rho R_d T_v \), so defining the 'virtual temperature', \( T_v \). It is easy to show that \( T_v \approx \left[ 1 + w(\alpha^{-1} - 1) \right] T \). Atmospheric GCMs often use a virtual temperature.

5 The form of (1.60) was suggested by de Szoeke (2003). More accurately, and with much more complication, the international equation of state of seawater (Unesco 1981) is an empirical equation that fits measurements to an accuracy of order \( 10^{-5} \) (see Fofonoff 1985). Generally accurate formulae are also available from Mellor (1991), Wright (1997), and (with particular attention to high accuracy) McDougall et al. (2002). These are all more easily computable than the UNESCO formula. The formulae of Wright and McDougall et al. are of the form:

\[
\rho = \frac{p + p_0}{\lambda + \alpha_0 (p + p_0)}
\]

where \( \alpha_0, p_0 \) and \( \lambda \) are expressed as polynomials in potential temperature and salinity, using the Gibbs function of Feistel and Hagen (1995), which is as or more accurate than the UNESCO formula. Wright’s formula used are used for Fig. 1.5 and Fig. 1.3. Bryden (1973) provides an accurate polynomial formula for potential temperature of seawater in terms of temperature, salinity and pressure, and this is used for Fig. 1.4. In most numerical ocean models potential temperature and salinity are the prognostic thermodynamic variables and an empirical equation of state is used to compute density and potential density.

6 For a development of thermodynamics from its fundamentals see, e.g., Callen (1985).

7 Claude-Louis-Marie-Henri Navier (1785–1836) was a French civil engineer, professor at the École Polytechnique and later at the École des Ponts et Chaussées. He was an expert in road and bridge building (he developed the theory of suspension bridges) and, relatedly, made lasting theoretical contributions to the theory of elasticity, being the first to publish a set of general equations for the dynamics of an elastic solid.
In fluid mechanics, he laid down the now-called Navier-Stokes equations, including the viscous terms, in 1822.

George Gabriel Stokes (1819–1903). Irish born (in Skreen, County Sligo), he held the Lucasian chair of Mathematics at Cambridge from 1849 until his retirement. As well as having a role in the development of fluid mechanics, especially through his considerations of viscous effects, Stokes worked on the dynamics of elasticity, fluorescence, the wave theory of light, and was (rather ill-advisedly in hindsight) a proponent of the idea of an ether permeating all space.

Some sources take incompressibility to mean that density is unaffected by pressure, but this alone is insufficient to guarantee that the mass conservation equation can be approximated by $\nabla \cdot \mathbf{v} = 0$.

Following de Szoeke [2003].

These results, usually known as Bernoulli’s theorem, were developed mainly by Daniel Bernoulli (1700–1782). They were based on earlier work on the conservation of energy that Daniel had done with his father, Johann Bernoulli (1667–1748), and so perhaps should be known as Bernoullis’ theorem.

Osborne Reynolds (1842-1912) was an Irish born (Belfast) physicist who was professor of engineering at Manchester University from 1868–1905. His early work was in electricity and magnetism, but he is now most famous for his work in hydrodynamics. The ‘Reynolds number,’ which determines the ratio of inertial to viscous forces, and the ‘Reynolds stress,’ which is the stress on the mean flow due to the fluctuating components, are both named after him. He was also one of the first scientists to think about the concept of group velocity.

Further Reading

There are numerous books on hydrodynamics, some of them being:


This is a classic text in the subject, although its notation is now too dated to make it really useful as an introduction. Another very well-known text is:


This mainly treats incompressible flow. It is rather heavy going for the true beginner, but nevertheless is a very useful reference for the fundamentals.

Two other useful references are:


Both are introductions written at the advanced undergraduate/beginning graduate level, and are easier-going than Batchelor. Kundu and Cohen’s book has more material on geophysical fluid dynamics.

There are also numerous books on thermodynamics, two clear and useful ones being:


Reif’s book has become something of a classic, and Callen provides a rather more axiomatic approach.

An introduction to thermodynamic effects in fluids, with an emphasis on fundamental properties, is provided by:

Problems

It is by the solution of problems that the investigator tests the temper of his steel; he finds new methods and new outlooks, and gains a wider and freer horizon.
David Hilbert (1862–1943).

1.1 For an infinitesimal volume, informally show that
\[
\frac{D}{Dt}(\rho \varphi \Delta V) = \rho \Delta V \frac{D \varphi}{Dt},
\]  
(P1.1)
where \( \varphi \) is some (differentiable) property of the fluid. Hence informally deduce that
\[
\frac{D}{Dt} \int_V \rho \varphi \, dV = \int_V \rho \frac{D \varphi}{Dt} \, dV.
\]  
(P1.2)

1.2 Show that the derivative of an integral is given by
\[
\frac{d}{dt} \int_{x_1(t)}^{x_2(t)} \varphi(x, t) \, dx = \int_{x_1}^{x_2} \frac{\partial \varphi}{\partial t} \, dx + \frac{dx_2}{dt} \varphi(x_2, t) - \frac{dx_1}{dt} \varphi(x_1, t).
\]  
(P1.3)
By generalizing to three-dimensions show that the material derivative of an integral of a fluid property is given by
\[
\frac{D}{Dt} \int_V \varphi(x, t) \, dV = \int_V \frac{\partial \varphi}{\partial t} \, dV + \int_S \varphi \dot{v} \cdot dS = \int_V \left[ \frac{\partial \varphi}{\partial t} + \nabla \cdot (\varphi \dot{v}) \right] \, dV,
\]  
(P1.4)
where the surface integral \( \int_S \) is over the surface bounding the volume \( V \). Hence deduce that
\[
\frac{D}{Dt} \int_V \rho \varphi \, dV = \int_V \rho \frac{D \varphi}{Dt} \, dV.
\]  
(P1.5)

1.3 Why is there no diffusion term in the mass continuity equation?

1.4 By invoking Galilean invariance we can often choose, without loss of generality, the basic state for problems in sound waves to be such that \( u_0 \equiv 0 \). The perturbation velocity is then certainly larger than the basic state velocity. How can we then justify ignoring the nonlinear term in the perturbation equation, as the term \( u' \frac{\partial u'}{\partial x} \) is certainly no smaller than the linear term \( u_0 \frac{\partial u'}{\partial x} \)?

1.5 What amplitude of sound wave is required for the nonlinear terms to become important? Is this achieved at a rock concert (120 dB), or near a jet aircraft that is taking off (160 dB).

1.6 Using the observed value of molecular diffusion of heat in water, estimate how long it would take for a temperature anomaly to mix from the top of the ocean to the bottom, assuming that molecular diffusion alone is responsible. Comment on whether you think the real ocean has reached equilibrium after the last ice age (which ended about 12,000 years ago).

1.7 Consider the following flow:
\[
u = V \sin[k(x - ct)]
\]  
(P1.6)
where \( \Gamma, V, k \) and \( c \) are positive constants. (This is similar to the flow in the mid-latitude troposphere — an eastward flow increasing with height, with a transverse wave superimposed.) Suppose that \( \Gamma z > c \) for the region of interest. Consider particles located along the \( y = 0 \) axis at \( t = 0 \), and compute their position at some later time \( t \). Compare this with the streamfunction for the flow at the same time.

(Hint: Show that the meridional particle displacement is \( \eta = \psi/(u - c) \), where \( \psi \) is the streamfunction and \( u \) and \( c \) are parameters.)
1.8 ♦ Consider the two-dimensional flow

\[ u = A(y) \sin \omega t, \quad v = A(y) \cos \omega t. \]  

(P1.7)

The time-mean of this at a fixed point flow is zero. If \( A \) is independent of \( y \), then fluid parcels move clockwise in a circle. What is its radius? If \( A \) does depend on \( y \), find an approximate expression for the average drift of a particle,

\[ \lim_{t \to \infty} \frac{x(a, t)}{t} \]

where \( a \) is a particle label and \( A \) is suitably ‘small’. Be precise about what small means.

1.9 (a) Suppose that a sealed, insulated container consists of two compartments, and that one of them is filled with an ideal gas and the other is a vacuum. The partition separating the compartments is removed. How does the temperature of the gas change? (Answer: It stays the same. Explain.) Obtain an expression for the final potential temperature, in terms of the initial temperature of the gas and the volumes of the two compartments.

Reconcile your answers with the first law of thermodynamics for an ideal gas for a reversible quasi-static process,

\[ dQ = T \, d\eta = c_p \frac{d\theta}{\theta} = dI + dW = c_v \, dT + p \, d\alpha. \]  

(P1.8)

(b) A dry parcel that is ascending adiabatically through the atmosphere will generally cool as it moves to lower pressure and expands, and its potential temperature stays the same. How can this be consistent with your answer to part (a)?

1.10 Show that adiabatic flow in an ideal gas satisfies \( p \rho^{-\gamma} = \text{constant} \), where \( \gamma = c_p/c_v \).

1.11 (a) Show that for an ideal gas (1.116) is equivalent to (1.100). You may use the Maxwell relation \( (\partial \alpha/\partial \eta)_p = (\partial T/\partial p)_\eta \).

(b) Show that for an ideal gas (1.127) is equivalent to (1.99). You may use the results of part (a).

1.12 Show that it follows directly from the equation of state, \( P = RT/\alpha \), that the internal energy of an ideal gas is a function of temperature only.

Solution: From (1.87) and \( p = RT/\alpha \) we have

\[ d\eta = \frac{1}{T} \left( \frac{\partial I}{\partial T} \right)_\alpha \, dT + \left[ \frac{1}{T} \left( \frac{\partial I}{\partial \alpha} \right)_T + \frac{R}{\alpha} \right] \, d\alpha. \]  

(P1.9)

But, mathematically,

\[ d\eta = \left( \frac{\partial \eta}{\partial T} \right)_\alpha \, dT + \left( \frac{\partial \eta}{\partial \alpha} \right)_T \, d\alpha. \]  

(P1.10)

Equating the coefficients of \( dT \) and \( d\alpha \) in these two expressions gives

\[ \left( \frac{\partial \eta}{\partial T} \right)_\alpha = \frac{1}{T} \left( \frac{\partial I}{\partial T} \right)_\alpha \quad \text{and} \quad \left( \frac{\partial \eta}{\partial \alpha} \right)_T = \frac{1}{T} \left( \frac{\partial I}{\partial \alpha} \right)_T + \frac{R}{\alpha}. \]  

(P1.11)

Noting that \( \partial^2 \eta/\partial \alpha \partial T = \partial^2 \eta/\partial T \partial \alpha \) we obtain

\[ \frac{\partial^2 I}{\partial \alpha \partial T} = \frac{\partial^2 I}{\partial T \partial \alpha} - \left( \frac{\partial I}{\partial \alpha} \right)_T. \]  

(P1.12)

Thus, \( (\partial I/\partial \alpha)_T = 0 \). Because, in general, the internal energy may be considered either a function of temperature and density or temperature and pressure, this proves that for an ideal gas the internal energy is a function only of temperature.
1.13 Show that it follows directly from the equation of state \( P = RT/\alpha \), that for an ideal gas the heat capacity at constant volume, \( c_v \), is, at most, a function of temperature.

1.14 Show that for an ideal gas
\[ Td\eta = c_v dT + pd\alpha. \] (P1.13)
and that its internal energy is given by \( I = \int c_v dT \).

Solution: Let us regard \( \eta \) as a function of \( T \) and \( \alpha \), where \( \alpha \) is the specific volume \( 1/\rho \). Then
\[ Td\eta = T \left( \frac{\partial \eta}{\partial T} \right)_\alpha dT + T \left( \frac{\partial \eta}{\partial \alpha} \right)_T d\alpha \]
by definition of \( c_v \). For an ideal gas the internal energy is a function of temperature alone (problem 1.12), so that using (1.70) the pressure of a fluid \( p = T(\partial \eta/\partial \alpha)_i = T(\partial \eta/\partial \alpha)_T \) and (P1.14) becomes
\[ Td\eta = c_v dT + p d\alpha \] (P1.15)
But, in general, the fundamental thermodynamic relation is
\[ Td\eta = dI + p d\alpha. \] (P1.16)
The terms on the right hand side of (P1.15) are identifiable as the change in the internal energy and the work done on a fluid, and so \( dI = c_v dT \). The heat capacity need not necessarily be constant, although for air it very nearly is, but it must be a function of temperature only.

1.15 (a) Beginning with the expression for potential temperature for an ideal gas, \( \theta = T(p_R/p)^\kappa \), where \( \kappa = R/c_p \), show that
\[ d\theta = \frac{\theta}{T}(dT - \alpha dp), \] (P1.17)
and therefore that the first law of thermodynamics may be written as
\[ dQ = Td\eta = c_p \frac{T}{\theta} d\theta. \] (P1.18)

(b) Show that (P1.18) is more generally true, and not just for an ideal gas.

1.16 Obtain an expression for the Gibbs function for an ideal gas in terms of pressure and temperature.

1.17 From (1.113) derive the conventional equation of state for an ideal gas, and obtain expressions for the heat capacities.

1.18 Consider an ocean at rest with known vertical profiles of potential temperature and salinity, \( \theta(z) \) and \( S(z) \). Suppose we also know the equation of state in the form \( \rho = \rho(\theta, S, p) \). Obtain an expression for the buoyancy frequency. Check your expression by substituting the equation of state for an ideal gas and recovering a known expression for the buoyancy frequency.

1.19 Consider a liquid, sitting in a container, with a free surface at the top (at \( z = H \)). The liquid obeys the equation of state \( \rho = \rho_0[1 - \beta_1(T - T_0)] \), and its internal energy, \( I \), is given by \( I = c_p T \). Suppose that the fluid is heated, so that it’s temperature rises uniformly by \( \Delta T \), and the free surface rises by a small amount \( \Delta H \). Obtain an expression for the ratio of the change in internal energy to the change in gravitational potential energy (GPE) of the ocean, and show that it is related to the scale height
1.131. If global warming increases the ocean temperature by 4 K, what is the ratio of the change of \( GPE \) to the change of \( I \)? Estimate also the average rise in sea level.

**Partial solution:** The change in internal energy and in GPE are

\[
\Delta I = C \rho_1 H_1 (T_2 - T_1), \quad \Delta GPE = \rho_1 g H_1 (H_2 - H_1)/2 = \rho_1 g H_1 \beta T H_1 (T_2 - T_1)/2. \tag{P1.19}
\]

(Derive these. Use mass conservation where necessary. The subscripts 1 and 2 denote initial and final states.) Hence \( \frac{\Delta GPE}{\Delta I} = g \beta H_1 / 2C. \)

1.20 ♦ Obtain an expression, in terms of temperature and pressure, for the potential temperature of a van der Waals gas, with equation of state \((p + a/\alpha^2)(\alpha - b) = RT\), where \(a\) and \(b\) are constants. Show that it reduces to the expression for an ideal gas in the limit \(a \to 0, \ b \to 0\).
If a body is moving in any direction, there is a force, arising from the earth's rotation, which always deflects it to the right in the northern hemisphere, and to the left in the southern.

William Ferrel, *The influence of the Earth's rotation upon the relative motion of bodies near its surface*, 1858.

**CHAPTER TWO**

**Effects of Rotation and Stratification**

The atmosphere and ocean are shallow layers of fluid on a sphere in that their thickness or depth is much less than their horizontal extent. Furthermore, their motion is strongly influenced by two effects: rotation and stratification, the latter meaning that there is a mean vertical gradient of (potential) density that is often large compared with the horizontal gradient. Here we consider how the equations of motion are affected by these effects. First, we consider some elementary effects of rotation on a fluid and derive the Coriolis and centrifugal forces, and then we write down the equations of motion appropriate for motion on a sphere. Then we discuss some approximations to the equations of motion that are appropriate for large-scale flow in the ocean and atmosphere, in particular the hydrostatic and geostrophic approximations. Following this we discuss gravity waves, a particular kind of wave motion that enabled by the presence of stratification, and finally we talk about how rotation leads to the production of certain types of boundary layers — Ekman layers — in rotating fluids.

**2.1 THE EQUATIONS OF MOTION IN A ROTATING FRAME OF REFERENCE**

Newton's second law of motion, that the acceleration on a body is proportional to the imposed force divided by the body's mass, applies in so-called inertial frames of reference. The earth rotates with a period of almost 24 hours (23h 56m) relative to the distant stars, and the surface of the earth manifestly is not, in that sense, an inertial frame. Nevertheless, because the surface of the earth is moving (in fact at speeds of up to a few hundreds of meters per second) it is very convenient to describe the flow relative to the earth's surface, rather than in some inertial frame. This necessitates recasting the equations into a form that is appropriate for a rotating frame of reference, and that is the subject of this section.
Figure 2.1 A vector $C$ rotating at an angular velocity $\Omega$. It appears to be a constant vector in the rotating frame, whereas in the inertial frame it evolves according to $(dC/dt)_I = \Omega \times C$.

2.1.1 Rate of change of a vector

Consider first a vector $C$ of constant length rotating relative to an inertial frame at a constant angular velocity $\Omega$. Then, in a frame rotating with that same angular velocity it appears stationary and constant. If in small interval of time $\delta t$ the vector $C$ rotates through a small angle $\delta \lambda$ then the change in $C$, as perceived in the inertial frame, is given by (see Fig. 2.1)

$$\delta C = |C| \cos \theta \delta \lambda \hat{m}, \quad (2.1)$$

where the vector $\hat{m}$ is the unit vector in the direction of change of $C$, which is perpendicular to both $C$ and $\Omega$. But the rate of change of the angle $\lambda$ is just, by definition, the angular velocity so that $\delta \lambda = |\Omega| \delta t$ and

$$\delta C = |C||\Omega| \sin \hat{\theta} \hat{m} \delta t = \Omega \times C \delta t. \quad (2.2)$$

using the definition of the vector cross product, where $\hat{\theta} = (\pi/2 - \theta)$ is the angle between $\Omega$ and $C$. Thus

$$\left(\frac{dC}{dt}\right)_I = \Omega \times C \quad (2.3)$$

where the left hand side is the rate of change of $C$ as perceived in the inertial frame.

Now consider a vector $B$ that changes in the inertial frame. In a small time $\delta t$ the change in $B$ as seen in the rotating frame is related to the change seen in the inertial frame by

$$(\delta B)_I = (\delta B)_R + (\delta B)_{rot} \quad (2.4)$$

where the terms are, respectively, the change seen in the inertial frame, the change due to the vector itself changing as measured in the rotating frame, and the change due to the rotation. Using (2.2) $(\delta B)_{rot} = \Omega \times B \delta t$, and so the rates of change of the vector $B$ in the inertial and rotating frames are related by

$$\left(\frac{dB}{dt}\right)_I = \left(\frac{dB}{dt}\right)_R + \Omega \times B \quad . \quad (2.5)$$
This relation applies to a vector \( B \) that, as measured at any one time, is the same in both inertial and rotating frames.

### 2.1.2 Velocity and acceleration in a rotating frame

The velocity of a body is not measured to be the same in the inertial and rotating frames, so care must be taken when applying (2.5) to velocity. First apply (2.5) to \( r \), the position of a particle to obtain

\[
\frac{d\mathbf{r}}{dt}_I = \left( \frac{d\mathbf{r}}{dt} \right)_R + \Omega \times \mathbf{r} \tag{2.6}
\]

or

\[
\mathbf{v}_I = \mathbf{v}_R + \Omega \times \mathbf{r}. \tag{2.7}
\]

We refer to \( \mathbf{v}_R \) and \( \mathbf{v}_I \) as the relative and inertial velocity, respectively, and (2.7) relates the two. Apply (2.5) again, this time to the velocity \( \mathbf{v}_R \) to give

\[
\frac{d\mathbf{v}_R}{dt}_I = \left( \frac{d\mathbf{v}_R}{dt} \right)_R + \Omega \times \mathbf{v}_R, \tag{2.8}
\]

or, using (2.7)

\[
\frac{d}{dt} (\mathbf{v}_I - \Omega \times \mathbf{r})_I = \left( \frac{d\mathbf{v}_R}{dt}_R \right) + \Omega \times \mathbf{v}_R, \tag{2.9}
\]

or

\[
\frac{d\mathbf{v}_I}{dt}_I = \left( \frac{d\mathbf{v}_R}{dt}_R \right) + \Omega \times \mathbf{v}_R + \frac{d\Omega}{dt} \times \mathbf{r} + \Omega \times \left( \frac{d\mathbf{r}}{dt}_I \right). \tag{2.10}
\]

Then, noting that

\[
\frac{d\mathbf{r}}{dt}_I = \left( \frac{d\mathbf{r}}{dt} \right)_R + \Omega \times \mathbf{r} = (\mathbf{v}_R + \Omega \times \mathbf{r}), \tag{2.11}
\]

and assuming that the rate of rotation is constant, (2.10) becomes

\[
\left( \frac{d\mathbf{v}_R}{dt} \right)_R = \left( \frac{d\mathbf{v}_I}{dt} \right)_I - 2\Omega \times \mathbf{v}_R - \Omega \times (\Omega \times \mathbf{r}). \tag{2.12}
\]

This equation may be interpreted as follows. The term on the left-hand side is the rate of change of the relative velocity as measure in the rotating frame. The first term on the right-hand side is the rate of change of the inertial velocity as measured in the inertial frame (or, loosely, the inertial acceleration). Thus, by Newton’s second law, it is equal to force on a fluid parcel divided by its mass. The second and third terms on the right-hand side (including the minus signs) are the **Coriolis force** and the **centrifugal force** per unit mass. Neither of these are true forces — they may be thought of as quasi-forces (i.e., ‘as if’ forces); that is, when a body is observed from a rotating frame it seems to behave as if unseen forces are present that affect its motion. If (2.12) is written, as is common, with the terms \( +2\Omega \times \mathbf{v}_R \) and \( +\Omega \times (\Omega \times \mathbf{r}) \) on the left-hand side then these terms should be referred to as the Coriolis and centrifugal accelerations.\(^1\)
Centrifugal force
If \( r_\perp \) is the perpendicular distance from the axis of rotation (see Fig. 2.1 and substitute \( r \) for \( C \)), then, because \( \Omega \) is perpendicular to \( r_\perp, \Omega \times r = \Omega \times r_\perp \). Then, using the vector identity \( \Omega \times (\Omega \times r_\perp) = (\Omega \cdot r_\perp)\Omega - (\Omega \cdot \Omega) r_\perp \) and noting that the first term is zero, we see that the centrifugal force per unit mass is just given by

\[
F_{cc} = -\Omega \times (\Omega \times r) = \Omega^2 r_\perp. \quad (2.13)
\]
This may usefully be written as the gradient of a scalar potential,

\[
F_{cc} = -\nabla \Phi_{ce}. \quad (2.14)
\]
where \( \Phi_{ce} = -(\Omega^2 r_\perp^2)/2 = -(\Omega \times r_\perp)^2/2 \).

Coriolis force
The Coriolis force per unit mass is:

\[
F_{co} = -2\Omega \times v_R. \quad (2.15)
\]
It plays a central role in much of geophysical fluid dynamics and will be considered extensively later on. For now, we just note three basic properties:

(i) There is no Coriolis force on bodies that are stationary in the rotating frame.
(ii) The Coriolis force acts to deflect moving bodies at right angles to their direction of travel.
(iii) The Coriolis force does no work on a body, a consequence of the fact that \( v_R \cdot (\Omega \times v_R) = 0 \).

2.1.3 Momentum equation in a rotating frame
Since \( (2.12) \) simply relates the accelerations of a particle in the inertial and rotating frames, then in the rotating frame of reference the momentum equation may be written

\[
\frac{Dv}{Dt} + 2\Omega \times v = -\frac{1}{\rho} \nabla p - \nabla \Phi. \quad (2.16)
\]
We have dropped the subscript \( R \); henceforth, unless ambiguity is present, all velocities without a subscript will be considered to be relative to the rotating frame.

2.1.4 Mass and tracer conservation in a rotating frame
Let \( \phi \) be a scalar field that, in the inertial frame, obeys

\[
\frac{D\phi}{Dt} + \phi \nabla \cdot v_I = 0. \quad (2.17)
\]
Now, observers in both the rotating and inertial frame measure the same value of \( \phi \). Further, \( D\phi/Dt \) is simply the rate of change of \( \phi \) associated with a material parcel, and therefore is reference frame invariant. Thus,

\[
\left( \frac{D\phi}{Dt} \right)_R = \left( \frac{D\phi}{Dt} \right)_I \quad (2.18)
\]
2.2 Equations of Motion in Spherical Coordinates

where \((D\phi/Dt)_R = (\partial\phi/\partial t)_R + v_R \cdot \nabla\phi\) and \((D\phi/Dt)_I = (\partial\phi/\partial t)_I + v_I \cdot \nabla\phi\) and the local temporal derivatives \((\partial\phi/\partial t)_R\) and \((\partial\phi/\partial t)_I\) are evaluated at fixed locations in the rotating and inertial frames, respectively.

Further, since \(\mathbf{v} = \mathbf{v}_I - \mathbf{\Omega} \times \mathbf{r}\), we have that

\[
\nabla \cdot \mathbf{v}_I = \nabla \cdot (\mathbf{v}_I - \mathbf{\Omega} \times \mathbf{r}) = \nabla \cdot \mathbf{v}_R
\]

since \(\nabla \cdot (\mathbf{\Omega} \times \mathbf{r}) = 0\). Thus, using \((2.18)\) and \((2.19)\), \((2.17)\) is equivalent to

\[
\frac{D\phi}{Dt} + \phi \nabla \cdot \mathbf{v} = 0
\]

where all observables are measured in the rotating frame. Thus, the equation for the evolution of a scalar whose measured value is the same in rotating and inertial frames is unaltered by the presence of rotation. In particular, the mass conservation equation is unaltered by the presence of rotation.

Although we have taken \((2.18)\) as true a priori, the individual components of the material derivative differ in the rotating and inertial frames. In particular

\[
\left(\frac{\partial\phi}{\partial t}\right)_I = \left(\frac{\partial\phi}{\partial t}\right)_R - (\mathbf{\Omega} \times \mathbf{r}) \cdot \nabla \phi
\]

because \(\mathbf{\Omega} \times \mathbf{r}\) is the velocity, in the inertial frame, of a uniformly rotating body. Similarly,

\[
\mathbf{v}_I \cdot \nabla \phi = (\mathbf{v}_R + \mathbf{\Omega} \times \mathbf{r}) \cdot \nabla \phi.
\]

Adding the last two equations reprises and confirms \((2.18)\).

2.2 EQUATIONS OF MOTION IN SPHERICAL COORDINATES

The earth is very nearly spherical and it might appear obvious that we should cast our equations in spherical coordinates. Although this does turn out to be true, the presence of a centrifugal force causes some complications which we must first discuss. The reader who is willing ab initio to treat the earth as a perfect sphere and to neglect the horizontal component of the centrifugal force may skip the next section.

2.2.1 The centrifugal force and spherical coordinates

The centrifugal force is a potential force, like gravity, and so we may therefore define an ‘effective gravity’ equal to the sum of the true, or Newtonian, gravity and the centrifugal force. The Newtonian gravitational force is directed approximately toward the center of the earth, with small deviations due mainly to the earth’s oblateness. The line of action of the effective gravity will in general differ slightly from this, and therefore have a component in the ‘horizontal’ plane, that is the plane perpendicular to the radial direction. The magnitude of the centrifugal force is \(\mathbf{\Omega}^2 \mathbf{r}_\perp\), and so the effective gravity is given by

\[
\mathbf{g} = \mathbf{g}_{\text{eff}} = \mathbf{g}_{\text{grav}} + \mathbf{\Omega}^2 \mathbf{r}_\perp
\]

where \(\mathbf{g}_{\text{grav}}\) is the Newtonian gravitational force due to the gravitational attraction of the earth and \(\mathbf{r}_\perp\) is normal to the rotation vector (in the direction \(\mathbf{C}\) in Fig.
Fig. 2.2 Left: Directions of forces and coordinates in true spherical geometry. \( g \) is the effective gravity (including the centrifugal force, \( C \)) and its horizontal component is evidently non-zero. Right: a modified coordinate system, in which the vertical direction is defined by the direction of \( g \), and so the horizontal component of \( g \) is identically zero. The dashed line schematically indicates a surface of constant geopotential. The differences between the direction of \( g \) and the direction of the radial coordinate, and between the sphere and the geopotential surface, are much exaggerated and in reality are similar to the thickness of the lines themselves.

\[
g = -\nabla \Phi. \tag{2.24}\]

Surfaces of constant \( \Phi \) are not quite spherical because \( r_\perp \), and hence the centrifugal force, vary with latitude (Fig. 2.2).

The components of the centrifugal force parallel and perpendicular to the radial direction are \( \Omega^2 r \cos^2 \vartheta \) and \( \Omega^2 r \cos \vartheta \sin \vartheta \). Newtonian gravity is much larger than either of these, and at the earth’s surface the ratio of centrifugal to gravitational terms is approximately, and no more than,

\[
\alpha \approx \frac{\Omega^2 a}{g} \approx \frac{(7.27 \times 10^{-5})^2 \times 6.4 \times 10^6}{10} \approx 3 \times 10^{-3}. \tag{2.25}\]

(Note that at the equator and pole the horizontal component of the centrifugal force is zero and the effective gravity is aligned with Newtonian gravity.) The angle between \( g \) and the line to the center of the earth is given by a similar expression and so is also small, typically around \( 3 \times 10^{-3} \) radians. However, the horizontal component of the centrifugal force is still large compared to the Coriolis force, their ratio in mid-latitudes being given by

\[
\frac{\text{Horizontal centrifugal force}}{\text{Coriolis force}} \approx \frac{\Omega^2 a \cos \vartheta \sin \vartheta}{2 |u|} \approx \frac{\Omega a}{4 |u|} \approx 10, \tag{2.26}\]

using \( u = 10 \text{ m s}^{-1} \). The centrifugal term therefore dominates over the Coriolis
2.2 Equations of Motion in Spherical Coordinates

term, and is largely balanced by a pressure gradient force. Thus, if we adhered to true spherical coordinates, both the horizontal and radial components of the momentum equation would be dominated by a static balance between a pressure gradient and gravity or centrifugal terms. Although in principle there is nothing wrong with writing the equations this way, it obscures the dynamical balances involving the Coriolis force and pressure that determine the large-scale horizontal flow.

A way around this problem is to use the direction of the geopotential force to define the vertical direction, and then to regard the surfaces of constant Φ as being true spheres. The horizontal component of effective gravity is then identically zero, and we have traded a potentially large dynamical error for a very small geometric error. In fact, over time, the earth has developed an equatorial bulge to compensate for and neutralize the centrifugal force, so that the effective gravity does act in a direction virtually normal to the earth's surface; that is, the surface of the earth is an oblate spheroid of nearly constant geopotential. The geopotential Φ is then a function of the vertical coordinate alone, and for many purposes we can just take Φ = gz; that is, the direction normal to geopotential surfaces, the local vertical, is, in this approximation, taken to be the direction of increasing r in spherical coordinates. It is because the oblateness is very small (the polar diameter is about 12,714 km, whereas the equatorial diameter is about 12,756 km) that using spherical coordinates is a very accurate way to map the spheroid, and if the angle between effective gravity and a natural direction of the coordinate system were not small then more heroic measures would be called for.

If the solid earth did not bulge at the equator, the behaviour of the atmosphere and ocean would differ significantly from that of the present system. For example, the surface of the ocean is nearly a geopotential surface, and if the solid earth were exactly spherical then the ocean would perforce become much deeper at low latitudes and the ocean basins would dry out completely at high latitudes. We could still choose to use the spherical coordinate system discussed above to describe the dynamics, but the shape of the surface of the solid earth would have to be represented by a topography, with the topographic height increasing monotonically polewards nearly everywhere.

2.2.2 Some identities in spherical coordinates

The location of a point is given by the coordinates (λ, ϑ, r) where λ is the angular distance eastward (i.e., longitude), ϑ is angular distance poleward (i.e., latitude) and r is the radial distance from the center of the earth — see Fig. 2.3. (In some other fields of study co-latitude is used as a spherical coordinate.) If a is the radius of the earth, then we also define z = r – a. At a given location we may also define the Cartesian increments (δx, δy, δz) = (r cos ϑ δλ, r δϑ, δr).

For a scalar quantity φ the material derivative in spherical coordinates is

\[ \frac{Dφ}{Dt} = \frac{∂φ}{∂t} + \frac{u}{r \cos ϑ} \frac{∂φ}{∂λ} + \frac{v}{r} \frac{∂φ}{∂ϑ} + \frac{w}{r} \frac{∂φ}{∂r}, \]

where the velocity components corresponding to the coordinates (λ, ϑ, r) are

\[ (u, v, w) = \left( r \cos ϑ \frac{Dλ}{Dt}, r \frac{Dϑ}{Dt}, \frac{Dr}{Dt} \right). \]
58 Chapter 2. Effects of Rotation and Stratification

Figure 2.3 The spherical coordinate system. The orthogonal unit vectors $i$, $j$ and $k$ point in the direction of increasing longitude $\lambda$, latitude $\vartheta$, and altitude $z$. Locally, one may apply a Cartesian system with variables $x$, $y$ and $z$ measuring distances along $i$, $j$ and $k$.

That is, $u$ is the zonal velocity, $v$ is the meridional velocity and $w$ the vertical velocity. If we define $(i,j,k)$ to be the unit vectors in the direction of increasing $(\lambda, \vartheta, r)$ then

$$v = iu + jv + kw. \quad (2.29)$$

Note also that $Dz/ Dt = Dz/ Dt$.

The divergence of a vector $B = iB_\lambda + jB_\vartheta + kB_z$ is

$$\nabla \cdot B = \frac{1}{\cos \vartheta} \left[ \frac{1}{r} \frac{\partial B_\lambda}{\partial \lambda} + \frac{1}{r} \frac{\partial}{\partial \vartheta} (B_\vartheta \cos \vartheta) + \frac{\cos \vartheta}{r^2} \frac{\partial}{\partial r} (r^2 B_r) \right]. \quad (2.30)$$

The vector gradient of a scalar is:

$$\nabla \phi = i \frac{1}{r \cos \vartheta} \frac{\partial \phi}{\partial \lambda} + j \frac{1}{r \cos \vartheta} \frac{\partial \phi}{\partial \vartheta} + k \frac{\partial \phi}{\partial r}. \quad (2.31)$$

The Laplacian of a scalar is:

$$\nabla^2 \phi \equiv \nabla \cdot \nabla \phi = \frac{1}{r^2 \cos \vartheta} \left[ \frac{1}{\cos \vartheta} \frac{\partial^2 \phi}{\partial \lambda^2} + \frac{\partial}{\partial \vartheta} \left( \cos \vartheta \frac{\partial \phi}{\partial \vartheta} \right) + \cos \vartheta \frac{\partial}{\partial r} \left( r^2 \frac{\partial \phi}{\partial r} \right) \right]. \quad (2.32)$$

The curl of a vector is:

$$\text{curl} B = \nabla \times B = \frac{1}{r^2 \cos \vartheta} \begin{vmatrix} i r \cos \vartheta & j r & k \\ \frac{\partial}{\partial \lambda} & \frac{\partial}{\partial \vartheta} & \frac{\partial}{\partial r} \\ B_\lambda r \cos \vartheta & B_\vartheta r & B_r \end{vmatrix}. \quad (2.33)$$
2.2 Equations of Motion in Spherical Coordinates

The vector Laplacian $\nabla^2 B$ (used for example when calculating viscous terms in the momentum equation) may be obtained from the vector identity:

$$\nabla^2 B = \nabla(\nabla \cdot B) - \nabla \times (\nabla \times B). \quad (2.34)$$

Only in Cartesian coordinates does this take the simple form:

$$\nabla^2 B = \frac{\partial^2 B}{\partial x^2} + \frac{\partial^2 B}{\partial y^2} + \frac{\partial^2 B}{\partial z^2}. \quad (2.35)$$

The expansion in spherical coordinates is, to most eyes, rather uninformative.

Rate of change of unit vectors

In spherical coordinates the defining unit vectors are $\textbf{i}$, the unit vector pointing eastward, parallel to a line of latitude; $\textbf{j}$ is the unit vector pointing polewards, parallel to a meridian; and $\textbf{k}$, the unit vector pointing radially outward. The directions of these vectors change with location, and in fact this is the case in nearly all coordinate systems, with the notable exception of the Cartesian one, and thus their material derivative is not zero. One way to evaluate this is to consider geometrically how the coordinate axes change with position (problem 2.5). We will approach the problem a little differently, by first obtaining the effective rotation rate $\Omega_{\text{flow}}$, relative to the earth, of a unit vector as it moves with the flow, and then applying (2.3).

Specifically, let the fluid velocity be $\textbf{v} = (u, v, w)$. The meridional component, $v$, produces a displacement $r \delta \theta = v \delta t$, and this gives rise a local effective vector rotation rate around the local zonal axis of $-(v/r) \textbf{i}$, the minus sign arising because a displacement in the direction of the north pole is produced by negative rotational displacement around the $\textbf{i}$ axis. Similarly, the zonal component, $u$, produces a displacement $\delta \lambda r \cos \theta = u \delta t$ and so an effective rotation rate, but now about the earth’s rotation axis, of $u/(r \cos \theta)$. Now, a rotation around the earth’s rotation axis may be written as (see Fig. 2.4)

$$\Omega = \Omega (j \cos \theta + k \sin \theta). \quad (2.36)$$

If the scalar rotation rate is not $\Omega$ but is $u/(r \cos \theta)$, then the vector rotation rate is

$$\frac{u}{r \cos \theta} (j \cos \theta + k \sin \theta) = j \frac{u}{r} + k \frac{u \tan \theta}{r}. \quad (2.37)$$

Thus, the total rotation rate of a vector that moves with the flow is

$$\Omega_{\text{flow}} = -i \frac{v}{r} + j \frac{u}{r} + k \frac{u \tan \theta}{r}. \quad (2.38)$$

Applying (2.3) to (2.38), we find

$$\frac{Di}{Dt} = \Omega_{\text{flow}} \times i = \frac{u}{r \cos \theta} (j \sin \theta - k \cos \theta), \quad (2.39a)$$

$$\frac{Dj}{Dt} = \Omega_{\text{flow}} \times j = -i \frac{u \tan \theta}{r} - k \frac{v}{r}, \quad (2.39b)$$

$$\frac{Dk}{Dt} = \Omega_{\text{flow}} \times k = i \frac{u}{r} + j \frac{v}{r}. \quad (2.39c)$$
2.2.3 Equations of motion

**Mass Conservation and Thermodynamic Equation**

The mass conservation equation, (1.36a), expanded in spherical co-ordinates, is

\[
\frac{\partial \rho}{\partial t} + \frac{u}{r \cos \theta} \frac{\partial (\rho u}){\partial \lambda} + \frac{v}{r \cos \theta} \frac{\partial (\rho v)}{\partial \vartheta} + \frac{w}{r} \frac{\partial (\rho w)}{\partial r} = 0. \tag{2.40}
\]

Equivalently, using the form (1.36b), this is

\[
\frac{\partial \rho}{\partial t} + \frac{1}{r \cos \theta} \frac{\partial (w \rho)}{\partial \lambda} + \frac{1}{r \cos \theta} \frac{\partial (v \rho \cos \theta)}{\partial \vartheta} + \frac{1}{r^2} \frac{\partial (w r^2 \rho)}{\partial r} = 0. \tag{2.41}
\]

The thermodynamic equation, (1.108), is a tracer advection equation. Thus, using (2.27), its (adiabatic) spherical coordinate form is

\[
\frac{\partial \theta}{\partial t} = \frac{u}{r \cos \theta} \frac{\partial \theta}{\partial \lambda} + \frac{v}{r \cos \theta} \frac{\partial \theta}{\partial \vartheta} + \frac{w}{r} \frac{\partial \theta}{\partial r}, \quad \tag{2.42}
\]

and similarly for tracers such as water vapour or salt.

**Momentum Equation**

Recall that inviscid momentum equation is:

\[
\frac{Dv}{Dt} + 2 \mathbf{\Omega} \times v = -\frac{1}{\rho} \nabla p - \nabla \Phi, \tag{2.43}
\]

where \( \Phi \) is the geopotential. In spherical coordinates the directions of the coordinate axes change with position and so the component expansion of (2.43) is

\[
\frac{Dv}{Dt} = \frac{Du}{Dt} \mathbf{i} + \frac{Dv}{Dt} \mathbf{j} + \frac{Dw}{Dt} \mathbf{k} + u \frac{Di}{Dt} + v \frac{Dj}{Dt} + w \frac{Dk}{Dt}, \tag{2.44a}
\]
2.2 Equations of Motion in Spherical Coordinates

\[ \frac{D}{Dt} \mathbf{i} + \frac{D}{Dt} \mathbf{j} + \frac{D}{Dt} \mathbf{k} + \Omega_{\text{flow}} \times \mathbf{v}, \]  (2.44b)

using (2.39). Using either (2.44a) and the expressions for the rates of change of the unit vectors given in (2.39), or (2.44b) and the expression for \( \Omega_{\text{flow}} \) given in (2.38), (2.44) becomes

\[
\frac{D}{Dt} \mathbf{v} = \mathbf{i} \left( \frac{D}{Dt} u - \frac{uv \tan \vartheta}{r} + \frac{uw}{r} \right) + \mathbf{j} \left( \frac{D}{Dt} v + \frac{u^2 \tan \vartheta}{r} + \frac{vw}{r} \right) + \mathbf{k} \left( \frac{D}{Dt} w - \frac{u^2 + v^2}{r} \right). \]  (2.45)

Using the definition of a vector cross product the Coriolis term is:

\[
2\Omega \times \mathbf{v} = \begin{vmatrix}
\mathbf{i} & \mathbf{j} & \mathbf{k} \\
0 & 2\Omega \cos \vartheta & 2\Omega \sin \vartheta \\
u & v & w
\end{vmatrix}
= \mathbf{i} (2\Omega v \cos \vartheta - 2\Omega u \sin \vartheta) + \mathbf{j} 2\Omega u \sin \vartheta - \mathbf{k} 2\Omega u \cos \vartheta. \]  (2.46)

Using (2.45) and (2.46), and the gradient operator given by (2.31), the momentum equation (2.43) becomes:

\[
\frac{D}{Dt} u - \left( 2\Omega + \frac{u}{r \cos \vartheta} \right) (v \sin \vartheta - w \cos \vartheta) = -\frac{1}{\rho r \cos \vartheta} \frac{\partial p}{\partial \vartheta}, \]  (2.47a)

\[
\frac{D}{Dt} v + \frac{wv}{r} + \left( 2\Omega + \frac{u}{r \cos \vartheta} \right) u \sin \vartheta = -\frac{1}{\rho r} \frac{\partial p}{\partial \vartheta}, \]  (2.47b)

\[
\frac{D}{Dt} w - \frac{u^2 + v^2}{r} - 2\Omega u \cos \vartheta = -\frac{1}{\rho} \frac{\partial p}{\partial r} - g. \]  (2.47c)

The terms involving \( \Omega \) are called Coriolis terms, and the quadratic terms on the left-hand sides involving \( 1/r \) are often called metric terms.

2.2.4 The primitive equations

The so-called primitive equations of motion are simplifications of the above equations frequently used in atmospheric and oceanic modelling. Three related approximations are involved; these are:

(i) The hydrostatic approximation. In the vertical momentum equation the gravitational term is assumed to be balanced by the pressure gradient term, so that

\[
\frac{\partial p}{\partial z} = -\rho g. \]  (2.48)

The advection of vertical velocity, the Coriolis terms, and the metric term \( (u^2 + v^2)/r \) are all neglected.

(ii) The shallow-fluid approximation. We write \( r = a + z \) where the constant \( a \) is the radius of the earth and \( z \) increases in the radial direction. The coordinate \( r \) is then replaced by \( a \) except where it used as the differentiating argument. Thus, for example,

\[
\frac{1}{r^2} \frac{\partial (r^2 u)}{\partial r} \rightarrow \frac{\partial w}{\partial z}. \]  (2.49)
(iii) The traditional approximation. Coriolis terms in the horizontal momentum equations involving the vertical velocity, and the still smaller metric terms \( \frac{u w}{r} \) and \( \frac{v w}{r} \), are neglected.

The second and third of these approximations should be taken, or not, together, the underlying reason being that they both relate to the presumed small aspect ratio of the motion, so the approximations succeed or fail together. If we make one approximation but not the other then we are being asymptotically inconsistent, and angular momentum and energy conservation are not assured [see section 2.2.7].

The hydrostatic approximation also depends on the small aspect ratio of the flow but in a slightly different way. For large-scale flow in the terrestrial atmosphere and ocean all three approximations are in fact all very accurate approximations. We defer a more complete treatment until section 2.7, in part because a treatment of the hydrostatic approximation is done most easily in the context of the Boussinesq equations, derived in section 2.4.

Making these approximations, the momentum equations become

\[
\frac{D u}{D t} - 2 \Omega \sin \theta v - \frac{u v}{a} \tan \theta = -\frac{1}{a \rho \cos \theta} \frac{\partial p}{\partial \lambda},
\]

\[\frac{D v}{D t} + 2 \Omega \sin \theta u + \frac{u^2 \tan \theta}{a} = -\frac{1}{\rho a} \frac{\partial p}{\partial \theta},
\]

\[0 = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g.
\]

\[\text{where } D \frac{D}{D t} = \left( \frac{\partial}{\partial t} + \frac{u}{a \cos \theta} \frac{\partial}{\partial \lambda} + \frac{v}{a} \frac{\partial}{\partial \theta} + \frac{w}{\partial z} \right).
\]

We note the ubiquity of the factor \( 2 \Omega \sin \theta \), and take the opportunity to define the Coriolis parameter, \( f \equiv 2 \Omega \sin \theta \).

The corresponding mass conservation equation for a shallow fluid layer is:

\[
\frac{\partial \rho}{\partial t} + \frac{u}{a \cos \theta} \frac{\partial \rho}{\partial \lambda} + \frac{v}{a} \frac{\partial \rho}{\partial \theta} + \frac{w}{\partial z} \rho = 0.
\]

or equivalently,

\[
\frac{\partial \rho}{\partial t} + \frac{1}{a \cos \theta} \frac{\partial (u \rho)}{\partial \lambda} + \frac{1}{a \cos \theta} \frac{\partial (v \rho \cos \theta)}{\partial \theta} + \frac{\partial (w \rho)}{\partial z} = 0.
\]

2.2.5 Primitive equations in vector form

The primitive equations may be written in a compact vector form provided we make a slight reinterpretation of the material derivative of the coordinate axes. Let \( \mathbf{u} = u \mathbf{i} + v \mathbf{j} + 0 \mathbf{k} \) be the horizontal velocity. The primitive equations (2.50a) and (2.50b) may be written as

\[
\frac{D \mathbf{u}}{D t} + f \times \mathbf{u} = -\frac{1}{\rho} \nabla_z p,
\]
where \( f = f_k = 2\Omega \sin \vartheta k \) and \( \nabla_z p = [\langle a \cos \vartheta \rangle^{-1} \partial p / \partial \lambda, a^{-1} \partial p / \partial \vartheta] \), the gradient operator at constant \( z \). In (2.54) the material derivative of the horizontal velocity is given by

\[
\frac{Du}{Dt} = \frac{Du}{Dt} + j \frac{Dv}{Dt} + u \frac{Di}{Dt} + v \frac{Dj}{Dt},
\]

(2.55)

where instead of (2.39) we have

\[
\frac{Di}{Dt} = \tilde{\Omega}_{\text{flow}} \times i = j u \tan \vartheta, \quad \frac{Dj}{Dt} = -i u \tan \vartheta / a,
\]

(2.56a, b)

where \( \tilde{\Omega}_{\text{flow}} = ku \tan \vartheta / a \) [which is the vertical component of (2.38), with \( r \) replaced by \( a \)]. The advection of the horizontal wind \( u \) is still by the three-dimensional velocity \( v \). The vertical momentum equation is the hydrostatic equation, (2.50c), and the mass conservation equation is

\[
\frac{D\rho}{Dt} + \rho \nabla \cdot v = 0 \quad \text{or} \quad \frac{\partial \rho}{\partial t} + \nabla \cdot (\rho v) = 0.
\]

(2.57)

where \( D/Dt \) on a scalar is given by (2.51), and the second expression is written out in full in (2.53).

### 2.2.6 The vector invariant form of the momentum equation

The ‘vector invariant’ form of the momentum equation is so-called because it appears to take the same form in all coordinate systems — there is no advective derivative of the coordinate system to worry about. With the aid of the identity

\[
(\mathbf{v} \cdot \nabla) \mathbf{v} = -\mathbf{v} \times \mathbf{\omega} + \nabla (\mathbf{v}^2 / 2),
\]

where \( \mathbf{\omega} = \nabla \times \mathbf{v} \) is the relative vorticity, the three dimensional momentum equation may be written:

\[
\frac{\partial \mathbf{v}}{\partial t} + (2\Omega + \mathbf{\omega}) \times \mathbf{v} = -\frac{1}{\rho} \nabla p - \frac{1}{2} \nabla \mathbf{v}^2 + \mathbf{g}.
\]

(2.58)

In spherical coordinates the relative vorticity is given by:

\[
\mathbf{\omega} = \nabla \times \mathbf{v} = \frac{1}{r^2 \cos \vartheta} \begin{vmatrix} i r \cos \vartheta & j r & k \\ \partial / \partial \lambda & \partial / \partial \vartheta & \partial / \partial r \\ ur \cos \vartheta & rv & w \end{vmatrix}
\]

\[
= i \frac{1}{r} \left( \frac{\partial w}{\partial \vartheta} - \frac{\partial (rv)}{\partial r} \right) - j \frac{1}{r \cos \vartheta} \left( \frac{\partial w}{\partial \lambda} - \frac{\partial (ur \cos \vartheta)}{\partial r} \right)
\]

\[
+ k \frac{1}{r \cos \vartheta} \left( \frac{\partial v}{\partial \lambda} - \frac{\partial (u \cos \vartheta)}{\partial \vartheta} \right).
\]

(2.59)

Making the traditional and shallow fluid approximations, the horizontal components of (2.58) may be written

\[
\frac{\partial \mathbf{u}}{\partial t} + (f + k \zeta) \times \mathbf{u} + w \frac{\partial \mathbf{u}}{\partial z} = -\frac{1}{\rho} \nabla_z p - \frac{1}{2} \nabla \mathbf{u}^2,
\]

(2.60)
where \( \mathbf{u} = (u, v, 0) \), \( f = k^2 \Omega \sin \vartheta \), \( \nabla_z \) is the horizontal gradient operator (the gradient at a constant value of \( z \)), and using (2.59), \( \zeta \) is given by

\[
\zeta = \frac{1}{a \cos \vartheta} \frac{\partial v}{\partial \lambda} - \frac{1}{a \cos \vartheta} \frac{\partial}{\partial \vartheta} (u \cos \vartheta) = \frac{1}{a \cos \vartheta} \frac{\partial v}{\partial \lambda} - \frac{1}{a} \frac{\partial u}{\partial \vartheta} + \frac{u}{a} \tan \vartheta. \tag{2.61}
\]

The separate components of the momentum equation are given by:

\[
\frac{\partial u}{\partial t} - (f + \zeta) v + w \frac{\partial u}{\partial z} = -\frac{1}{a \rho \cos \vartheta} \left( \frac{1}{\rho} \frac{\partial p}{\partial \lambda} + \frac{1}{2} \frac{\partial u^2}{\partial \lambda} \right), \tag{2.62}
\]

and

\[
\frac{\partial v}{\partial t} + (f + \zeta) u + w \frac{\partial v}{\partial z} = -\frac{1}{a} \left( \frac{1}{\rho} \frac{\partial p}{\partial \vartheta} + \frac{1}{2} \frac{\partial u^2}{\partial \vartheta} \right). \tag{2.63}
\]

Related expressions are given problem 2.2 and we treat vorticity at greater length in chapter 4.

### 2.2.7 Angular Momentum

The zonal momentum equation can be usefully expressed as a statement about axial angular momentum; that is, angular momentum about the rotation axis. The zonal angular momentum per unit mass is the component of angular momentum in the direction of the axis of rotation and it is given by, without making any shallow atmosphere approximation,

\[
m = (u + \Omega r \cos \vartheta) r \cos \vartheta. \tag{2.64}
\]

The evolution equation for this quantity follows from the zonal momentum equation and has the simple form

\[
\frac{Dm}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial \lambda}, \tag{2.65}
\]

where the material derivative is

\[
\frac{D}{Dt} = \frac{\partial}{\partial t} + \frac{\mathbf{u}}{r \cos \vartheta} \frac{\partial}{\partial \lambda} + \frac{v}{r \sin \vartheta} \frac{\partial}{\partial \vartheta} + \frac{w}{r} \frac{\partial}{\partial r}. \tag{2.66}
\]

Using the mass continuity equation, this can be written as

\[
\frac{D \rho m}{Dt} + \rho m \nabla \cdot \mathbf{v} = -\frac{\partial p}{\partial \lambda} \tag{2.67}
\]

or

\[
\frac{\partial \rho m}{\partial t} + \frac{1}{r \cos \vartheta} \frac{\partial (\rho um)}{\partial \lambda} + \frac{1}{r \cos \vartheta} \frac{\partial}{\partial \vartheta} (\rho um \cos \vartheta) + \frac{\partial}{\partial z} (\rho mw) = -\frac{\partial p}{\partial \lambda}. \tag{2.68}
\]

This is an angular momentum conservation equation.

If the fluid is confined to a shallow layer near the surface of a sphere, then we may replace \( r \), the radial coordinate, by \( a \), the radius of the sphere, in the definition of \( m \), and we define \( \tilde{m} = (u + \Omega a \cos \vartheta) a \cos \vartheta \). Then (2.65) is replaced by

\[
\frac{D \tilde{m}}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial \lambda}. \tag{2.69}
\]
2.2 Equations of Motion in Spherical Coordinates

where now

\[
\frac{D}{Dt} = \frac{\partial}{\partial t} + \frac{u}{a \cos \theta} \frac{\partial}{\partial \lambda} + \frac{v}{a} \frac{\partial}{\partial \theta} + w \frac{\partial}{\partial z}.
\]

(2.70)

Using mass continuity this may be written as

\[
\frac{\partial \rho \tilde{m}}{\partial t} + \frac{u}{a \cos \theta} \frac{\partial \tilde{m}}{\partial \lambda} + \frac{v}{a} \frac{\partial \tilde{m}}{\partial \theta} + w \frac{\partial \tilde{m}}{\partial z} = -\frac{1}{\rho} \frac{\partial \rho}{\partial \lambda}.
\]

(2.71)

which is the appropriate angular momentum conservation equation for a shallow atmosphere.

*From angular momentum to the spherical component equations*

An alternative way to derive the three components of the momentum equation in spherical polar coordinates is to begin with (2.65) and the principle of conservation of energy. That is, we take the equations for conservation of angular momentum and energy as true a priori and demand that the forms of the momentum equation be constructed to satisfy these. Expanding the material derivative in (2.65), noting that \(D_{r}/Dt = w\) and \(D \cos \theta /Dt = -(v/r) \sin \theta\), immediately gives (2.47a). Multiplication by \(u\) then yields

\[
u \frac{Du}{Dt} - 2\Omega uv \sin \theta + 2\Omega uw \cos \theta - \frac{u^2 v \tan \theta}{r} + \frac{u^2 w}{r} = -\frac{u}{\rho r \cos \theta} \frac{\partial p}{\partial \lambda}.
\]

(2.72)

Now suppose that the meridional and vertical momentum equations are of the form

\[\begin{align*}
\frac{Dv}{Dt} + \text{Coriolis and metric terms} &= -\frac{1}{\rho r} \frac{\partial p}{\partial \theta} \quad (2.73a) \\
\frac{Dw}{Dt} + \text{Coriolis and metric terms} &= -\frac{1}{\rho} \frac{\partial p}{\partial \theta}, \quad (2.73b)
\end{align*}\]

but that we do not know what form the Coriolis and metric terms take. To determine that form, construct the kinetic energy equation by multiplying (2.73) by \(v\) and \(w\), respectively. Now, the metric terms must vanish when we sum the resulting equations along with (2.72), so that (2.73a) must contain the Coriolis term \(2\Omega u \sin \theta\) as well as the metric term \(u^2 \tan \theta /r\), and (2.73b) must contain \(-2\Omega u \cos \theta\) as well as the metric term \(u^2 /r\). But if (2.73b) contains the term \(u^2 /r\) it must also contain the term \(v^2 /r\) by isotropy, and therefore (2.73a) must also contain the term \(v w /r\). In this way, (2.47) is precisely reproduced, although the skeptic might argue that the uniqueness of the form has not been proven.

A particular advantage of this approach arises in determining the appropriate momentum equations that conserve angular momentum and energy in the shallow-fluid approximation. We begin with (2.69) and expand to obtain (2.50a). Multiplying by \(u\) gives

\[
\frac{Du}{Dt} - 2\Omega uv \sin \theta - \frac{u^2 v \tan \theta}{a} = -\frac{u}{\rho a \cos \theta} \frac{\partial p}{\partial \lambda}.
\]

(2.74)

To ensure energy conservation, the meridional momentum equation must contain the Coriolis term \(2\Omega u \sin \theta\) and the metric term \(u^2 \tan \theta /a\), but the vertical momentum
equation must have neither of the metric terms appearing in (2.47c). Thus we deduce the following equations:

\[
\begin{align*}
\frac{D u}{D t} - \left( 2\Omega \sin \vartheta + \frac{u \tan \vartheta}{a} \right) v &= -\frac{1}{\rho a \cos \vartheta} \frac{\partial p}{\partial \lambda}, \\
\frac{D v}{D t} + \left( 2\Omega \sin \vartheta + \frac{u \tan \vartheta}{a} \right) u &= -\frac{1}{\rho a} \frac{\partial p}{\partial \vartheta}, \\
\frac{D w}{D t} &= -\frac{1}{\rho} \frac{\partial \rho}{\partial r} - g.
\end{align*}
\] (2.75)

This equation set, when used in conjunction with the thermodynamic and mass continuity equations, conserves appropriate forms of angular momentum and energy. In the hydrostatic approximation the material derivative of \( w \) in (2.75b) is additionally neglected. Thus, the hydrostatic approximation is mathematically and physically consistent with the shallow-fluid approximation, but it is an additional approximation with slightly different requirements that one may choose, rather than being required, to make. From an asymptotic perspective, the difference lies in the small parameter necessary for either approximation to hold, namely

Shallow fluid and traditional approximations: \( \gamma = \frac{H}{a} \ll 1 \) (2.76a)

Small aspect ratio for hydrostatic approximation: \( \alpha = \frac{H}{L} \ll 1 \). (2.76b)

where \( L \) is the horizontal scale of the motion and \( a \) is the radius of the earth. For hemispheric or global scale phenomena \( L \sim a \) and the two approximations coincide. (Requirement (2.76b) for the hydrostatic approximation is derived in section 2.7.)

2.3 CARTESIAN APPROXIMATIONS: THE TANGENT PLANE

2.3.1 The f-plane

Although the rotation of the earth is central for many dynamical phenomena, the sphericity of the earth is not always so. This is especially true for phenomena on a scale somewhat smaller than global where the use of spherical coordinates becomes awkward, and it is more convenient to use a locally Cartesian representation of the equations. Referring to Fig. 2.4 we will define a plane tangent to the surface of the earth at a latitude \( \vartheta_0 \), and then use a Cartesian coordinate system \((x, y, z)\) to describe motion on that plane. For small excursions on the plane, \((x, y, z) \approx (a\lambda \cos \vartheta_0, a(\vartheta - \vartheta_0), z)\). Consistently, the velocity is \( \mathbf{v} = (u, v, w) \), so that \( u, v \) and \( w \) are the components of the velocity in the tangent plane. These are approximately in the east-west, north-south and vertical directions, respectively.

The momentum equations for flow in this plane are then

\[
\begin{align*}
\frac{\partial u}{\partial t} + (\mathbf{v} \cdot \nabla) u + 2\Omega_y w - 2\Omega_z v &= -\frac{1}{\rho} \frac{\partial p}{\partial x}, \\
\frac{\partial v}{\partial t} + (\mathbf{v} \cdot \nabla) v + 2\Omega_x u &= -\frac{1}{\rho} \frac{\partial p}{\partial y}, \\
\frac{\partial w}{\partial t} + (\mathbf{v} \cdot \nabla) w + 2(\Omega_x v - \Omega_y u) &= -\frac{1}{\rho} \frac{\partial p}{\partial z} - g.
\end{align*}
\] (2.77)
where the rotation vector $\Omega = \Omega_x \mathbf{i} + \Omega_y \mathbf{j} + \Omega_z \mathbf{k}$ and $\Omega_x = 0$, $\Omega_y = \Omega \cos \theta_0$ and $\Omega_z = \Omega \sin \theta_0$. If we make the traditional approximation, and so ignore the components of $\Omega$ not in the direction of the local vertical, then

$$
\frac{Du}{Dt} - f_0 v = -\frac{1}{\rho} \frac{\partial p}{\partial x}, \quad (2.78a)
$$

$$
\frac{Dv}{Dt} + f_0 u = -\frac{1}{\rho} \frac{\partial p}{\partial y}, \quad (2.78b)
$$

$$
\frac{Dw}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial z} - \rho g. \quad (2.78c)
$$

where $f_0 = 2\Omega_z \sin \theta_0$. Defining the horizontal velocity vector $\mathbf{u} = (u, v, 0)$, the first two equations may also be written as

$$
\frac{Du}{Dt} + f_0 \times \mathbf{u} = -\frac{1}{\rho} \nabla_z p, \quad (2.79)
$$

where $\mathbf{u}/Dt = \partial \mathbf{u}/\partial t + \mathbf{v} \cdot \nabla \mathbf{u}$, $f_0 = 2\Omega \sin \theta_0 \mathbf{k} = f_0 \mathbf{k}$, and $\mathbf{k}$ is the direction perpendicular to the plane (it does not change its orientation with latitude). These equations are, evidently, exactly the same as the momentum equations in a system in which the rotation vector is aligned with the local vertical, as illustrated in the right panel in Fig. 2.4. They will describe flow on the surface of a rotating sphere to a good approximation provided the flow is of limited latitudinal extent so that the effects of sphericity are unimportant. This is known as the \textit{f-plane} approximation since the Coriolis parameter is a constant. We may in addition make the hydrostatic approximation, in which case (2.78c) becomes the familiar $\partial p/\partial z = -\rho g$.

### 2.3.2 The beta-plane approximation

The magnitude of the vertical component of rotation varies with latitude, and this has important dynamical consequences. We can approximate this effect by allowing the effective rotation vector to vary. Thus, noting that, for small variations in latitude,

$$
f = 2\Omega \sin \phi \approx 2\Omega \sin \phi_0 + 2\Omega (\phi - \phi_0) \cos \theta_0, \quad (2.80)
$$

then on the tangent plane we may mimic this by allowing the Coriolis parameter to vary as

$$
f = f_0 + \beta y, \quad (2.81)
$$

where $f_0 = 2\Omega \sin \theta_0$ and $\beta = \partial f/\partial y = (2\Omega \cos \theta_0)/a$. This important approximation is known as the \textit{beta-plane}, or \textit{β-plane}, approximation; it captures the the most important dynamical effects of sphericity, without the complicating geometric effects, which are not essential to describe many phenomena. The momentum equations (2.78a), (2.78b) and (2.78c) (or its hydrostatic counterpart) are unaltered, save that $f_0$ is replaced by $f_0 + \beta y$ to represent a varying Coriolis parameter. Thus, sphericity combined with rotation is dynamically equivalent to a differentially rotating system. For future reference, we write down the \textit{β-plane} horizontal momentum equations:

$$
\frac{Du}{Dt} + f \times \mathbf{u} = -\frac{1}{\rho} \nabla_z p, \quad (2.82)
$$
where \( f = (f_0 + \beta y) \hat{k} \). In component form this equation becomes

\[
\begin{align*}
\frac{D}{Dt} u - f v &= - \frac{1}{\rho} \frac{\partial p}{\partial x}, \\
\frac{D}{Dt} v + f u &= - \frac{1}{\rho} \frac{\partial p}{\partial y},
\end{align*}
\tag{2.83a,b}
\]

The mass conservation, thermodynamic and hydrostatic equations in the \( \beta \)-plane approximation are the same as the usual Cartesian, \( f \)-plane, forms of those equations.

### 2.4 EQUATIONS FOR A STRATIFIED OCEAN: THE BOUSSINESQ APPROXIMATION

The density variations in the ocean are quite small compared to the mean density, and we may exploit this to derive somewhat simpler but still quite accurate equations of motion. Let us first examine how much density does vary in the ocean.

#### 2.4.1 Variation of density in the ocean

The variations of density in the ocean are due to three effects: the compression of water by pressure (which we denote as \( \Delta p/\rho \)), the thermal expansion of water if its temperature changes \( \Delta T \), and the haline contraction if its salinity changes \( \Delta S \). How big are these? An appropriate equation of state to approximately evaluate these effects is the linear one

\[
\rho = \rho_0 \left[ 1 - \beta_T (T - T_0) + \beta_S (S - S_0) + \frac{p}{\rho_0 c_s^2} \right],
\tag{2.84}
\]

where \( \beta_T \approx 2 \times 10^{-4} \text{K}^{-1}, \beta_S \approx 10^{-3} \text{psu}^{-1} \) and \( c_s \approx 1500 \text{m s}^{-1} \) (see the table on page 37). The three effects are then:

**Pressure compressibility:** We have \( \Delta p/\rho \approx \Delta p/c_s^2 \approx \rho_0 gH/c_s^2 \) where \( H \) is the depth and the pressure change is quite accurately evaluated using the hydrostatic approximation. Thus,

\[
\frac{|\Delta p/\rho|}{\rho_0} \ll 1 \quad \text{if} \quad \frac{gH}{c_s^2} \ll 1,
\tag{2.85}
\]

or if \( H \ll c_s^2/g \). The quantity \( c_s^2/g \approx 200 \text{km} \) is the density scale height of the ocean. Thus, the pressure at the bottom of the ocean (say \( H = 10 \text{km} \) in the deep trenches), enormous as it is, is insufficient to compress the water enough to make a significant change in its density. Changes in density due to dynamical variations of pressure are small if the Mach number is small, and this is also the case.

**Thermal expansion:** We have \( \Delta T \rho \approx -\beta_T \rho_0 \Delta T \) and therefore

\[
\frac{|\Delta T \rho|}{\rho_0} \ll 1 \quad \text{if} \quad \beta_T \Delta T \ll 1.
\tag{2.86}
\]

For \( \Delta T = 20 \text{K} \), \( \beta_T \Delta T \approx 4 \times 10^{-3} \), and evidently we would require temperature differences of order \( \beta_T^{-1} \), or 5000 K to obtain order one variations in density.
Saline contraction: We have $\Delta_S \rho \approx \beta_S \rho_0 \Delta S$ and therefore

$$\frac{|\Delta_S \rho|}{\rho_0} \ll 1 \quad \text{if} \quad \beta_S \Delta S \ll 1. \tag{2.87}$$

As changes in salinity in the ocean rarely exceed 5 psu, for which $\beta_S \Delta S = 5 \times 10^{-3}$, the fractional change in the density of seawater is correspondingly very small.

Evidently, fractional density changes in the ocean are very small.

### 2.4.2 The Boussinesq equations

The Boussinesq equations are a set of equations that exploit the smallness of density variations in many liquids.⁴ To set notation we write

$$\rho = \rho_0 + \delta \rho(x, y, z, t) \tag{2.88a}$$

$$= \rho_0 + \hat{\rho}(z) + \rho'(x, y, z, t) \tag{2.88b}$$

$$= \tilde{\rho}(z) + \rho'(x, y, z, t), \tag{2.88c}$$

where $\rho_0$ is a constant and we assume that

$$|\hat{\rho}|, |\rho'|, |\delta \rho| \ll \rho_0. \tag{2.89}$$

We need not assume that $|\rho'| \ll |\hat{\rho}|$, but this is often the case in the ocean. To obtain the Boussinesq equations we will just use (2.88a), but (2.88c) will be useful for the anelastic equations considered later.

Associated with the reference density is a reference pressure that is defined to be in hydrostatic balance with it. That is,

$$p = p_0(z) + \delta p(x, y, z, t) \tag{2.90a}$$

$$= \tilde{p}(z) + p'(x, y, z, t), \tag{2.90b}$$

where $|\delta p| \ll p_0$, $|p'| \ll \tilde{p}$ and

$$\frac{dp_0}{dz} = -g \rho_0, \quad \frac{d\tilde{p}}{dz} = -g \tilde{p}. \tag{2.91a,b}$$

Note that $\nabla_z p = \nabla_z p' = \nabla_z \delta p$ and that $p_0 \approx \tilde{p}$ if $|\hat{\rho}| \ll \rho_0$.

**Momentum equations**

To obtain the Boussinesq equations we use $\rho = \rho_0 + \delta \rho$, and assume $\delta \rho/\rho_0$ is small. Without approximation, the momentum equation can be written as

$$(\rho_0 + \delta \rho) \left( \frac{Dv}{Dt} + 2\Omega \times v \right) = -\nabla \delta p - \frac{\partial p_0}{\partial z} k - g(\rho_0 + \delta \rho) k, \tag{2.92}$$

and using (2.91a) this becomes, again without approximation,

$$(\rho_0 + \delta \rho) \left( \frac{Dv}{Dt} + 2\Omega \times v \right) = -\nabla \delta p - g \delta \rho k. \tag{2.93}$$
If density variations are small this becomes

\[ \left( \frac{D}{Dt} + 2\Omega \times \mathbf{v} \right) = -\nabla \phi + bk, \]  

(2.94)

where \( \phi = \delta p/\rho_0 \) and \( b = -g\delta \rho/\rho_0 \) is the buoyancy. Note that we should not and do not neglect the term \( g\delta \rho \), for there is no reason to believe it to be small (\( \delta \rho \) may be small, but \( g \) is big). Eq. (2.94) is the momentum equation in the Boussinesq approximation, and it is common to say that the Boussinesq approximation ignores all variations of density of a fluid in the momentum equation, except when associated with the gravitational term.

For most large-scale motions in the ocean the deviation pressure and density fields are also approximately in hydrostatic balance, and in that case the vertical component of (2.94) becomes

\[ \frac{\partial \phi}{\partial z} = b. \]  

(2.95)

A condition for (2.95) to hold is that vertical accelerations are small compared to \( g\delta \rho/\rho_0 \), and not compared to the acceleration due to gravity itself. For more discussion of this point, see section 2.7.

**Mass Conservation**

The unapproximated mass conservation equation is

\[ \frac{D\delta \rho}{Dt} + (\rho_0 + \delta \rho) \nabla \cdot \mathbf{v} = 0. \]  

(2.96)

Provided that time scales advectively — that is to say that \( D/Dt \) scales in the same way as \( \mathbf{v} \cdot \nabla \) — then we may approximate this equation by

\[ \nabla \cdot \mathbf{v} = 0, \]  

(2.97)

which is the same as that for a constant density fluid. This absolutely does not allow one to go back and use (2.96) to say that \( D\delta \rho/Dt = 0; \) the evolution of density is given by the thermodynamic equation in conjunction with an equation of state, and this should not be confused with the mass conservation equation. Note also that in eliminating the time-derivative of density we eliminate the possibility of sound waves.

**Thermodynamic equation and equation of state**

The Boussinesq equations are closed by the addition of an equation of state, a thermodynamic equation and, as appropriate, a salinity equation. Neglecting salinity for the moment, a useful starting point is to write the thermodynamic equation, (1.116), as

\[ \frac{D\rho}{Dt} - \frac{1}{c_s^2} \frac{D p}{Dt} = \frac{\dot{Q}}{(\partial \eta/\partial \rho)_p} \approx -\dot{Q} \left( \frac{\rho_0 \beta_T}{c_p} \right) \]  

(2.98)

using \( (\partial \eta/\partial \rho)_p = (\partial \eta/\partial T)_p (\partial T/\partial \rho)_p \approx c_p/(T\rho_0 \beta_T). \)
Given the expansions (2.88a) and (2.90a) this can be written as

\[
\frac{D\delta \rho}{Dt} - \frac{1}{c_s^2} \frac{Dp_0}{Dt} = -\dot{Q} \left( \frac{\rho_0 \beta T}{c_p} \right),
\]

(2.99)
or, using (2.91a),

\[
\frac{D}{Dt} \left( \delta \rho + \frac{\rho_0 \beta T}{c_s^2} z \right) = -\dot{Q} \left( \frac{\rho_0 \beta T}{c_p} \right),
\]

(2.100)
as in (1.119). The severest approximation to this is to neglect the second term in brackets, and noting that \( b = -g\delta \rho/\rho_0 \) we obtain

\[
\frac{Db}{Dt} = \dot{b},
\]

(2.101)

where \( \dot{b} = g\beta T\dot{Q}/c_p \). The momentum equation (2.94), mass continuity equation (2.97) and thermodynamic equation (2.101) then form a closed set, called the simple Boussinesq equations.

A somewhat more accurate approach is to include the compressibility of the fluid that is due to the hydrostatic pressure. From (2.100), the potential density is given by \( \delta \rho_{pot} = \delta \rho + \rho_0 g z / c_s^2 \), and so the potential buoyancy, that is the buoyancy based on potential density, is given by

\[
b_\sigma \equiv -g \delta \rho_{pot}/\rho_0 = -g\rho_0 \left( \delta \rho + \frac{\rho_0 g z}{c_s^2} \right) = b - \frac{g z}{H_\rho},
\]

(2.102)

where \( H_\rho = c_s^2 / g \). The thermodynamic equation, (2.100), may then be written

\[
\frac{Db_\sigma}{Dt} = \dot{b}_\sigma,
\]

(2.103)

where \( \dot{b}_\sigma = \dot{b} \). Buoyancy itself is obtained from \( b_\sigma \) by the ‘equation of state’, \( b = b_\sigma + g z / H_\rho \).

In many applications we may need to use a still more accurate equation of state. In that case (and see section 1.5.5) we replace (2.101) by the thermodynamic equations

\[
\frac{D\theta}{Dt} = \dot{\theta}, \quad \frac{DS}{Dt} = \dot{S},
\]

(2.104a,b)

where \( \theta \) is the potential temperature and \( S \) is salinity, along with an equation of state. The equation of state has the general form \( b = b(\theta, S, p) \), but to be consistent with the level of approximation in the other Boussinesq equations we should replace \( p \) by the hydrostatic pressure calculated with the reference density, that is by \( -\rho_0 g z \), and the equation of state then takes the general form

\[
b = b(\theta, S, z).
\]

(2.105)

An example of (2.105) is (1.174), taken with the definition of buoyancy \( b = -g\delta \rho/\rho_0 \). The closed set of equations (2.94), (2.97), (2.104) and (2.105) are the
Summary of Boussinesq Equations

The simple Boussinesq equations are, for an inviscid fluid:

Momentum equations: \[ \frac{Dv}{Dt} + f \times v = -\nabla \phi + bk \] (B.1)

Mass conservation: \[ \nabla \cdot v = 0 \] (B.2)

Buoyancy equation: \[ \frac{Db}{Dt} = \dot{b} \] (B.3)

A more general form replaces the buoyancy equation by:

Thermodynamic equation: \[ \frac{D\theta}{Dt} = \dot{\theta} \] (B.4)

Salinity equation: \[ \frac{DS}{Dt} = \dot{S} \] (B.5)

Equation of state: \[ b = b(\theta, S, \phi) \] (B.6)

Energy conservation is only assured if \( b = b(\theta, S, z) \).

general Boussinesq equations. Using an accurate equation of state and the Boussinesq approximation is the procedure used in many comprehensive ocean general circulation models. The Boussinesq equations, which with the hydrostatic and traditional approximations are often considered to be the oceanic primitive equations, are summarized in the shaded box above.

Mean stratification and the buoyancy frequency

The processes that cause density to vary in the vertical often differ from those that cause it to vary in the horizontal. For this reason it is sometimes useful to write \( \rho = \rho_0 + \tilde{\rho}(z) + \rho'(x, y, z, t) \) and define \( \tilde{b}(z) \equiv -g\tilde{\rho}/\rho_0 \) and \( b' \equiv -g\rho'/\rho_0 \). Using the hydrostatic equation to evaluate pressure, the thermodynamic equation (2.98) becomes, to a good approximation,

\[ \frac{Db'}{Dt} + N^2 w = 0, \] (2.106)

where \( \frac{D}{Dt} \) remains a three-dimensional operator and

\[ N^2(z) = \left( \frac{d\tilde{b}}{dz} - \frac{g^2}{c_s^2} \right) = \frac{d\tilde{b}_\sigma}{dz}, \] (2.107)

where \( \tilde{b}_\sigma = \tilde{b} - gz/H \rho \). The quantity \( N^2 \) is a measure of the mean stratification of the fluid, and is equal to the vertical gradient of the mean potential buoyancy. \( N \) is known as the buoyancy frequency, something we return to in section 2.9. Equations (2.106) and (2.107) also hold in the simple Boussinesq equations, but with \( c_s^2 = \infty \).
2.4.3 Energetics of the Boussinesq system

In a uniform gravitational field but with no other forcing or dissipation, we write the simple Boussinesq equations as

\[
\frac{Dv}{Dt} + 2\Omega \times v = bk - \nabla \phi, \quad \nabla \cdot v = 0, \quad \frac{Db}{Dt} = 0. \tag{2.108a,b,c}
\]

From (2.108a) and (2.108b) the kinetic energy density evolution (c.f., section 1.9) is given by

\[
\frac{1}{2} \frac{Dv^2}{Dt} = bw - \nabla \cdot (\phi v), \tag{2.109}
\]

where the constant reference density \( \rho_0 \) is omitted. Let us now define the potential \( \Phi \) such that \( \nabla \Phi = -k \) (so \( \Phi = -z \)) and so

\[
\frac{D\Phi}{Dt} = \nabla \cdot (v\Phi) = -w. \tag{2.110}
\]

Using this and (2.108c) gives

\[
\frac{D}{Dt}(b\Phi) = -wb. \tag{2.111}
\]

Adding this to (2.109) and expanding the material derivative gives

\[
\frac{\partial}{\partial t} \left( \frac{1}{2}v^2 + b\Phi \right) + \nabla \cdot \left[ v \left( \frac{1}{2}v^2 + b\Phi + \phi \right) \right] = 0. \tag{2.112}
\]

This constitutes an energy equation for the Boussinesq system, and may be compared to (1.189). (See also problem 2.12.) The energy density (divided by \( \rho_0 \)) is just \( v^2/2 + b\Phi \). What does the second term represent? Its integral, multiplied by \( \rho_0 \), is the potential energy of the flow minus that of the basic state, or \( \int g(\rho - \rho_0)z \, dz \).

If there were a heating term on the right-hand side of (2.108c) this would directly provide a source of potential energy, rather than internal energy as in the compressible system. Because the fluid is incompressible, there is no conversion from kinetic and potential energy into internal energy.

* Energetics with a general equation of state

Now consider the energetics of the general Boussinesq equations. Suppose first that we allow the equation of state to be a function of pressure; the equations are motion then (2.108) except that (2.108c) is replaced by

\[
\frac{D\theta}{Dt} = 0, \quad \frac{DS}{Dt} = 0, \quad b = b(\theta, S, \phi). \tag{2.113a,b,c}
\]

A little algebraic experimentation will reveal that no energy conservation law of the form (2.112) generally exists for this system! The problem arises because, by requiring that the fluid be incompressible, we eliminate the proper conversion of internal energy to kinetic energy. However, if we use the consistent approximation \( b = b(\theta, S, z) \), the system does conserve an energy, as we now show.\(^5\)

Define the potential, \( \Pi \), by the integral of \( b \) at constant potential temperature and salinity

\[
\Pi(\theta, S, z) \equiv -\int b \, dz. \tag{2.114}
\]
Taking its material derivative gives
\[
\frac{D\Pi}{Dt} = \left( \frac{\partial \Pi}{\partial \theta} \right)_{S,z} \frac{D\theta}{Dt} + \left( \frac{\partial \Pi}{\partial S} \right)_{\theta,z} \frac{DS}{Dt} + \left( \frac{\partial \Pi}{\partial z} \right)_{\theta,S} \frac{Dz}{Dt} = -b w, \tag{2.115}
\]
using (2.113a,b). Combining (2.115) and (2.109) gives
\[
\frac{\partial}{\partial t} \left( \frac{1}{2} v^2 + \Pi \right) + \nabla \cdot \left[ v \left( \frac{1}{2} v^2 + \Pi + \phi \right) \right] = 0. \tag{2.116}
\]
Thus, energetic consistency is maintained with an arbitrary equation of state, provided the pressure is replaced by a function of \( z \). If \( b \) is not an explicit function of \( z \) in the equation of state, the conservation law is identical to (2.112).

### 2.5 EQUATIONS FOR A STRATIFIED ATMOSPHERE: THE ANELASTIC APPROXIMATION

#### 2.5.1 Preliminaries

In the atmosphere the density varies significantly, especially in the vertical. However deviations of both \( \rho \) and \( p \) from a statically balanced state are often quite small, and the relative vertical variation of potential temperature is also small. We can usefully exploit these observations to give a somewhat simplified set of equations, useful both for theoretical and numerical analysis because sound waves are eliminated by way of an ‘anelastic’ approximation. To begin we set
\[
\rho = \bar{\rho}(z) + \delta\rho(x,y,z,t), \quad p = \bar{p}(z) + \delta p(x,y,z,t), \tag{2.117a,b}
\]
where we assume that \( |\delta\rho| \ll |\bar{\rho}| \) and we define \( \bar{p} \) such that
\[
\frac{\partial \bar{p}}{\partial z} = -g \bar{\rho}(z). \tag{2.118}
\]
The notation is similar to that for the Boussinesq case except that, importantly, the density basic state is now a (given) function of vertical coordinate. As with the Boussinesq case, the idea is to ignore dynamic variations of density (i.e., of \( \delta\rho \)) except where associated with gravity. First recall a couple of ideal gas relationships involving potential temperature, \( \theta \), and entropy \( s \) (divided by \( c_p \), so \( s = \log \theta \)), namely
\[
s = \log \theta = \log T - \frac{R}{c_p} \log p = \frac{1}{y} \log p - \log \rho, \tag{2.119}
\]
where \( y = c_p/c_v \), implying
\[
\delta s = \frac{1}{y} \frac{\delta p}{p} - \frac{\delta \rho}{\rho} \approx \frac{1}{y} \frac{\delta p}{\bar{p}} - \frac{\delta \rho}{\bar{\rho}}. \tag{2.120}
\]
Further, if \( \tilde{s} \equiv y^{-1} \log \bar{p} - \log \bar{\rho} \) then
\[
\frac{d\tilde{s}}{dz} = \frac{1}{yp} \frac{d\bar{p}}{dz} - \frac{1}{p} \frac{d\bar{\rho}}{dz} = -\frac{g}{y} \frac{\bar{\rho}}{\bar{p}} - \frac{1}{\bar{p}} \frac{d\bar{\rho}}{dz}. \tag{2.121}
\]
In the atmosphere, the left-hand side is, typically, much smaller than either of the two terms on the right-hand side.
2.5.2 The Momentum equation

The exact inviscid horizontal momentum equation is

\[(\tilde{\rho} + \rho') \frac{D\mathbf{u}}{Dt} + f \times \mathbf{u} = -\nabla_z \delta p.\]  

(2.122)

Neglecting \(\rho'\) where it appears with \(\tilde{\rho}\) leads to

\[\frac{D\mathbf{u}}{Dt} + f \times \mathbf{u} = -\nabla_z \phi,\]  

(2.123)

where \(\phi = \delta p / \tilde{\rho}\), and this is similar to the corresponding equation in the Boussinesq approximation.

The vertical component of the inviscid momentum equation is, without approximation,

\[\tilde{\rho} + \delta \rho \frac{Dw}{Dt} = -\frac{\partial \tilde{p}}{\partial z} - \frac{\partial \delta p}{\partial z} - g\tilde{\rho} - g\delta \rho = -\frac{\partial \delta p}{\partial z} - g\delta \rho.\]  

(2.124)

using (2.117). Neglecting \(\delta \rho\) on the left-hand side we obtain

\[\frac{Dw}{Dt} = -\frac{1}{\tilde{\rho}} \frac{\partial \tilde{p}}{\partial z} - g \frac{\delta \rho}{\tilde{\rho}} = -\frac{\partial}{\partial z} \left( \frac{\delta p}{\tilde{\rho}} \right) - \frac{\delta p}{\tilde{\rho}^2} \frac{\partial \tilde{p}}{\partial z} - g \frac{\delta \rho}{\tilde{\rho}}.\]  

(2.125)

This is not a useful form for a gaseous atmosphere, since the variation of the mean density cannot be ignored. However, we may eliminate \(\delta \rho\) in favour of \(\delta s\) using (2.120) to give

\[\frac{Dw}{Dt} = g\delta s - \frac{\partial}{\partial z} \left( \frac{\delta p}{\tilde{\rho}} \right) - g \frac{\delta p}{\tilde{\rho}^2} \frac{\partial \tilde{p}}{\partial z},\]  

(2.126)

and using (2.121) gives

\[\frac{Dw}{Dt} = g\delta s - \frac{\partial}{\partial z} \left( \frac{\delta p}{\tilde{\rho}} \right) + \frac{d\tilde{\rho}}{dz} \frac{\delta p}{\tilde{\rho}}.\]  

(2.127)

What have these manipulations gained us? Two things:

(i) The gravitational term now involves \(\delta s\) rather than \(\delta \rho\) which enables a more direct connection with the thermodynamic equation.

(ii) The potential temperature scale height (\(~100\) km) in the atmosphere is much larger than the density scale height (\(~10\) km), and so the last term in (2.127) is small.

The second item thus suggests that we choose our reference state to be one of constant potential temperature (see also problem 2.17). The term \(d\tilde{s}/dz\) then vanishes and the vertical momentum equation becomes

\[\frac{Dw}{Dt} = g\delta s - \frac{\partial \phi}{\partial z},\]  

(2.128)

where \(\phi = \delta p / \tilde{\rho}\), \(\delta s = \delta \theta / \tilde{\theta}\) and \(\tilde{\theta} = \theta_0\), a constant. If we define a buoyancy by \(b_a = g\delta s = g\delta \theta / \tilde{\theta}\), then (2.123) and (2.128) have the same form as the Boussinesq momentum equations, but with a slightly different definitions of \(b\).
2.5.3 Mass conservation

Using (2.117a) the mass conservation equation may be written, without approximation, as

\[
\frac{\partial \delta \rho}{\partial t} + \nabla \cdot [(\tilde{\rho} + \delta \rho)v] = 0. \tag{2.129}
\]

We neglect \( \delta \rho \) where it appears with \( \tilde{\rho} \) in the divergence term. Further, the local time derivative will be small if time itself is scaled advectively (i.e., \( T \sim L/U \) and sound waves do not dominate), giving

\[
\nabla \cdot u + \frac{1}{\tilde{\rho}} \frac{\partial}{\partial z}(\tilde{\rho} w) = 0. \tag{2.130}
\]

It is here that the eponymous ‘anelastic approximation’ arises: the elastic compressibility of the fluid is neglected, and this serves to eliminate sound waves. For reference, in spherical coordinates the equation is

\[
\frac{1}{\cos \theta} \frac{\partial}{\partial \lambda} u + \frac{1}{\cos \theta} \frac{\partial}{\partial \theta} (v \cos \theta) + \frac{1}{\tilde{\rho}} \frac{\partial (w \tilde{\rho})}{\partial z} = 0. \tag{2.131}
\]

In an ideal gas, the choice of constant potential temperature determines how the reference density \( \tilde{\rho} \) varies with height. In some circumstances it is convenient to let \( \tilde{\rho} \) be a constant, \( \rho_0 \) (effectively choosing a different equation of state), in which case the anelastic equations become identical with the Boussinesq equations, albeit with the buoyancy interpreted in terms of potential temperature in the former and density in the latter. Because of their similarity, the Boussinesq and anelastic approximations are sometimes referred to as the strong and weak Boussinesq approximations, respectively.

2.5.4 Thermodynamic equation

The thermodynamic equation for an ideal gas may be written

\[
\frac{D \ln \theta}{Dt} = \frac{\dot{Q}}{Tc_p}. \tag{2.132}
\]

In the anelastic equations, \( \theta = \tilde{\theta} + \delta \theta \) where \( \tilde{\theta} \) is constant, and the thermodynamic equation is

\[
\frac{D \delta s}{Dt} = \frac{\tilde{\theta}}{Tc_p} \dot{Q}. \tag{2.133}
\]

Summarizing, the complete set of anelastic equations, with rotation but with no dissipation or diabatic terms, is

\[
\begin{align*}
\frac{Dv}{Dt} + 2 \Omega \times v &= kb_a - \nabla \phi \\
\frac{Db_a}{Dt} &= 0 \\
\nabla \cdot (\tilde{\rho}v) &= 0
\end{align*} \tag{2.134}
\]
where \( b_a = g \delta s = g \delta \theta/\tilde{\theta} \). The main difference between the anelastic and Boussinesq sets of equations is in the mass continuity equation, and when \( \tilde{\rho} = \rho_0 = \) constant the two equation sets are formally identical. However, whereas the Boussinesq approximation is a very good one for ocean dynamics, the anelastic approximation is much less so for large-scale atmosphere flow: the constancy of the reference potential temperature state is then not a particularly good approximation and so the deviations in density from its reference profile are not especially small, leading to inaccuracies in the momentum equation. Nevertheless, the anelastic equations have been used very productively in limited area ‘large-eddy-simulations’ where one does not wish to make the hydrostatic approximation but where sound waves are unimportant. The equations also provide a good jumping-off point for theoretical studies and the still simpler models that will be considered in the chapter.

### 2.5.5 Energetics of the anelastic equations

Conservation of energy follows in much the same way as for the Boussinesq equations, except that \( \tilde{\rho} \) enters. Take the dot product of (2.134a) with \( \tilde{\rho} \mathbf{v} \) to obtain

\[
\tilde{\rho} \frac{D}{Dt} \left( \frac{1}{2} \mathbf{v}^2 \right) = -\nabla \cdot (\phi \tilde{\rho} \mathbf{v}) + b_a \tilde{\rho} \tilde{w}.
\] (2.135)

Now, define a potential \( \Phi(z) \) such that \( \nabla \Phi = -k \), and so

\[
\tilde{\rho} \frac{D\Phi}{Dt} = -w \tilde{\rho}.
\] (2.136)

Combining this with the thermodynamic equation (2.134b) gives

\[
\tilde{\rho} \frac{D(b_a \Phi)}{Dt} = -w b_a \tilde{\rho}.
\] (2.137)

Adding this to (2.135) gives

\[
\tilde{\rho} \frac{D}{Dt} \left( \frac{1}{2} \mathbf{v}^2 + b_a \Phi \right) = -\nabla \cdot (\phi \tilde{\rho} \mathbf{v}),
\] (2.138)

or, expanding the material derivative,

\[
\frac{\partial}{\partial t} \left[ \tilde{\rho} \left( \frac{1}{2} \mathbf{v}^2 + b_a \Phi \right) \right] + \nabla \cdot \left[ \tilde{\rho} \mathbf{v} \left( \frac{1}{2} \mathbf{v}^2 + b_a \Phi + \phi \right) \right] = 0.
\] (2.139)

This equation has the form

\[
\frac{\partial E}{\partial t} + \nabla \cdot \left[ \mathbf{v} (E + \tilde{\rho} \phi) \right] = 0,
\] (2.140)

where \( E = \tilde{\rho}(\mathbf{v}^2/2 + b_a \Phi) \) is the energy density of the flow. This is a consistent energetic equation for the system, and when integrated over a closed domain the total energy is evidently conserved. The total energy density comprises the kinetic energy and a term \( \tilde{\rho} b_a \Phi \), which is analogous to the potential energy of a Boussinesq system. However, it is not exactly equal to that because \( b_a \) is the buoyancy based on potential temperature, not density; rather, the term combines contributions from both the internal energy and the potential energy.
2.6 CHANGING VERTICAL COORDINATE

Although using $z$ as a vertical coordinate is a natural choice given our Cartesian worldview, it is not the only option, nor is it always the most useful one. Any variable that has a one-to-one correspondence with $z$ in the vertical, so any variable that varies monotonically with $z$, could be used; pressure and, perhaps surprisingly, entropy, are common choices. In the atmosphere pressure almost always falls monotonically with height, and using it instead of $z$ provides a useful simplification of the mass conservation and geostrophic relations, as well as a more direct connection with observations, which are often taken at fixed values of pressure. (In the ocean pressure is almost the same as height, because density is almost constant.) Entropy seems an exotic vertical coordinate, but it is very useful in adiabatic flow, and we consider that in chapter 3.

2.6.1 Pressure coordinates

The primitive equations of motion for an ideal gas can be written,

$$\frac{Du}{Dt} + f \times u = -\frac{1}{\rho} \nabla p, \quad \frac{\partial p}{\partial z} = -\rho g,$$

(2.141a)

$$\frac{D\theta}{Dt} = 0, \quad \frac{D\rho}{Dt} + \rho \nabla \cdot v = 0,$$

(2.141b)

where $p = \rho RT$ and $\theta = T (p_R/p)^{R/c_p}$, and $p_R$ is the reference pressure. These are respectively the horizontal momentum, hydrostatic, thermodynamic and mass continuity equations. They can be put into a form similar to the Boussinesq equations by transforming from Cartesian \([x,y,z]\) to pressure coordinates \([x,y,p]\).

The analog of the vertical velocity is $\omega \equiv \partial p/\partial t$, and the advective derivative itself is given by

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u \cdot \nabla p + \omega \frac{\partial}{\partial p}.$$  

(2.142)

The horizontal and time derivatives are taken at constant pressure. However, $x$ and $y$ are still purely horizontal coordinates, and $u = ui + vj$ is still a strictly horizontal velocity, perpendicular to the vertical ($z$) axis. The operator $D/Dt$ is of course the same in pressure or height coordinates because is simply the total derivative of some property of a fluid parcel. However, the individual terms comprising it in general differ between height and pressure coordinates.

To obtain an expression for the pressure force, first consider a general vertical coordinate, $\xi$. Then the chain rule gives

$$\left( \frac{\partial}{\partial x} \right)_{\xi} = \left( \frac{\partial}{\partial x} \right)_z + \left( \frac{\partial z}{\partial x} \right)_\xi \frac{\partial}{\partial z}.$$  

(2.143)

Now let $\xi = p$ and apply the relationship to $p$ itself to give

$$0 = \left( \frac{\partial p}{\partial x} \right)_z + \left( \frac{\partial z}{\partial x} \right)_p \frac{\partial p}{\partial z},$$  

(2.144)

which, using the hydrostatic relationship, gives

$$\left( \frac{\partial p}{\partial x} \right)_z = \rho \left( \frac{\partial \phi}{\partial x} \right)_p,$$  

(2.145)
where $\Phi = \rho z$ is the geopotential. Thus, the horizontal pressure force in the momentum equations is

$$\frac{1}{\rho} \nabla_z p = \nabla_p \Phi,$$  

(2.146)

where the subscripts on the gradient operator indicate that the horizontal derivatives are taken at constant $z$ or constant $p$. Also, from (2.141a), the hydrostatic equation is just

$$\frac{\partial \Phi}{\partial p} = -\alpha.$$  

(2.147)

### Mass Continuity

The mass conservation equation simplifies attractively in pressure coordinates, if the hydrostatic approximation is used. Recall that the mass conservation equation can be derived from the Lagrangian form

$$\frac{D}{Dt} \rho \delta V = 0,$$  

(2.148)

where $\delta V = \delta x \delta y \delta z$ is a volume element. But by the hydrostatic relationship $\rho \delta z = (1/g) \delta p$ and thus

$$\frac{D}{Dt} (\delta x \delta y \delta p) = 0.$$  

(2.149)

This is completely analogous to the expression for the Lagrangian conservation of volume in an incompressible fluid, (1.15). Thus, without further ado, we write the mass conservation in pressure coordinates as

$$\nabla_p \cdot \mathbf{u} + \frac{\partial \omega}{\partial p} = 0,$$  

(2.150)

where the horizontal derivative is taken at constant pressure. The primitive equations in pressure coordinates, equivalent to (2.141) in height coordinates, are thus:

| $\frac{D}{Dt} \mathbf{u}$ | $\mathbf{f} \times \mathbf{u} = -\nabla_p \Phi$, | $\frac{\partial \Phi}{\partial p} = -\alpha$, |
| $\frac{D}{Dt} \theta$ | $= 0$, | $\nabla_p \cdot \mathbf{u} + \frac{\partial \omega}{\partial p} = 0$, |

(2.151)

where $D/Dt$ is given by (2.142). The equation set is completed with the addition of the ideal gas equation and the definition of potential temperature. These equations are isomorphic to the hydrostatic general Boussinesq equations (see shaded box on page 72) with $z \rightarrow -p$, $w \rightarrow -\omega$, $\Phi \rightarrow \Phi$, $b \rightarrow \alpha$, and an equation of state $b = b(\theta, z) \rightarrow \alpha = \alpha(\theta, p)$. In an ideal gas, for example, $\alpha = -(\theta R/p_R)(p_R/p)^{1/\gamma}$.

The main practical difficulty with the pressure-coordinate equations is the lower boundary condition. Using

$$w = \frac{Dz}{Dt} = \frac{\partial z}{\partial t} + \mathbf{u} \cdot \nabla_p z + \omega \frac{\partial z}{\partial p},$$  

(2.152)
and (2.147), the boundary condition of $w = 0$ at $z = z_s$ becomes

$$\frac{\partial \Phi}{\partial t} + \mathbf{u} \cdot \nabla_p \Phi - \alpha \omega = 0$$

(2.153)

at $p(x, y, z_s, t)$. In theoretical studies, it is common to assume that the lower boundary is in fact a constant pressure surface and simply assume that $\omega = 0$, or sometimes the condition $\omega = -\alpha^{-1} \frac{\partial \Phi}{\partial t}$ is used. For realistic studies (with general circulation models, say) the fact that the level $z = 0$ is not a coordinate surface must be properly accounted for. For this reason, and especially if the lower boundary is uneven because of the presence of topography, so-called sigma coordinates are sometimes used, in which the vertical coordinate is chosen so that the lower boundary is a coordinate surface. Sigma coordinates may use height itself as a measure of displacement (typical in oceanic applications) or use pressure (typical in atmospheric applications). In the latter case the vertical coordinate is $\sigma = \frac{p}{p_s}$, where $p_s(x, y, t)$ is the surface pressure. The difficulty of applying (2.153) is replaced by a prognostic equation for the surface pressure, which is derived from the mass conservation equation (problem 2.21).

2.6.2 Log-pressure coordinates

A variant of pressure coordinates is log-pressure coordinates, in which the vertical coordinate is $Z = -H \ln \left( \frac{p}{p_R} \right)$ where $p_R$ is a reference pressure (say 1000 mb) and $H$ a constant (for example the scale height $RT_s / g$) so that $Z$ has units of length. (Capital letters are conventionally used for some variables in log-pressure coordinates, and these are not to be confused with scaling parameters.) The 'vertical velocity' for the system is now

$$W \equiv \frac{DZ}{Dt},$$

(2.154)

and the advective derivative is now

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + \mathbf{u} \cdot \nabla_p + W \frac{\partial}{\partial Z}. $$

(2.155)

It is straightforward to show (problem 2.22) that the primitive equations of motion in these coordinates are:

$$\frac{Du}{Dt} + f \times \mathbf{u} = -\nabla_Z \Phi, \quad \frac{\partial \Phi}{\partial Z} = \frac{RT}{H},$$

(2.156a)

$$\frac{D\theta}{Dt} = 0, \quad \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial W}{\partial Z} - \frac{W}{H} = 0. $$

(2.156b)

The last equation may be written $\nabla_Z \cdot \mathbf{u} + \rho_R^{-1} \frac{\partial}{\partial z} (\rho_R W) / \partial z = 0$, where $\rho_R = \exp(-z/H)$, so giving a form similar to the mass conservation equation in the anelastic equations.

2.7 HYDROSTATIC BALANCE

In this section we consider one of the most fundamental balances in geophysical fluid dynamics — hydrostatic balance, and in the next section we consider another fundamental balance, geostrophic balance. Neither hydrostasy (the state of
2.7 Hydrostatic Balance

geostrophic balance) nor geostrophy (the state of geostrophic balance) are exactly realized in the atmosphere or ocean, but their approximate satisfaction has profound consequences on the behaviour of the atmosphere and ocean. We first encountered hydrostatic balance in section 1.3.4; we now look in more detail at the conditions required for it to hold.

2.7.1 Preliminaries

Consider the relative sizes of terms in (2.77c):

\[
\frac{W}{T} + \frac{UW}{L} + \frac{W^2}{H} + \Omega U \sim \left| \frac{1}{\rho} \frac{\partial p}{\partial z} \right| + g. \tag{2.157}
\]

For most large-scale motion in the atmosphere and ocean the terms on the right-hand side are orders of magnitude larger than those on the left, and therefore must be approximately equal. Explicitly, suppose \( W \sim 1 \text{ cm s}^{-1}, L \sim 10^5 \text{ m}, H \sim 10^3 \text{ m}, U \sim 10 \text{ m s}^{-1}, T = L/U \). Then by substituting into (2.157) it seems that the pressure term is the only one which could balance the gravitational term, and we are led to approximate (2.77c) by,

\[
\frac{\partial p}{\partial z} = -\rho g. \tag{2.158}
\]

This equation, which is a vertical momentum equation, is known as hydrostatic balance.

However, (2.158) is not always a useful equation! Let us suppose that the density is a constant, \( \rho_0 \). We can then write the pressure as

\[
p(x, y, z, t) = p_0(z) + p'(x, y, z, t), \tag{2.159}
\]

where

\[
\frac{\partial p_0}{\partial z} = -\rho_0 g. \tag{2.160}
\]

That is, \( p_0 \) and \( \rho_0 \) are in hydrostatic balance. The inviscid vertical momentum equation becomes, without approximation,

\[
\frac{Dw}{Dt} = -\frac{1}{\rho_0} \frac{\partial p'}{\partial z}. \tag{2.161}
\]

Thus, for constant density fluids, the gravitational term has no dynamical effect; there is no buoyancy force, and the pressure term in the horizontal momentum equations can be replaced by \( p' \). Hydrostatic balance, and in particular (2.160), is certainly not an appropriate vertical momentum equation in this case. If the fluid is stratified, we should therefore subtract off the hydrostatic pressure associated with the mean density before we can determine whether hydrostasy is a useful dynamical approximation, accurate enough to determine the horizontal pressure gradients. This is automatic in the Boussinesq equations, where the vertical momentum equation is

\[
\frac{Dw}{Dt} = -\frac{\partial \phi}{\partial z} + b, \tag{2.162}
\]

and the hydrostatic balance of the basic state is already subtracted out. In the more general equation,

\[
\frac{Dw}{Dt} = -\frac{1}{\rho} \frac{\partial p}{\partial z} - g, \tag{2.163}
\]
we need to compare the advective term on the left-hand side with the pressure variations arising from horizontal flow in order to determine whether hydrostasy is an appropriate vertical momentum equation. Nevertheless, if we simply need to determine the pressure for use in an equation of state then we simply need to compare the sizes of the dynamical terms in (2.77c) with \( g \) itself in order to determine whether a hydrostatic approximation will suffice.

### 2.7.2 Scaling and the aspect ratio

In a Boussinesq fluid we write the horizontal and vertical momentum equations as

\[
\frac{Du}{Dt} - f \times u = -\nabla \phi, \quad \frac{Dw}{Dt} = -\frac{\partial \phi}{\partial z} - b. \tag{2.164a,b}
\]

With \( f = 0 \), (2.164a) implies the scaling

\[
\phi \sim U^2. \tag{2.165}
\]

If we use mass conservation, \( \nabla_z \cdot u + \partial w/\partial z = 0 \), to scale vertical velocity then

\[
w \sim W = \frac{H}{L}U = \alpha U, \tag{2.166}
\]

where \( \alpha \equiv H/L \) is the aspect ratio. The advective terms in the vertical momentum equation all scale as

\[
\frac{Dw}{Dt} \sim \frac{UW}{L} = \frac{U^2H}{L^2}. \tag{2.167}
\]

Using (2.165) and (2.167) the ratio of the advective term to the pressure gradient term in the vertical momentum equations then scales as

\[
\frac{|Dw/Dt|}{|\partial \phi/\partial z|} \sim \frac{U^2H/L^2}{U^2/H} \sim \left(\frac{H}{L}\right)^2. \tag{2.168}
\]

Thus, the condition for hydrostasy, that \( |Dw/Dt|/|\partial \phi/\partial z| \ll 1 \), is:

\[
\alpha^2 = \left(\frac{H}{L}\right)^2 \ll 1. \tag{2.169}
\]

The advective term in the vertical momentum may then be neglected. Thus, hydrostasy is a small aspect ratio approximation.

We can obtain the same result more formally by nondimensionalizing the momentum equations. Using uppercase symbols to denote scaling values we write

\[
(x, y) = L(\hat{x}, \hat{y}), \quad z = H\hat{z}, \quad u = U\hat{u}, \quad w = W\hat{w} = \frac{HU}{L}\hat{w}, \tag{2.170}
\]

\[
t = T\hat{t}, \quad \phi = \Phi\hat{\phi} = U^2\hat{\phi}, \quad b = B\hat{b} = \frac{U^2}{H}\hat{b},
\]

where the hatted variables are nondimensional and the scaling for \( w \) is suggested
by the mass conservation equation, \( \nabla_z \cdot \mathbf{u} + \partial w/\partial z = 0 \). Substituting (2.170) into (2.164) (with \( f = 0 \)) gives us the nondimensional equations

\[
\frac{D\hat{u}}{Dt} = -\nabla_z \hat{\phi}, \quad \alpha^2 \frac{D\hat{w}}{Dt} = -\frac{\partial \hat{\phi}}{\partial \hat{z}} - \hat{b},
\]

(2.171a,b)

where \( D/Dt = \partial/\partial t + \hat{u} \partial/\partial \hat{x} + \hat{v} \partial/\partial \hat{y} + \hat{w} \partial/\partial \hat{z} \) and we use the convention that when \( \nabla \) operates on nondimensional quantities the operator itself is nondimensional. From (2.171b) it is clear that hydrostatic balance pertains when \( \alpha^2 \ll 1 \).

### 2.7.3 Effects of stratification on hydrostatic balance

To include the effects of stratification we need to involve the thermodynamic equation, so let us first write down the complete set of non-rotating dimensional equations:

\[
\frac{Du}{Dt} = -\nabla_z \phi, \quad \frac{Dw}{Dt} = -\frac{\partial \phi}{\partial z} + b',
\]

(2.172a,b)

\[
\frac{Db'}{Dt} + wN^2 = 0, \quad \nabla \cdot \mathbf{u} = 0.
\]

(2.173a,b)

We have written, without approximation, \( b = b'(x,y,z,t) + \tilde{b}(z) \), with \( N^2 = \partial \tilde{b}/\partial z \); this separation is useful because the horizontal and vertical buoyancy variations may scale in different ways, and often \( N^2 \) may be regarded as given. (We also redefine \( \phi \) by subtracting off a static component in hydrostatic balance with \( \tilde{b} \).) We nondimensionalize (2.173) by first writing

\[
(x,y) = L(\hat{x}, \hat{y}), \quad z = H\hat{z}, \quad \mathbf{u} = U\hat{u}, \quad w = W\hat{w} = \epsilon \frac{HU}{L} \hat{w},
\]

(2.174)

\[
t = \hat{t} = \frac{L}{U} \hat{t}, \quad \phi = U^2 \hat{\phi}, \quad b' = \Delta b \hat{b} = \frac{U^2}{H} \hat{b}', \quad N^2 = \frac{N^2}{\hat{N}^2} \hat{N}^2,
\]

where \( \epsilon \) is, for the moment, undetermined, \( \hat{N} \) is a representative, constant, value of the buoyancy frequency and \( \Delta b \) scales only the horizontal buoyancy variations. Substituting (2.174) into (2.172) and (2.173) gives

\[
\frac{D\hat{u}}{Dt} = -\nabla_z \hat{\phi}, \quad \epsilon \alpha^2 \frac{D\hat{w}}{Dt} = -\frac{\partial \hat{\phi}}{\partial \hat{z}} + \hat{b'},
\]

(2.175a,b)

\[
\frac{U^2}{N^2 H^2} \frac{D\hat{b}'}{Dt} + \epsilon \hat{w} \hat{N}^2 = 0, \quad \nabla \cdot \hat{\mathbf{u}} + \epsilon \frac{\partial \hat{w}}{\partial \hat{z}} = 0.
\]

(2.176a,b)

where now \( D/D\hat{t} = \partial/\partial \hat{t} + \hat{u} \cdot \nabla \hat{z} + \epsilon \partial/\partial \hat{z} \). To obtain a nontrivial balance in (2.176a) we choose \( \epsilon = U^2/(N^2 H^2) \equiv Fr^2 \), where \( Fr \) is the Froude number, a measure of the stratification of the flow. The vertical velocity then scales as

\[
W = \frac{Fr \ U \ H}{L}
\]

(2.177)

and if the flow is highly stratified the vertical velocity will be even smaller than a pure aspect ratio scaling might suggest. (There must, therefore, be some cancellation in horizontal divergence in the mass continuity equation; that is, \( |\nabla_z \cdot \mathbf{u}| \ll \)
With this choice of $\epsilon$ the nondimensional Boussinesq equations may be written:

$$
\frac{Du}{Dt} = -\nabla z \hat{\phi}, \quad Fr^2 \alpha^2 \frac{D\hat{w}}{Dt} = -\frac{\partial \hat{\phi}}{\partial \hat{z}} + \hat{b}' \quad (2.178a,b)
$$

$$
\frac{D\hat{b}'}{Dt} + \hat{w} N^2 = 0, \quad \nabla \cdot \hat{u} + Fr^2 \frac{\partial \hat{w}}{\partial \hat{z}} = 0. \quad (2.179a,b)
$$

The nondimensional parameters in the system are the aspect ratio and the Froude number (in addition to $N$, but by construction this is just an order one function of $z$). From (2.178b) condition for hydrostatic balance to hold is evidently that

$$Fr^2 \alpha^2 \ll 1 \quad (2.180),$$

so generalizing the aspect ratio condition (2.169) to a stratified fluid. Because $Fr$ is a measure of stratification, (2.180) formalizes our intuitive expectation that the more stratified a fluid the more vertical motion is suppressed and therefore the more likely hydrostatic balance is to hold. Note also that (2.180) is equivalent to $U^2/(L^2 N^2) \ll 1$.

Suppose we solve the hydrostatic equations; that is, we omit the advective derivative in the vertical momentum equation, and by numerical integration we obtain $u, w$ and $b$. This flow is the solution of the nonhydrostatic equations in the small aspect ratio limit. The solution never violates the scaling assumptions, even if $w$ seems large, because we can always rescale the variables in order that condition (2.180) is satisfied.

Why bother with any of this scaling? Why not just say that hydrostatic balance holds when $|Dw/Dt| \ll |\partial \phi/\partial z|$? One reason is that we don’t have a good idea of the value of $w$ from direct measurements, and it may change significantly in different oceanic and atmospheric parameter regimes. On the other hand the Froude number and the aspect ratio are familiar nondimensional parameters with a wide applicability in other contexts, and which we can control in a laboratory setting or estimate in the ocean or atmosphere. Still, in scaling theory it is common that ascertaining which parameters are to be regarded as given and which should be derived is a choice, rather than being set a priori.

### 2.7.4 Hydrostasy in the ocean and atmosphere

Is the hydrostatic approximation in fact a good one in the ocean and atmosphere?

**In the ocean**

For the large scale ocean circulation, let $N \sim 10^{-2} s^{-1}, U \sim 0.1 m s^{-1}$ and $H \sim 1 km$. Then $Fr = U/(NH) \sim 10^{-2} \ll 1$. Thus, $Fr^2 \alpha^2 \ll 1$ even for unit aspect-ratio motion. In fact, for larger scale flow the aspect ratio is also small; for basin-scale flow $L \sim 10^6 m$ and $Fr^2 \alpha^2 \sim 0.01^2 \times 0.001^2 = 10^{-10}$ and hydrostatic balance is an extremely good approximation.

For intense convection, for example in the Labrador Sea, the hydrostatic approximation may be less appropriate, because the intense descending plumes may have an aspect ratio $(H/L)$ of one or greater and the stratification is very weak. The hydrostatic condition then often becomes the requirement that the Froude number
is small. Representative orders of magnitude are $U \sim W \sim 0.1 \text{ ms}^{-1}$, $H \sim 1 \text{ km}$ and $N \sim 10^{-3} \text{ s}^{-1} - 10^{-4} \text{ s}^{-1}$. For these values $Fr$ ranges between 0.1 and 1, and at the upper end of this range hydrostatic balance is violated.

**In the atmosphere**

Over much of the troposphere $N \sim 10^{-2} \text{ s}^{-1}$ so that with $U = 10 \text{ m s}^{-1}$ and $H = 1 \text{ km}$ we find $Fr \sim 1$. Hydrostacy is then maintained because the aspect ratio $H/L$ is much less than unity. For larger scale synoptic activity a larger vertical scale is appropriate, and with $H = 10 \text{ km}$ both the Froude number and the aspect ratio are much smaller than one; indeed with $L = 1000 \text{ km}$ we find $Fr^2 \alpha^2 \sim 0.1^2 \times 0.1^2 = 10^{-4}$ and the flow is hydrostatic to a very good approximation indeed. However, for smaller scale atmospheric motion associated with fronts and, especially, convection, there can be little expectation that hydrostatic balance will be a good approximation.

### 2.8 GEOSTROPHIC AND THERMAL WIND BALANCE

We now consider the dominant dynamical balance in the horizontal components of the momentum equation. In the horizontal plane (meaning along geopotential surfaces) we find that the Coriolis term is much larger than the advective terms and the dominant balance is between it and the horizontal pressure force. This balance is called *geostrophic balance*, and it occurs when the Rossby number is small, as we now investigate.

#### 2.8.1 The Rossby Number

The *Rossby number* characterizes the importance of rotation in a fluid. It is, essentially, the ratio of the magnitude of the relative acceleration to the Coriolis acceleration, and it is of fundamental importance in geophysical fluid dynamics. It arises from a simple scaling of horizontal momentum equation, namely

\[
\begin{align*}
\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{u} + f \times \mathbf{u} &= -\frac{1}{\rho} \nabla \cdot \mathbf{p}, \\
U^2/L & \quad fU
\end{align*}
\]

where $U$ is the approximate magnitude of the horizontal velocity and $L$ is a typical lengthscale over which that velocity varies. (We assume that $W/H \ll U/L$, so that vertical advection does not dominate the advection.) The ratio of the sizes of the advective and Coriolis terms is defined to be the Rossby number,

\[
\text{Ro} = \frac{U}{fL}.
\]

If the Rossby number is small then rotation effects are important, and as the values in table 2.1 indicate this is the case for large-scale flow in both ocean and atmosphere.

Another intuitive way to think about the Rossby number is in terms of timescales. The Rossby number based on a timescale is

\[
\text{Ro}_T = \frac{1}{fT},
\]
Table 2.1 Scales of large-scale flow in atmosphere and ocean. The choices given are representative of large-scale eddying motion in both systems.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Scaling Symbol</th>
<th>Meaning</th>
<th>Atmos. value</th>
<th>Ocean value</th>
</tr>
</thead>
<tbody>
<tr>
<td>((x,y))</td>
<td>L (x,y)</td>
<td>Horizontal length</td>
<td>10⁶ m</td>
<td>10⁵ m</td>
</tr>
<tr>
<td>(t)</td>
<td>(T) (T)</td>
<td>Timescale</td>
<td>1 day (10⁵ s)</td>
<td>10 days (10⁶ s)</td>
</tr>
<tr>
<td>((u,v))</td>
<td>(U) (u,v)</td>
<td>Horizontal velocity</td>
<td>10 m s⁻¹</td>
<td>0.1 m s⁻¹</td>
</tr>
<tr>
<td>(Ro)</td>
<td>(U/\Omega L)</td>
<td>Rossby number, (U/\Omega L)</td>
<td>0.1</td>
<td>0.01</td>
</tr>
</tbody>
</table>

where \(T\) is a timescale associated with the dynamics at hand. If the timescale is an advective one, meaning that \(T \sim L/U\), then this definition is equivalent to (2.182). Now, \(f = 2\Omega \sin \vartheta\), where \(\Omega\) is the angular velocity of the rotating frame and equal to \(2\pi \sin \vartheta / T_p\) where \(T_p\) is the period of rotation (24 hours). Thus,

\[
\text{Ro}_T = \frac{T_p}{4\pi \sin \vartheta} = \frac{T_i}{T},
\]  

(2.184)

where \(T_i = 1/f\) is the ‘inertial timescale’, about three hours in midlatitudes. Thus, for phenomena with timescales much longer than this, such as the motion of the Gulf Stream or a mid-latitude atmospheric weather system, the effects of the earth’s rotation can be expected to be important, whereas a short-lived phenomena, such as a cumulus cloud or tornado, may be oblivious to such rotation. The expressions (2.182) and (2.183) of course, just approximate measures of the importance of rotation.

2.8.2 Geostrophic Balance

If the Rossby number is sufficiently small in (2.181a) then the rotation term will dominate the nonlinear advection term, and if the time period of the motion scales advectively then the rotation term also dominates the local time derivative. The only term which can then balance the rotation terms is the pressure term, and therefore we must have

\[
f \times u \approx -\frac{1}{\rho} \nabla_z p,
\]  

(2.185)

or, in Cartesian component form

\[
f u \approx -\frac{1}{\rho} \frac{\partial p}{\partial y}, \quad f v \approx \frac{1}{\rho} \frac{\partial p}{\partial x}.
\]  

(2.186)

This balance is known as \textit{geostrophic balance}, and its consequences are profound, giving geophysical fluid dynamics a special place in the broader field of fluid dynamics. We define the geostrophic velocity by

\[
f u_\theta = -\frac{1}{\rho} \frac{\partial p}{\partial y}, \quad f v_\theta = \frac{1}{\rho} \frac{\partial p}{\partial x},
\]  

(2.187)
and for low Rossby number flow \( u \approx u_\theta \) and \( v \approx v_\theta \). In spherical coordinates the geostrophic velocity is

\[
\begin{align*}
      f u_\theta &= -\frac{1}{\rho a} \frac{\partial p}{\partial \vartheta}, \\
      f v_\theta &= \frac{1}{a \rho \cos \vartheta} \frac{\partial p}{\partial \lambda},
\end{align*}
\]

where \( f = 2\Omega \sin \vartheta \). Geostrophic balance has a number of immediate ramifications:

- Geostrophic flow is parallel to lines of constant pressure (isobars). If \( f > 0 \) the flow is anti-clockwise round a region of low pressure and clockwise around a region of high pressure (see Fig. 2.5).
- If the Coriolis force is constant and if the density does not vary in the horizontal the geostrophic flow is horizontally non-divergent and

\[
\nabla_z \cdot \mathbf{u}_\theta = \frac{\partial u_\theta}{\partial x} + \frac{\partial v_\theta}{\partial y} = 0.
\]

We may define the geostrophic streamfunction, \( \psi \), by

\[
\psi \equiv \frac{p}{f_0 \rho_0},
\]

whence

\[
\begin{align*}
      u_\theta &= -\frac{\partial \psi}{\partial y}, \\
      v_\theta &= \frac{\partial \psi}{\partial x}.
\end{align*}
\]

The vertical component of vorticity, \( \zeta \), is then given by

\[
\zeta = \mathbf{k} \cdot \nabla \times \mathbf{v} = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \nabla^2 \psi.
\]
If the Coriolis parameter is not constant, then cross-differentiating (2.187) gives, for constant density geostrophic flow,

\[ v \frac{\partial f}{\partial y} + f \nabla_z \cdot u_g = 0, \]  
(2.193)

which implies, using mass continuity,

\[ \beta v_g = f \frac{\partial w}{\partial z}. \]  
(2.194)

where \( \beta \equiv \frac{\partial f}{\partial y} = \frac{2 \Omega}{a} \cos \theta \). This geostrophic vorticity balance is sometimes known as Sverdrup balance, although that expression is better restricted to the case when the vertical velocity results from external agents, and specifically a wind stress, as considered in chapter 14.

### 2.8.3 Taylor-Proudman effect

If \( \beta = 0 \), then (2.194) implies that the vertical velocity is not a function of height. In fact, in that case none of the components of velocity vary with height if density is also constant. To show this, in the limit of zero Rossby number we first write the three-dimensional momentum equation as

\[ f_0 \times \mathbf{v} = -\nabla \phi - \nabla \chi, \]  
(2.195)

where \( f_0 = 2 \Omega = 2 \Omega k \), \( \phi = p/\rho_0 \), and \( \nabla \chi \) represents other potential forces. If \( \chi = gz \) then the vertical component of this equation represents hydrostatic balance, and the horizontal components represent hydrostatic balance. On taking the curl of this equation, the terms on the right-hand side vanish and the left-hand side becomes

\[ (f_0 \cdot \nabla)\mathbf{v} - f_0 \nabla \cdot \mathbf{v} - (\mathbf{v} \cdot \nabla) f_0 + \mathbf{v} \nabla \cdot f_0 = 0. \]  
(2.196)

But \( \nabla \cdot \mathbf{v} = 0 \) by mass conservation, and because \( f_0 \) is constant both \( \nabla \cdot f_0 \) and \( (\mathbf{v} \cdot \nabla)f_0 \) vanish. Thus

\[ (f_0 \cdot \nabla)\mathbf{v} = 0, \]  
(2.197)

which, since \( f_0 = f_0 k \), implies \( f_0 \partial \mathbf{v} / \partial z = 0 \), and in particular we have

\[ \frac{\partial u}{\partial z} = 0, \quad \frac{\partial v}{\partial z} = 0, \quad \frac{\partial w}{\partial z} = 0. \]  
(2.198)

A different presentation of this argument proceeds as follows. If the flow is exactly in geostrophic and hydrostatic balance then

\[ v = \frac{1}{f_0} \frac{\partial \phi}{\partial x}, \quad u = -\frac{1}{f_0} \frac{\partial \phi}{\partial y}, \quad \frac{\partial \phi}{\partial z} = -g. \]  
(2.199a,b,c)

Differentiating (2.199a,b) with respect to \( z \), and using (2.199c) yields

\[ \frac{\partial v}{\partial z} = -\frac{1}{f_0} \frac{\partial g}{\partial x} = 0, \quad \frac{\partial u}{\partial z} = \frac{1}{f_0} \frac{\partial g}{\partial y} = 0. \]  
(2.200)
2.8 Geostrophic and Thermal Wind Balance

Noting that the geostrophic velocities are horizontally non-divergent \( \nabla_z \cdot \mathbf{u} = 0 \), and using mass continuity then gives \( \partial w / \partial z = 0 \), as before.

If there is a solid horizontal boundary anywhere in the fluid, for example at the surface, then \( w = 0 \) at that surface and thus \( w = 0 \) everywhere. Hence the motion occurs in planes that lie perpendicular to the axis of rotation, and the flow is effectively two-dimensional. This result is known as the Taylor-Proudman effect, namely that for constant density flow in geostrophic and hydrostatic balance the vertical derivatives of the horizontal and the vertical velocities are zero. At zero Rossby number, if the vertical velocity is zero somewhere in the flow, it is zero everywhere in that vertical column; furthermore, the horizontal flow has no vertical shear, and the fluid moves like a slab. The effects of rotation have provided a stiffening of the fluid in the vertical.

In neither the atmosphere nor the ocean do we observe precisely such vertically coherent flow, mainly because of the effects of stratification. However, it is typical of geophysical fluid dynamics that the assumptions underlying a derivation are not fully satisfied, yet there are manifestations of it in real flow. Thus, one might have naively expected, because \( \partial w / \partial z = - \nabla_z \cdot \mathbf{u} \), that the scales of the various variables would be related by \( W/H \sim U/L \). However, if the flow is rapidly rotating we expect that the horizontal flow will be in near geostrophic balance and therefore nearly divergence free; thus \( \nabla_z \cdot \mathbf{u} \ll U/L \), and \( W \ll HU/L \).

2.8.4 Thermal wind balance

Thermal wind balance arises by combining the geostrophic and hydrostatic approximations, and this is most easily done in the context of the anelastic (or Boussinesq) equations, or in pressure coordinates. For the anelastic equations, geostrophic balance may be written

\[
-f v_g = -\frac{\partial \phi}{\partial x} = -\frac{1}{a \cos \vartheta} \frac{\partial \phi}{\partial \lambda}, \quad f u_g = -\frac{\partial \phi}{\partial y} = -\frac{1}{a} \frac{\partial \phi}{\partial \vartheta}.
\]

Combining these relations with hydrostatic balance, \( \partial \phi / \partial z = b \), gives

\[
\begin{align*}
-f \frac{\partial v_g}{\partial z} &= -\frac{\partial b}{\partial x} = -\frac{1}{a \cos \vartheta} \frac{\partial b}{\partial \lambda} \\
+f \frac{\partial u_g}{\partial z} &= -\frac{\partial b}{\partial y} = -\frac{1}{a} \frac{\partial b}{\partial \vartheta}
\end{align*}
\]

These equations represent thermal wind balance, and the vertical derivative of the geostrophic wind is the ‘thermal wind’. Eq. (2.202b) may be written in terms of the zonal angular momentum as

\[
\frac{\partial m_g}{\partial z} = -\frac{a}{2 \Omega \tan \vartheta} \frac{\partial b}{\partial y},
\]

where \( m_g = (u_g + \Omega a \cos \vartheta) a \cos \vartheta \). Potentially more accurate than geostrophic balance is the so-called cyclostrophic or gradient-wind balance, which retains a centrifugal term in the momentum equation. Thus, we omit only the material derivative in the meridional momentum equation (2.50b) and obtain

\[
2 u \Omega \sin \vartheta + \frac{u^2}{a} \tan \vartheta = -\frac{\partial \phi}{\partial y} = -\frac{1}{a} \frac{\partial \phi}{\partial \vartheta}.
\]
The mechanism of thermal wind. A cold fluid is denser than a warm fluid, so by hydrostasy the vertical pressure gradient is greater where the fluid is cold. Thus, the pressure gradients form as shown, where ‘higher’ and ‘lower’ mean relative to the average at that height. The horizontal pressure gradients are balanced by the Coriolis force, producing (for $f > 0$) the horizontal winds shown (⊗ into the paper, and ◎ out of the paper). Only the wind shear is given by the thermal wind.

For large-scale flow this only differs significantly from geostrophic balance very close to the equator. Combining cyclostrophic and hydrostatic balance gives a modified thermal wind relation, and this takes a simple form when expressed in terms of angular momentum, namely

$$\frac{\partial m^2}{\partial z} \approx -\frac{a^3 \cos^2 \vartheta}{\sin \vartheta} \frac{\partial b}{\partial y}.$$  \hfill (2.205)

If the density or buoyancy is constant then there is no shear and (2.202) or (2.205) give the Taylor-Proudman result. But suppose that the temperature falls in the polewards direction. Then thermal wind balance implies that the (eastwards) wind will increase with height — just as is observed in the atmosphere! In general a vertical shear of the horizontal wind is associated with a horizontal temperature gradient, and this is one of the most simple and far-reaching effects in geophysical fluid dynamics. The basic physical effect is illustrated in Fig. 2.6.

**Pressure coordinates**

In pressure coordinates geostrophic balance is just

$$f \times u_g = -\nabla_p \Phi,$$ \hfill (2.206)

where $\Phi$ is the geopotential and $\nabla_p$ is the gradient operator taken at constant pressure. If $f$ is constant, it follows from (2.206) that the geostrophic wind is non-divergent on pressure surfaces. Taking the vertical derivative of (2.206) (that is, its derivative with respect to $p$) and using the hydrostatic equation, $\partial \Phi / \partial p = -\alpha$, gives the thermal wind equation

$$f \times \frac{\partial u_g}{\partial p} = \nabla_p \alpha = \frac{R}{p} \nabla_p T,$$ \hfill (2.207)
where the last equality follows using the ideal gas equation and because the horizontal derivative is at constant pressure. In component form this is

\[ -f \frac{\partial v}{\partial p} = \frac{R}{p} \frac{\partial T}{\partial x}, \quad f \frac{\partial u}{\partial p} = \frac{R}{p} \frac{\partial T}{\partial y}. \tag{2.208} \]

In log-pressure coordinates, with \( Z = -H \ln(p/p_R) \), thermal wind is

\[ f \times \frac{\partial u}{\partial Z} = -\frac{R}{H} \nabla Z T. \tag{2.209} \]

The physical meaning in all these cases is the same: a horizontal temperature gradient, or a temperature gradient along an isobaric surface, is accompanied by a vertical shear of the horizontal wind.

### 2.8.5 Effects of rotation on hydrostatic balance

Because rotation inhibits vertical motion, we might expect it to affect the requirements for hydrostasy. The simplest setting in which to see this is the rotating Boussinesq equations, \(2.164\). Let us nondimensionalize these by writing

\[
\begin{align*}
(x, y) &= L(\hat{x}, \hat{y}), \quad z = H\hat{z}, \quad u = U\hat{u}, \quad t = T\hat{t}, \quad f = f_0\hat{f}, \\
w &= \frac{\beta HU}{f_0} \hat{w} = \beta \frac{H}{L} \hat{w}, \quad \phi = \Phi\hat{\phi} = f_0 U L \hat{\phi}, \quad b = B\hat{b} = \frac{f_0 u L}{H} \hat{b},
\end{align*}
\]

where \( \hat{\beta} \equiv \beta L / f_0 \). (If \( f \) is constant, then \( \hat{f} \) is a unit vector in the vertical direction.) These relations are the same as \(2.170\), except for the scaling for \( w \), which is suggested by \(2.194\), and the scaling for \( \phi \) and \( b' \), which are suggested by geostrophic and thermal wind balance.

Substituting into \(2.164\) we obtain the following nondimensional momentum equations:

\[
\begin{align*}
\text{Ro} \frac{D\hat{u}}{Dt} + \hat{f} \times \hat{u} &= -\nabla \hat{\phi}, \quad \text{Ro} \hat{\beta} \alpha^2 \frac{D\hat{w}}{Dt} = -\frac{\partial \hat{\phi}}{\partial \hat{z}} - \hat{b},
\end{align*}
\]

Here, \( D/D\hat{t} = \partial/\partial \hat{t} + \hat{u} \cdot \nabla \hat{z} + \hat{\beta} \partial/\partial \hat{z} \) and \( \text{Ro} = U/(f_0 L) \). There are two noteable aspects to these equations. First and most obviously, when \( \text{Ro} \ll 1 \), \(2.211a\) reduces to geostrophic balance, \( f \times u = -\nabla \phi \). Second, the material derivative in \(2.211b\) is multiplied by three nondimensional parameters, and we can understand the appearance of each as follows.

(i) The aspect ratio dependence \( (\alpha^2) \) arises in the same way as for non-rotating flows — that is, because of the presence of \( w \) and \( z \) in the vertical momentum equation as opposed to \((u,v)\) and \((x,y)\) in the horizontal equations.

(ii) The Rossby number dependence \( (\text{Ro}) \) arises because in rotating flow the pressure gradient is balanced by the Coriolis force, which is Rossby number larger than the advective terms.

(iii) The factor \( \hat{\beta} \) arises because in rotating flow \( w \) is smaller than \( u \) by the \( \hat{\beta} \) times the aspect ratio.
The factor $Ro \hat{\beta} \alpha^2$ is very small for large-scale flow; the reader is invited to calculate representative values. Evidently, a rapidly rotating fluid is more likely to be in hydrostatic balance than a non-rotating fluid, other conditions being equal. The combined effects of rotation and stratification are, not surprisingly, quite subtle and we leave that topic for chapter 5.

2.9 STATIC INSTABILITY AND THE PARCEL METHOD

In this and the next couple of sections we consider how a fluid might oscillate if it were perturbed away from a resting state. Our focus is on vertical displacements, and the restoring force is gravity, and we will neglect the effects of rotation, and indeed initially we will neglect horizontal motion entirely. Given that, the simplest and most direct way to approach the problem is to consider from first principles the pressure and gravitational forces on a displaced parcel. To this end, consider a fluid at rest in a constant gravitational field, and therefore in hydrostatic balance. Suppose that a small parcel of the fluid is adiabatically displaced upwards by the small distance $\delta z$, without altering the overall pressure field — that is, the fluid parcel instantly assumes the pressure of its environment. If after the displacement the parcel is lighter than its environment, it will accelerate upwards, because the upward pressure gradient force is now greater downwards gravity force on the parcel — that is, the parcel is *buoyant* (a manifestation of Archimedes’ principle) and the fluid is *statically unstable*. If on the other hand the fluid parcel finds itself heavier than its surroundings, the downward gravitational force will be greater than the upward pressure force and the fluid will sink back towards its original position and an oscillatory motion will develop. Such an equilibrium is *statically stable*. Using such simple ‘parcel’ arguments we will now develop criteria for the stability of the environmental profile.

2.9.1 A simple special case: a density-conserving fluid

Consider first the simple case of an incompressible fluid in which the density of the displaced parcel is conserved, that is $D\rho /Dt = 0$ (and refer to Fig. 2.7 setting $\rho_0 = \rho$). If the environmental profile is $\tilde{\rho}(z)$ and the density of the parcel is $\rho$ then a parcel displaced from a level $z$ [where its density is $\tilde{\rho}(z)$] to a level $z + \delta z$ [where the parcel's density is still $\tilde{\rho}(z)$] will find that its density then differs from its surroundings by the amount

$$\delta \rho = \rho(z + \delta z) - \tilde{\rho}(z + \delta z) = \tilde{\rho}(z) - \tilde{\rho}(z + \delta z) = - \frac{\partial \tilde{\rho}}{\partial z} \delta z. \quad (2.212)$$

The parcel will be heavier than its surroundings, and therefore the parcel displacement will be stable, if $\partial \tilde{\rho} / \partial z < 0$. Similarly, it will be unstable if $\partial \tilde{\rho} / \partial z > 0$. The upward force (per unit volume) on the displaced parcel is given by

$$F = -g \delta \rho = g \frac{\partial \tilde{\rho}}{\partial z} \delta z, \quad (2.213)$$

and thus Newton’s second law implies that the motion of the parcel is determined by

$$\rho(z) \frac{\partial^2 \delta z}{\partial t^2} = g \frac{\partial \tilde{\rho}}{\partial z} \delta z, \quad (2.214)$$
2.9 Static Instability and the Parcel Method

Figure 2.7 A parcel is adiabatically displaced upward from level \( z \) to \( z + \delta z \). If the resulting density difference, \( \delta \rho \), between the parcel and its new surroundings is positive the displacement is stable, and conversely. If \( \tilde{\rho} \) is the environmental values, and \( \rho_\theta \) is potential density, we see that

\[
\delta \rho = \tilde{\rho}_\theta(z) - \tilde{\rho}_\theta(z + \delta z),
\]

or

\[
\frac{\partial^2 \delta z}{\partial t^2} = \frac{g}{\tilde{\rho}} \frac{\partial \tilde{\rho}}{\partial z} \delta z = -N^2 \delta z, \tag{2.215}
\]

where

\[
N^2 = -\frac{g}{\tilde{\rho}} \frac{\partial \tilde{\rho}}{\partial z}, \tag{2.216}
\]

is the buoyancy frequency, or the Brunt-Väisälä frequency, for this problem. If \( N^2 > 0 \) then a parcel displaced upward is heavier than its surroundings, and thus experiences a restoring force; the density profile is said to be stable and \( N \) is the frequency at which the fluid parcel oscillates. If \( N^2 < 0 \), the density profile is unstable and the parcel continues to ascend and convection ensues. In liquids it is often a good approximation to replace \( \tilde{\rho} \) by \( \rho_0 \) in the denominator of (2.216).

2.9.2 The general case: using potential density

More generally, in an adiabatic displacement it is potential density, \( \rho_\theta \), and not density itself that is materially conserved. Consider a parcel that is displaced adiabatically a vertical distance from \( z \) to \( z + \delta z \); the parcel preserves its potential density, and let us use the pressure at level \( z + \delta z \) as the reference level. The in situ density of the parcel at \( z + \delta z \), namely \( \rho(z + \delta z) \), is then equal to its potential density \( \rho_\theta(z + \delta z) \) and, because \( \rho_\theta \) is conserved, this is equal to the potential density of the environment at \( z \), \( \tilde{\rho}_\theta(z) \). The difference in in situ density between the parcel and the environment at \( z + \delta z \), \( \delta \rho \), is thus equal to the difference between the potential density of the environment at \( z \) and at \( z + \delta z \). Putting this together (and see Fig. 2.7) we have

\[
\delta \rho = \rho(z + \delta z) - \tilde{\rho}(z + \delta z) = \rho_\theta(z + \delta z) - \tilde{\rho}_\theta(z + \delta z)
\]

\[
= \rho_\theta(z) - \tilde{\rho}_\theta(z + \delta z) = \tilde{\rho}_\theta(z) - \tilde{\rho}_\theta(z + \delta z),
\]

and therefore

\[
\delta \rho = -\frac{\partial \tilde{\rho}_\theta}{\partial z} \delta z, \tag{2.218}
\]
where the derivative on the right-hand side is the environmental gradient of potential density. If the right-hand side is positive, the parcel is heavier than its surroundings and the displacement is stable. Thus, the conditions for stability are:

<table>
<thead>
<tr>
<th>Stability</th>
<th>( \frac{\partial \tilde{\rho}_\theta}{\partial z} &lt; 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instability</td>
<td>( \frac{\partial \tilde{\rho}_\theta}{\partial z} &gt; 0 )</td>
</tr>
</tbody>
</table>

(2.219a,b)

The equation of motion of the fluid parcel is

\[
\frac{\partial^2 \delta z}{\partial t^2} = \frac{\theta}{\rho} \left( \frac{\partial \tilde{\rho}_\theta}{\partial z} \right) \delta z = -N^2 \delta z,
\]

(2.220)

where, noting that \( \rho(z) = \tilde{\rho}_\theta(z) \) to within \( O(\delta z) \),

\[
N^2 = -\frac{\theta}{\rho_0} \left( \frac{\partial \tilde{\rho}_\theta}{\partial z} \right).
\]

(2.221)

This is a general expression for the buoyancy frequency, true in both liquids and gases. It is important to realize that the quantity \( \tilde{\rho}_\theta \) is the locally-referenced potential density of the environment, as will become more clear below.

**An ideal gas**

In the atmosphere potential density is related to potential temperature by \( \rho_\theta = p_R/\theta R \). Using this in (2.221) gives

\[
N^2 = \frac{\theta}{\theta_0} \left( \frac{\partial \tilde{\theta}}{\partial z} \right),
\]

(2.222)

where \( \tilde{\theta} \) refers to the environmental profile of potential temperature. The reference value \( p_R \) does not appear, and we are free to choose this value arbitrarily — the surface pressure is a common choice. The conditions for stability, (2.219), then correspond to \( N^2 > 0 \) for stability and \( N^2 < 0 \) for instability. In the troposphere (the lowest several kilometers of the atmosphere) the average \( N \) is about 0.01 s\(^{-1}\), with a corresponding period, \( (2\pi/N) \), of about 10 minutes. In the stratosphere (which lies above the troposphere) \( N^2 \) is a few times higher than this.

**A liquid ocean**

No simple, accurate, analytic expression is available for computing static stability in the ocean. If the ocean had no salt, then the potential density referenced to the surface would generally be a measure of the sign of stability of a fluid column, if not of the buoyancy frequency. However, in the presence of salinity, the surface-referenced potential density is not necessarily even a measure of the sign of stability, because the coefficients of compressibility \( \beta_T \) and \( \beta_S \) vary in different ways with pressure. To see this, suppose two neighbouring fluid elements at the surface have the same potential density, but different salinities and temperatures. Displace them
both adiabatically to the deep ocean. Although their potential densities (referenced to the surface) are still equal, we can say little about their actual densities, and hence their stability relative to each other, without doing a detailed calculation because they will each have been compressed by different amounts. It is the profile of the \textit{locally-referenced} potential density that determines the stability.

A sometimes-useful expression for stability arises by noting that in an adiabatic displacement

\[ \delta \rho \theta = \delta \rho - \frac{1}{c_s^2} \delta p = 0. \]  

(2.223)

If the fluid is hydrostatic \( \delta p = -\rho g \delta z \) so that if a parcel is displaced adiabatically its density changes according to

\[ \left( \frac{\partial \rho}{\partial z} \right)_{\rho_0} = -\frac{\rho g}{c_s^2}. \]  

(2.224)

If a parcel is displaced a distance \( \delta z \) upwards then the density difference between it and its new surroundings is

\[ \delta \rho = -\left[ \left( \frac{\partial \rho}{\partial z} \right)_{\rho_0} - \left( \frac{\partial \tilde{\rho}}{\partial z} \right) \right] \delta z = \left[ \rho g \frac{1}{c_s^2} + \left( \frac{\partial \tilde{\rho}}{\partial z} \right) \right] \delta z, \]  

(2.225)

where the tilde again denotes the environmental field. It follows that the stratification is given by

\[ N^2 = -g \left[ \frac{g}{c_s^2} + \frac{1}{\tilde{\rho}} \left( \frac{\partial \tilde{\rho}}{\partial z} \right) \right]. \]  

(2.226)

This expression holds for both liquids and gases, and for ideal gases it is precisely the same as (2.222) (problem 2.8). In liquids, a good approximation is to use a reference value \( \rho_0 \) for the undifferentiated density in the denominator, whence it becomes equal to the Boussinesq expression (2.107). Typical values of \( N \) in the upper ocean where the density is changing most rapidly (i.e., in the pycnocline — ‘pycno’ for density, ‘cline’ for changing) are about 0.01 s\(^{-1}\), falling to 0.001 s\(^{-1}\) in the more homogeneous abyssal ocean. These correspond to periods of about 10 and 100 minutes, respectively.

\* \textit{Cabbeling}

\textit{Cabbeling} is an instability that arises because of the nonlinear equation of state of seawater. From Fig. 1.3 we see that the contours are slightly convex, bowing upward, especially in the plot at sea level. Suppose we mix two parcels of water, each with the same density (\( \sigma_0 = 28 \), say), but with different initial values of temperature and salinity. Then the resulting parcel of water will have a temperature and a salinity equal to the average of the two parcels, but its density will be higher than either of the two original parcels. In the appropriate circumstances such mixing may thus lead to a convective instability; this may, for example, be an important source of ‘bottom water’ formation in the Weddell Sea, off Antarctica.
2.9.3 Lapse rates in dry and moist atmospheres

A dry ideal gas

The negative of the rate of change of the temperature in the vertical is known as the temperature lapse rate, or often just the lapse rate, and the lapse rate corresponding to \( \partial \theta / \partial z = 0 \) is called the dry adiabatic lapse rate and denoted \( \Gamma_d \). Using \( \theta = T(p_0/p)^{R/c_p} \) and \( \partial p / \partial z = -\rho g \) we find that the lapse rate and the potential temperature lapse rate are related by

\[
\frac{\partial T}{\partial z} = T \frac{\partial \theta}{\partial z} - \frac{g}{c_p},
\]

so that the dry adiabatic lapse rate is given by

\[
\Gamma_d = \frac{g}{c_p},
\]

as in fact we derived in (1.134). (We use the subscript \( d \), for dry, to differentiate it from the moist lapse rate considered below.) The conditions for static stability corresponding to (2.219) are thus:

\[
\begin{align*}
\text{Stability:} & \quad \frac{\partial \tilde{\theta}}{\partial z} > 0; \quad \text{or} \quad -\frac{\partial \tilde{T}}{\partial z} < \Gamma_d, \\
\text{Instability:} & \quad \frac{\partial \tilde{\theta}}{\partial z} < 0; \quad \text{or} \quad -\frac{\partial \tilde{T}}{\partial z} > \Gamma_d,
\end{align*}
\]

where a tilde indicates that the values are those of the environment. The atmosphere is in fact generally stable by this criterion: the observed lapse rate, corresponding to an observed buoyancy frequency of about \( 10^{-2} \text{s}^{-1} \), is often about \( 7 \text{Kkm}^{-1} \), whereas a dry adiabatic lapse rate is about \( 10 \text{Kkm}^{-1} \). Why the discrepancy? One reason, particularly important in the tropics, is that the atmosphere contains water vapour.

Effects of water vapour on the lapse rate of an ideal gas

The amount of water vapour that can be contained in a given volume is an increasing function of temperature (with the presence or otherwise of dry air in that volume being largely irrelevant). Thus, if a parcel of water vapour is cooled, it will eventually become saturated and water vapour will condense into liquid water. A measure of the amount of water vapour in a unit volume is its partial pressure, and the partial pressure of water vapour at saturation, \( e_s \), is given by the Clausius-Clapeyron equation,

\[
\frac{de_s}{dT} = \frac{L_v e_s}{R_v T^2},
\]

where \( L_v \) is the latent heat of condensation or vapourization (per unit mass) and \( R_v \) is the gas constant for water vapour. If a parcel rises adiabatically it will cool, and at some height (known as the 'lifting condensation level', a function of its initial temperature and humidity only) the parcel will become saturated and any further ascent will cause the water vapour to condense. The ensuing condensational heating causes the parcel's temperature, and buoyancy, to increase; the parcel thus rises
further, causing more water vapour to condense, and so on, and the consequence of this is that an environmental profile that is stable if the air is dry may be unstable if saturated. Let us now derive an expression for the lapse rate of a saturated parcel that is ascending adiabatically apart from the affects of condensation.

Let \( w \) denote the mass of water vapour per unit mass of dry air, the mixing ratio, and let \( w_s \) be the saturation mixing ratio. \( (w_s = \alpha e_s/(p - e_s) \approx \alpha_w e_s/p \) where \( \alpha_w = 0.62 \), the ratio of the mass of a water molecule to one of dry air.) The diabatic heating associated with condensation is then given by

\[
Q_{\text{cond}} = -L_c \frac{Dw_s}{Dt},
\]

so that the thermodynamic equation is

\[
c_p \frac{D\ln \theta}{Dt} = -\frac{L_c}{T} \frac{Dw_s}{Dt},
\]

or, in terms of \( p \) and \( T \)

\[
c_p \frac{D\ln T}{Dt} - R \frac{D\ln p}{Dt} = -\frac{L_c}{T} \frac{Dw_s}{Dt}.
\]

If these material derivatives are due to the parcel ascent then

\[
\frac{d \ln T}{dz} - \frac{R}{c_p} \frac{d \ln p}{dz} = -\frac{L_c}{T c_p} \frac{dw_s}{dz}.
\]

and using the hydrostatic relationship and the fact that \( w_s \) is a function of \( T \) and \( p \) we obtain

\[
\frac{dT}{dz} + \frac{g}{c_p} = -\frac{L_c}{c_p} \left[ \frac{(\partial w_s/\partial T)_T}{\partial p} \frac{dT}{dz} - \left( \frac{\partial w_s}{\partial p}_T \right) \rho g \right].
\]

Solving for \( dT/dz \), the lapse rate, \( \Gamma_s \), of an ascending saturated parcel is given by

\[
\Gamma_s = -\frac{dT}{dz} = \frac{g}{c_p} \left[ 1 + \frac{L_c}{c_p} \left( \frac{\partial w_s}{\partial T}_T \right) \frac{dT}{dz} \right] \approx \frac{g}{c_p} \left[ 1 + \frac{L_c w_s}{c_p} \left( \frac{RT}{c_p} \right) \right].
\]

where the last near-equality follows with use of the Clausius-Clapeyron relation. This \( (\Gamma_s) \) is variously called the pseudoadiabatic or moist adiabatic or saturated adiabatic lapse rate. Because \( g/c_p \) is the dry adiabatic lapse rate \( \Gamma_d, \Gamma_s < \Gamma_d \), and values of \( \Gamma_s \) are typically around 6 Kkm\(^{-1}\) in the lower atmosphere; however, \( dw_s/dT \) is an increasing function of \( T \) so that \( \Gamma_s \) decreases with increasing temperature and can be as low as 3.5 Kkm\(^{-1}\). For a saturated parcel, the stability conditions analogous to (2.229) are

\[
\text{Stability : } -\frac{\partial \bar{T}}{\partial z} < \Gamma_s, \quad (2.237a)
\]

\[
\text{Instability : } -\frac{\partial \bar{T}}{\partial z} > \Gamma_s. \quad (2.237b)
\]

where \( \bar{T} \) is the environmental temperature. The observed environmental profile in convecting situations is often a combination of the dry adiabatic and moist adiabatic profiles: an unsaturated parcel that is is unstable by the dry criterion will
rise and cool following a dry adiabat, $\Gamma_d$, until it becomes saturated at the lifting condensation level, above which it will rise following a saturation adiabat, $\Gamma_s$. Such convection will proceed until the atmospheric column is stable and, especially in low latitudes, the lapse rate of the atmosphere is largely determined by such convective processes.

* Equivalent potential temperature

Suppose that all the moisture in a parcel of air condenses, and that all the heat released goes into heating the parcel. The equivalent potential temperature, $\theta_{eq}$, is the potential temperature that the parcel then achieves. We may obtain an approximate analytic expression for it by noting that the first law of thermodynamics, $dQ = T\,d\eta$, then implies, by definition of potential temperature,

$$- L_c \, dw = c_p \, T \, d \ln \theta,$$

(2.238)

where $dw$ is the change in water vapour mixing ratio, so that a reduction of $w$ via condensation leads to heating. Integrating gives, by definition of equivalent potential temperature,

$$- \int_w^0 \frac{L_c \, w}{c_p \, T} \, dw = \int_{\theta_{eq}}^{\theta} \, d \ln \theta,$$

(2.239)

and so, if $T$ and $L_c$ are assumed constant,

$$\theta_{eq} = \theta \exp \left( \frac{L_c \, w}{c_p \, T} \right).$$

(2.240)

The equivalent potential temperature so defined is approximately conserved during condensation, the approximation arising going from (2.239) to (2.240). It is a useful expression for diagnostic purposes, and in constructing theories of convection, but it is not accurate enough to use as a prognostic variable in a putatively realistic numerical model. The ‘equivalent temperature’ may be defined in terms of the equivalent potential temperature by

$$T_{eq} = \theta_{eq} \left( \frac{p}{p_R} \right)^{\kappa}.$$

(2.241)

2.10 GRAVITY WAVES

The parcel approach to oscillations and stability, while simple and direct, is divorced from the fluid-dynamical equations of motion, making it hard to include other effects such as rotation, or to explore the effects of possible differences between the hydrostatic and non-hydrostatic cases. To remedy this, we now use the equations of motion to analyze the motion resulting from a small disturbance.

2.10.1 Gravity waves and convection in a Boussinesq fluid

Let us consider a Boussinesq fluid, at rest, in which the buoyancy varies linearly with height and the buoyancy frequency, $N$, is a constant. Linearizing the equations of
2.10 Gravity Waves

motion about this basic state gives the linear momentum equations,
\[
\frac{\partial u'}{\partial t} = -\frac{\partial \phi'}{\partial x}, \hspace{1cm} \frac{\partial w'}{\partial t} = -\frac{\partial \phi'}{\partial z} + b',
\] (2.242a,b)

and mass continuity and thermodynamic equations,
\[
\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} = 0, \hspace{1cm} \frac{\partial b'}{\partial t} + w'N^2 = 0,
\] (2.243a,b)

where for simplicity we assume that the flow is a function only of \(x\) and \(z\). A little algebra gives a single equation for \(w'\),
\[
\left[ \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} \right) \frac{\partial^2}{\partial t^2} + N^2 \frac{\partial^2}{\partial x^2} \right] w' = 0.
\] (2.244)

Seeking solutions of the form \(w' = \text{Re} W \exp[i(kx + mz - \omega t)]\) (where \(\text{Re}\) denotes the real part) yields the dispersion relationship for gravity waves:
\[
\omega^2 = \frac{k^2 N^2}{k^2 + m^2}.
\] (2.245)

The frequency (see Fig. 2.8) is thus always less than \(N\), approaching \(N\) for small horizontal scales, \(k \gg m\). If we neglect pressure perturbations, as in the parcel argument, then the two equations,
\[
\frac{\partial w'}{\partial t} = b', \hspace{1cm} \frac{\partial b'}{\partial t} + w'N^2 = 0,
\] (2.246)

form a closed set, and give \(\omega^2 = N^2\).

If the basic state density increases with height then \(N^2 < 0\) and we expect this state to be unstable. Indeed, the disturbance grows exponentially according to \(\exp(\sigma t)\) where
\[
\sigma = i\omega = \frac{\pm kN}{(k^2 + m^2)^{1/2}},
\] (2.247)
where $\tilde{N}^2 = -N^2$. Most convective activity in the ocean and atmosphere is, ultimately, related to an instability of this form, although of course there are many complicating issues — water vapour in the atmosphere, salt in the ocean, the effects of rotation and so forth.

**Hydrostatic gravity waves and convection**

Let us now suppose that the fluid satisfies the hydrostatic Boussinesq equations. The linearized two-dimensional equations of motion become

\begin{align}
\frac{\partial u'}{\partial t} &= -\frac{\partial \phi'}{\partial x}, \quad 0 = -\frac{\partial \phi'}{\partial z} + b', \\
\frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} &= 0, \\
\frac{\partial b'}{\partial t} + w'N^2 &= 0,
\end{align}

being the horizontal and vertical momentum equations, mass continuity, and the thermodynamic equation respectively. Then a little algebra gives the dispersion relation,

\[ \omega^2 = k^2 \frac{N^2}{m^2}. \]

The frequency and, if $N^2$ is negative the growth rate, is unbounded for as $k/m \to \infty$, and the hydrostatic approximation thus has quite unphysical behaviour for small horizontal scales (see also problem 2.10).

### 2.11 ACOUSTIC-GRAVITY WAVES IN AN IDEAL GAS

We now consider wave motion in a stratified, compressible fluid such as the earth’s atmosphere. The complete problem is complicated and uninformative; we will specialize to the case of an isothermal, stationary atmosphere and ignore the effects of rotation and sphericity, but otherwise we will make few approximations. In this section we will denote the unperturbed state with a subscript 0 and the perturbed state with a prime (‘); we will also omit many algebraic details. Because it is at rest, the basic state is in hydrostatic balance,

\[ \frac{\partial p_0}{\partial z} = -\rho_0(z)g. \]

Ignoring variations in the $y$-direction for algebraic simplicity, the linearized equations of motion are:

\begin{align}
\text{u-momentum:} \quad \rho_0 \frac{\partial u'}{\partial t} &= -\frac{\partial p'}{\partial x} \\
\text{w-momentum:} \quad \rho_0 \frac{\partial w'}{\partial t} &= -\frac{\partial p'}{\partial z} - \rho' g \\
\text{mass conservation:} \quad \frac{\partial \rho'}{\partial t} + w' \frac{\partial \rho_0}{\partial z} &= -\rho_0 \left( \frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} \right) \\
\text{thermodynamic:} \quad \frac{\partial \theta'}{\partial t} + w' \frac{\partial \theta_0}{\partial z} &= 0
\end{align}
2.11 * Acoustic-Gravity Waves in an Ideal Gas

The equation of state is given by:

\[
\frac{\theta'}{\theta_0} + \frac{\rho'}{\rho_0} = \frac{1}{\gamma} \frac{p'}{p_0}. \tag{2.251e}
\]

For an isothermal basic state we have \( p_0 = \rho_0 RT_0 \) where \( T_0 \) is a constant, so that \( \rho_0 = \rho_s e^{-z/H} \) and \( p_0 = p_s e^{-z/H} \) where \( H = RT_0/g \). Further, using \( \theta = T(p_s/p)^\kappa \) where \( \kappa = R/c_p \), we have that \( \theta_0 = T_0 e^{z/H} \) and so \( N^2 = \kappa g/H \). It is also convenient to use (1.100) on page 22 to write the linear thermodynamic equation in the form:

\[
\frac{\partial p'}{\partial t} - w' p_0 = -\gamma p_0 \left( \frac{\partial u'}{\partial x} + \frac{\partial w'}{\partial z} \right). \tag{2.251f}
\]

Differentiating (2.251b) with respect to time and using (2.251f) leads to:

\[
\left( \frac{\partial^2}{\partial t^2} - c_s^2 \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{\gamma} \frac{\partial}{\partial z} \right) \right) w' = c_s^2 \left( \frac{\partial}{\partial z} - \frac{\kappa}{H} \right) \frac{\partial u'}{\partial x}, \tag{2.252a}
\]

where \( c_s^2 = (\gamma p/\gamma \rho)_0 = \gamma RT_0 = \gamma p_0/\rho_0 \) is the square of the speed of sound, and \( \gamma = c_p/c_v = 1/(1 - \kappa) \). Similarly, differentiating (2.251b) with respect to time and using (2.251f) and (2.251f) leads to:

\[
\left( \frac{\partial^2}{\partial t^2} - c_s^2 \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{\gamma} \frac{\partial}{\partial z} \right) \right) w' = c_s^2 \left( \frac{\partial}{\partial z} - \frac{\kappa}{H} \right) \frac{\partial u'}{\partial x}\tag{2.252b}
\]

Eqs. (2.252a) and (2.252b) combine to give, after some cancellation,

\[
\frac{\partial^4 w'}{\partial t^4} - c_s^2 \frac{\partial^2}{\partial t^2} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{\gamma} \frac{\partial}{\partial z} \right) w' - c_s^2 \frac{k g}{H} \frac{\partial^2 w'}{\partial x^2} = 0. \tag{2.253}
\]

If we set \( w' = W(x, z, t) e^{\pm i(2Ht)} \), so that \( W = (\rho_0/\rho_s)^{1/2} w \), then the term with the single \( z \)-derivative is eliminated, giving

\[
\frac{\partial^4 W}{\partial t^4} - c_s^2 \frac{\partial^2}{\partial t^2} \left( \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial z^2} - \frac{1}{4H^2} \right) W - c_s^2 \frac{k g}{H} \frac{\partial^2 w'}{\partial x^2} = 0. \tag{2.254}
\]

Although superficially complicated, this equation has constant coefficients and we may seek wavelike solutions of the form

\[
W = \text{Re} \tilde{W} e^{i(k x + m z - \omega t)}, \tag{2.255}
\]

where \( \tilde{W} \) is the complex wave amplitude. Using (2.255) in (2.254) leads to the dispersion relation for acoustic-gravity waves, namely

\[
\omega^4 - c_s^2 \omega^2 \left( k^2 + m^2 + \frac{1}{4H^2} \right) + c_s^2 N^2 k^2 = 0, \tag{2.256}
\]

with solution

\[
\omega^2 = \frac{1}{2} c_s^2 K^2 \left[ 1 \pm \left( 1 - \frac{4N^2 k^2}{c_s^2 K^4} \right)^{1/2} \right], \tag{2.257}
\]

where \( K^2 = k^2 + m^2 + 1/(4H^2) \). (The factor \( [1 - 4N^2 k^2/(c_s^2 K^4)] \) is always positive — see problem 2.23.) For an isothermal, ideal-gas, atmosphere \( 4N^2 H^2/c_s^2 \approx 0.8 \) and so this may be written

\[
\frac{\omega^2}{N^2} \approx 2.5 \tilde{k}^2 \left[ 1 \pm \left( 1 - \frac{0.8 \hat{k}^2}{\hat{K}^4} \right)^{1/2} \right], \tag{2.258}
\]

where \( \tilde{k}^2 = \hat{k}^2 + \hat{m}^2 + 1/4 \), and \((\hat{k}, \hat{m}) = (kH, mH)\).
Fig. 2.9 Dispersion diagram for acoustic gravity waves in an isothermal atmosphere, calculated using (2.258). The frequency is given in units of the buoyancy frequency $N$, and the wavenumbers are non-dimensionalized by the inverse of the scale height, $H$. The solid curves indicate acoustic waves, whose frequency is always higher than that of the corresponding Lamb wave at the same wavenumber (i.e. $ck$), and of the base acoustic frequency $\approx 1.12N$. The dashed curves indicate internal gravity waves, whose frequency asymptotes to $N$ at small horizontal scales.

### 2.11.1 Interpretation

**Acoustic and gravity waves**

There are two branches of roots in (2.257), corresponding to acoustic waves (using the plus sign in the dispersion relation) and internal gravity waves (using the minus sign). These (and the Lamb wave, described below) are plotted in Fig. 2.9. If $4N^2k^2/c_s^2K^4 \ll 1$ then the two sets of waves are well separated. From (2.258) this is satisfied when

$$\frac{4k}{\gamma} (kH)^2 \approx 0.8(kH)^2 \ll \left[ (kH)^2 + (mH)^2 + \frac{1}{4} \right]^2 ; \quad (2.259)$$

that is, when either $mH \gg 1$ or $kH \gg 1$. The two roots of the dispersion relation are then

$$\omega_a^2 \approx c_s^2k^2 = c_s^2\left(k^2 + m^2 + \frac{1}{4H^2}\right) \quad (2.260)$$

and

$$\omega_g^2 \approx \frac{N^2k^2}{k^2 + m^2 + 1/(4H^2)} ; \quad (2.261)$$
corresponding to acoustic and gravity waves, respectively. The acoustic waves owe their existence to the presence of compressibility in the fluid, and they have no counterpart in the Boussinesq system. On the other hand, the internal gravity waves are just modified forms of those found in the Boussinesq system, and if we take the limit \((kH, mH) \to \infty\) then the gravity wave branch reduces to
\[
\omega^2 = \frac{N^2 k^2}{k^2 + m^2},
\]
which is the dispersion relationship for gravity waves in the Boussinesq approximation. We may consider this to be the limit of infinite scale height or (equivalently) the case in which wavelengths of the internal waves are sufficiently small that the fluid is essentially incompressible.

**Vertical structure**
Recall that
\[
w' = W(x, z, t)e^{z/(2H)}
\]
and, by inspection of (2.252), \(u'\) has the same vertical structure. That is,
\[
w' \propto e^{z/(2H)}, \quad u' \propto e^{z/(2H)},
\]
and the amplitude of the velocity field of the internal waves increases with height. The pressure and density perturbation amplitudes fall off with height, varying like
\[
p' \propto e^{-z/(2H)}, \quad \rho' \propto e^{-z/(2H)}.
\]
The kinetic energy of the perturbation, \(\rho_0(u'^2 + w'^2)\) is constant with height, because \(\rho_0 = \rho_s e^{-z/H}\).

**Hydrostatic approximation and Lamb waves**
Equations (2.252) also admit to a solution with \(w' = 0\). We then have
\[
\left(\frac{\partial^2}{\partial t^2} - c_s^2 \frac{\partial^2}{\partial x^2}\right) u' = 0 \quad \text{and} \quad \left(\frac{\partial}{\partial z} - \frac{\kappa}{H}\right) \frac{\partial u'}{\partial x} = 0,
\]
and these have solutions of the form
\[
u' = \text{Re} \tilde{U} e^{\kappa x/H} e^{i(kx - \omega t)}, \quad \omega = ck,
\]
where \(\tilde{U}\) is the wave amplitude. These are horizontally propagating sound waves, known as Lamb waves after the hydrodynamicist Horace Lamb. Their velocity perturbation amplitude increases with height, but the pressure perturbation falls with height; that is
\[
u' \propto e^{\kappa x/H} \approx e^{2z/(7H)}, \quad p' \propto e^{(\kappa - 1)z/H} \approx e^{-5z/(7H)}.
\]
Their kinetic energy density \(\rho_0 u'^2\) varies as
\[
\text{K.E.} \propto e^{-z/H + 2\kappa x/H} = e^{(2R - c_P)z/(c_P H)} = e^{(R - c_v)z/(c_P H)} \approx e^{-3z/(7H)}
\]
for an ideal gas. (In a simple ideal gas, \(c_v = nR/2\) where \(n\) is the number of excited degrees of freedom, 5 for a diatomic molecule.) The kinetic energy density thus falls away exponentially from the surface, and in this sense Lamb waves are an example of edge waves or surface-trapped waves.

Consider now case in which we make the hydrostatic approximation \textit{ab initio},
but do not restrict the perturbation to have \( w' = 0 \). The linearized equations are identical to (2.251), except that (2.251b) is replaced by

\[
\frac{\partial p'}{\partial z} = -\rho' g. \tag{2.268}
\]

The consequence of this is that first term \( (\partial^2 w' / \partial t^2) \) in (2.252b) disappears, as do the first two terms in (2.253) [the terms \( \delta^2 w' / \partial t^4 - c^2 (\partial^2 / \partial t^2)(\partial^2 w' / \partial x^2) \)]. It is a simple matter to show that the dispersion relation is then

\[
\omega^2 = \frac{N^2 k^2}{m^2 + 1/(4H^2)}. \tag{2.269}
\]

These are long gravity waves, and may be compared with the corresponding Boussinesq result (2.249). Again, the frequency increases without bound as the horizontal wavelength diminishes. The Lamb wave, of course, still exists in the hydrostatic model, because (2.264) is still a valid solution. Thus, horizontally propagating sound waves still exist in hydrostatic (primitive equation) models, but vertically propagating sound waves do not — essentially because the term \( \partial w / \partial t \) is absent from the vertical momentum equation.

### 2.12 THE EKMAN LAYER

In the final topic of this chapter, we return to geostrophic flow and consider the effects of friction. The fluid fields in the interior of a domain are often set by different physical processes than those occurring at a boundary, and consequently often change rapidly in a thin boundary layer, as in Fig. 2.10. Such boundary layers nearly always involve one or both of viscosity and diffusion, because these appear in the terms of highest differential order in the equations of motion, and so are responsible for the number and type of boundary conditions that the equations must satisfy — for example, the presence of molecular viscosity leads to the condition that the tangential flow (as well as the normal flow) must vanish at a rigid surface.

In many boundary layers in non-rotating flow the dominant balance in the momentum equation is between the advective and viscous terms. In some contrast,
in large-scale atmospheric and oceanic flow the effects of rotation are large, and this results in a boundary layer, known as the *Ekman layer*, in which the dominant balance is between Coriolis and frictional terms.\(^{12}\) Now, the direct effects of molecular viscosity and diffusion are nearly always negligible at distances more than a few millimeters away from a solid boundary, but it is inconceivable that the entire boundary layer between the free atmosphere (or ocean) and the surface is only a few millimeters thick. Rather, in practice a balance occurs between the Coriolis terms and the stress due to small-scale turbulent motion, and this gives rise to a boundary layer that has a typical depth of tens to hundreds of meters. Because the stress arises from the turbulence we cannot with confidence determine its precise form; thus, we should try to determine what general properties Ekman layers may have that are *independent* of the precise form of the friction.

The atmospheric Ekman layer occurs near the ground, and the stress at the ground itself is due to the surface wind (and its vertical variation). In the ocean the main Ekman layer is near the surface, and the stress at ocean surface is largely due to the presence of the overlying wind. There is also a weak Ekman layer at the bottom of the ocean, analogous to the atmospheric Ekman layer. To analyze all these layers, let us assume that:

* The Ekman layer is Boussinesq. This is a very good assumption for the ocean, and a reasonable one for the atmosphere if the boundary layer is not too deep.
* The Ekman layer has a finite depth that is less than the total depth of the fluid, this depth being given by the level at which the frictional stresses essentially vanish. Within the Ekman layer, frictional terms are important, whereas geostrophic balance holds beyond it.
* The nonlinear and time dependent terms in the equations of motion are negligible, hydrostatic balance holds in the vertical, and buoyancy is constant, not varying in the horizontal.
* As needed, we shall assume that friction can be parameterized by a viscous term of the form \(\frac{1}{\rho_0} \frac{\partial \tau}{\partial z} = A \frac{\partial^2 u}{\partial z^2}\), where \(A\) is constant and \(\tau\) is the stress. In laboratory settings \(A\) may be the molecular viscosity, whereas in the atmosphere and ocean it is a so-called *eddy viscosity*. (In turbulent flows momentum is transferred by the near-random motion of small parcels of fluid and, by analogy with the motion of molecules that produces a molecular viscosity, the associated stress is approximately represented, or parameterized, using a turbulent or eddy viscosity that may be orders of magnitude larger than the molecular one.) In all cases it is the vertical derivative of the stress that dominates.

### 2.12.1 Equations of motion and scaling

Frictional-geostrophic balance in the horizontal momentum equation is:

\[
\mathbf{f} \times \mathbf{u} = -\nabla_z \phi + \frac{\partial \tilde{\tau}}{\partial z}.
\]  

\[\text{(2.270)}\]
where \( \tilde{\tau} \equiv \tau/\rho_0 \) is the kinematic stress (and \( \tau \) is the stress itself), and \( f \) is allowed to vary with latitude. If we model the stress with an eddy viscosity (2.270) becomes

\[
\tilde{f} \times \tilde{u} = -\nabla \tilde{\phi} + A \frac{\partial^2 \tilde{u}}{\partial \tilde{z}^2}.
\]  

(2.271)

The vertical momentum equation is hydrostatic balance, \( \partial \phi / \partial z = b \), and, because buoyancy is constant, we may without loss of generality write this as

\[
\frac{\partial \phi}{\partial z} = 0.
\]  

(2.272)

The equation set is completed by the mass continuity equation \( \nabla \cdot \mathbf{v} = 0 \).

The \textit{Ekman number}

We non-dimensionalize the equations by setting

\[
(u, v) = U(\hat{u}, \hat{v}), \quad (x, y) = L(\hat{x}, \hat{y}), \quad f = f_0 \hat{f}, \quad z = H \hat{z}, \quad \phi = \Phi \hat{\phi},
\]  

(2.273)

where hatted variables are non-dimensional. \( H \) is a scaling for the height, and at this stage we will suppose it to be some height scale in the free atmosphere or ocean, not the height of the Ekman layer itself. Geostrophic balance suggests that \( \Phi = f_0 UL \). Substituting (2.273) into (2.271) we obtain

\[
\hat{f} \times \hat{u} = -\hat{\nabla} \hat{\phi} + \text{Ek} \frac{\partial^2 \hat{u}}{\partial \hat{z}^2},
\]  

(2.274)

where the parameter

\[
\text{Ek} \equiv \left( \frac{A}{f_0 H^2} \right),
\]  

(2.275)

is the \textit{Ekman number}, and it determines the importance of frictional terms in the horizontal momentum equation. If \( \text{Ek} \ll 1 \) then the friction is small in the flow interior where \( \hat{z} = \mathcal{O}(1) \). However, the friction term cannot necessarily be neglected in the boundary layer because it is of the highest differential order in the equation, and so determines the boundary conditions; if \( \text{Ek} \) is small the vertical scales become small and the second term on the right-hand side of (2.274) remains finite. The case when this term is simply omitted from the equation is therefore a \textit{singular limit}, meaning that it differs from the case with \( \text{Ek} \to 0 \). If \( \text{Ek} \gtrsim 1 \) friction is important everywhere, but it is usually the case that \( \text{Ek} \) is small for atmospheric and oceanic large-scale flow, and the interior flow is very nearly geostrophic. (In part this is because \( A \) itself is only large near a rigid surface where the presence of a shear creates turbulence and a significant eddy viscosity.)

\textit{Momentum balance in the Ekman layer}

For definiteness, suppose the fluid lies above a rigid surface at \( z = 0 \). Sufficiently far away from the boundary the velocity field is known, and we suppose this flow to be in geostrophic balance. We then write the velocity field and the pressure field as the sum of the interior geostrophic part, plus a boundary layer correction:

\[
\hat{u} = \hat{u}_g + \hat{u}_E, \quad \hat{\phi} = \hat{\phi}_g + \hat{\phi}_E,
\]  

(2.276)
where the Ekman layer corrections, denoted with a subscript $E$, are negligible away from the boundary layer. Now, in the fluid interior we have, by hydrostatic balance, \( \partial \hat{\phi}_g / \partial \hat{z} = 0 \). In the boundary layer we have still \( \partial \hat{\phi}_g / \partial \hat{z} = 0 \) so that, to satisfy hydrostasy, \( \partial \hat{\phi}_E / \partial \hat{z} = 0 \). But because \( \hat{\phi}_E \) vanishes away from the boundary we have \( \hat{\phi}_E = 0 \) everywhere. This is an important result: there is no boundary layer in the pressure field. Note that this is a much stronger result than saying that pressure is continuous, which is nearly always true in fluids; rather, it is a special result about Ekman layers.

Using (2.276) with \( \hat{\phi}_E = 0 \), the dimensional horizontal momentum equation (2.270) becomes, in the Ekman layer,

\[
f \times u_E = \frac{\partial \tau}{\partial \hat{z}}. \tag{2.277}
\]

The dominant force balance in the Ekman layer is thus between the Coriolis force and the friction. We can determine the thickness of the Ekman layer if we model the stress with an eddy viscosity so that

\[
f \times u_E = A \frac{\partial^2 u_E}{\partial \hat{z}^2}, \tag{2.278}
\]

or, non-dimensionally,

\[
\hat{f} \times \hat{u}_E = \hat{E}k \frac{\partial^2 \hat{u}_E}{\partial \hat{z}^2}. \tag{2.279}
\]

It is evident this equation can only be satisfied if \( \hat{z} \neq O(1) \), implying that \( H \) is not a proper scaling for \( z \) in the boundary layer. Rather, if the vertical scale in the Ekman layer is \( \hat{\delta} \) (meaning \( \hat{z} \sim \hat{\delta} \)) we must have \( \hat{\delta} \sim \hat{E}k^{1/2} \). In dimensional terms this means the thickness of the Ekman layer is

\[
\delta = H \hat{\delta} = H \hat{E}k^{1/2} \tag{2.280}
\]

or

\[
\delta = \left( \frac{A}{f_0} \right)^{1/2}. \tag{2.281}
\]

[This estimate also emerges directly from (2.278).] Note that (2.280) can be written as

\[
\hat{E}k = \left( \frac{\delta}{H} \right)^2. \tag{2.282}
\]

That is, the Ekman number is equal to the square of the ratio of the depth of the Ekman layer to an interior depth scale of the fluid motion. In laboratory flows where \( A \) is the molecular viscosity we can thus estimate the Ekman layer thickness, and if we know the eddy viscosity of the ocean or atmosphere we can estimate the thickness of their respective Ekman layers. We can invert this argument and obtain an estimate of \( A \) if we know the Ekman layer depth. In the atmosphere, deviations from geostrophic balance are very small in the atmosphere above 1 km, and using this gives \( A \approx 10^2 \text{ m}^2 \text{ s}^{-1} \). In the ocean Ekman depths are about 50 m or less, and eddy viscosities about 0.1 m$^2$ s$^{-1}$. 

2.12.2 Integral properties of the Ekman layer

What can we deduce about the Ekman layer without specifying the detailed form of the frictional term? Using dimensional notation we recall frictional-geostrophic balance,

\[ f \times u = -\nabla \phi + \frac{\partial \tilde{\tau}}{\partial z}, \] (2.283)

where \( \tilde{\tau} \) is zero at the edge of the Ekman layer. In the Ekman layer itself we have

\[ f \times u_E = \frac{\partial \tilde{\tau}}{\partial z}. \] (2.284)

Consider either a top or bottom Ekman layer, and integrate over its thickness. From (2.284) we obtain

\[ f \times M_E = \tilde{\tau}_t - \tilde{\tau}_b, \] (2.285)

where

\[ M_E = \int_{\text{Ek}} u_E \, dz \] (2.286)

is the ageostrophic transport in the Ekman layer, and where \( \tilde{\tau}_t \) and \( \tilde{\tau}_b \) is the stress at the top and the bottom of the respective layer. The former (latter) will be zero in a bottom (top) Ekman layer. More explicitly, (2.285) may be written as:

| Top Ekman Layer: | \( M_E = -\frac{1}{f} k \times \tilde{\tau}_t \) |
| Bottom Ekman Layer: | \( M_E = \frac{1}{f} k \times \tilde{\tau}_b \) |

(2.287a,b)

The transport in the Ekman layer is thus at right-angles to the stress at the surface. This has a simple physical explanation: integrated over the depth of the Ekman layer the surface stress must be balanced by the Coriolis force, which in turn acts at right angles to the mass transport. This result is particularly useful in the ocean, where the stress at the upper surface is primarily due to the wind, and can be regarded as independent of the interior flow. If \( f \) is positive, as in the Northern hemisphere, then an Ekman transport is induced 90° to the right of the direction of the wind stress. This has innumerable important consequences — for example, in inducing coastal upwelling when, as is not uncommon, the wind blows parallel to the coast. Upwelling off the coast of California is one example. In the atmosphere, however, the stress arises as a consequence of the interior flow, and we need to parameterize the stress in terms of the flow in order to calculate the surface stress.

Finally, we obtain an expression for the vertical velocity induced by an Ekman layer. The mass conservation equation is

\[ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0. \] (2.288)

Integrating this over an Ekman layer gives

\[ \nabla \cdot M_t = -(w_t - w_b), \] (2.289)
where $M_t$ is the total (Ekman plus geostrophic) transport in the Ekman layer,

$$M_t = \int_{Ek} u \, dz = \int_{Ek} (u_g + u_E) \, dz \equiv M_g + M_E,$$  

(2.290)

and $w_t$ and $w_b$ are the vertical velocities at the top and bottom of the Ekman layer; the former (latter) is zero in a top (bottom) Ekman layer. From (2.285)

$$k \times (M_t - M_g) = \frac{1}{f} (\tilde{\tau}_t - \tilde{\tau}_b).$$  

(2.291)

Taking the curl of this (i.e., cross-differentiating) gives

$$\nabla \cdot (M_t - M_g) = \text{curl}_z [(\tilde{\tau}_t - \tilde{\tau}_b)/f],$$  

(2.292)

where the curl$_z$ operator on a vector $A$ is defined by curl$_z A \equiv \partial x A_y - \partial y A_x$. Using (2.289) we obtain, for bottom and top Ekman layers respectively,

$$w_b = \text{curl}_z \frac{\tilde{\tau}_t}{f} + \nabla \cdot M_g, \quad w_t = \text{curl}_z \frac{\tilde{\tau}_b}{f} - \nabla \cdot M_g,$$  

(2.293a,b)

where $\nabla \cdot M_g = -\beta M_g/f$ is the divergence of the geostrophic transport in the Ekman layer, which is often small compared to the other terms in these equations. Thus, friction induces a vertical velocity at the edge of the Ekman layer, proportional to the curl of the stress at the surface, and this is perhaps the most used result in Ekman layer theory. Numerical models sometimes do not have the vertical resolution to explicitly resolve an Ekman layer, and (2.293) provides a means of parameterizing the Ekman layer in terms of resolved or known fields. It is particularly useful for the top Ekman layer in the ocean, where the stress can be regarded as a given function of the overlying wind.

### 2.12.3 Explicit solutions. I: A bottom boundary layer

We now assume that the frictional terms can be parameterized as an eddy viscosity and calculate the explicit form of the solution in the boundary layer. The frictional-geostrophic balance may be written as

$$f \times (u - u_g) = A \frac{\partial^2 u}{\partial z^2},$$  

(2.294a)

where

$$f(u_g, v_g) = \left( -\frac{\partial \phi}{\partial y}, \frac{\partial \phi}{\partial x} \right).$$  

(2.294b)

We continue to assume there are no horizontal gradients of temperature, so that, via thermal wind, $\partial u_g/\partial z = \partial v_g/\partial z = 0$.

**Boundary conditions and solution**

Appropriate boundary conditions for a bottom Ekman layer are:

At $z = 0$: \quad $u = 0, \quad v = 0$ (the no slip condition)  

(2.295a)
Figure 2.11 The idealised Ekman layer solution in the lower atmosphere, plotted as a hodograph of the wind components: the arrows show the velocity vectors at a particular heights, and the curve traces out the continuous variation of the velocity. The values on the curve are of the nondimensional variable $z/d$, where 

$$d = \left(\frac{2A}{f}\right)^{1/2},$$

and $v_\theta$ is chosen to be zero.

As $z \to \infty$: \quad $u = u_\theta$, \quad $v = v_\theta$ \quad (a geostrophic interior). \quad (2.295b)

Let us seek solutions to (2.294a) of the form

$$u = u_\theta + A_0 e^{\alpha z}, \quad v = v_\theta + B_0 e^{\alpha z},$$ \quad (2.296)

where $A_0$ and $B_0$ are constants. Substituting into (2.294a) gives two homogeneous algebraic equations

$$A_0 f - B_0 A \alpha^2 = 0, \quad -A_0 A \alpha^2 - B_0 f = 0.$$ \quad (2.297a,b)

For non-trivial solutions the solvability condition $\alpha^4 = -f^2/A^2$ must hold, from which we find $\alpha = \pm (1 \pm i) \sqrt{f/2A}$. Using the boundary conditions we then obtain the solution

$$u = u_\theta - e^{-z/d} \left[ u_\theta \cos(z/d) + v_\theta \sin(z/d) \right],$$ \quad (2.298a)

$$v = v_\theta + e^{-z/d} \left[ u_\theta \sin(z/d) - v_\theta \cos(z/d) \right],$$ \quad (2.298b)

where $d = \sqrt{2A/f}$ is, within a constant factor, the depth of the Ekman layer obtained from scaling considerations. The solution decays exponentially from the surface with this e-folding scale, so that $d$ is a good measure of the Ekman layer thickness. Note that the boundary layer correction depends on the interior flow, since the boundary layer serves to bring the flow to zero at the surface.

To illustrate the solution, suppose that the pressure force is directed in the $y$-direction (northward), so that the geostrophic current is eastward. Then the solution, the now-famous Ekman spiral, is plotted in Fig. 2.11 and Fig. 2.12. The wind falls to zero at the surface, and its direction just above the surface is northeastward; that is, it is rotated by 45° to the left of its direction in the free atmosphere. Although this result is independent of the value of the frictional coefficients, it is dependent on the form of the friction chosen. The force balance in the Ekman layer is between the Coriolis force, the stress, and the pressure force. At the surface the Coriolis force is zero, and the balance is entirely between the northward pressure force and the southwards stress force.
2.12 The Ekman Layer

Figure 2.12 Solutions for a bottom Ekman layer with a given flow in the fluid interior (left), and for a top Ekman layer with a given surface stress (right), both with \( d = 1 \). On the left we have \( u_g = 1, v_g = 0 \). On the right we have \( u_g = v_g = 0, \tilde{\tau}_y = 0 \) and \( \sqrt{2\tilde{\tau}_x/(fd)} = 1 \).

Mass transport, force balance and vertical velocity

The cross-isobaric flow is given by (for \( v_g = 0 \))

\[
V = \int_0^\infty v \, dz = \int_0^\infty u_g e^{-z/d} \sin(z/d) \, dz = \frac{u_g d}{2}.
\]  
(2.299)

For positive \( f \), this is to the left of the geostrophic flow — that is, down the pressure gradient. In the general case (\( v_g \neq 0 \)) we obtain

\[
V = \int_0^\infty (v - v_g) \, dz = \frac{d}{2}(u_g - v_g).
\]  
(2.300)

Similarly, the additional zonal transport produced by frictional effects are, for \( v_g = 0 \),

\[
U = \int_0^\infty (u - u_g) \, dz = -\int_0^\infty e^{-z/d} \sin(z/d) \, dz = -\frac{u_g d}{2},
\]  
(2.301)

and in the general case

\[
U = \int_0^\infty (u - u_g) \, dz = -\frac{d}{2}(u_g + v_g).
\]  
(2.302)

Thus, the total transport caused by frictional forces is

\[
M_E = \frac{d}{2} \left[ -i(u_g + v_g) + j(u_g - v_g) \right].
\]  
(2.303)

The total stress at the bottom surface \( z = 0 \) induced by frictional forces is

\[
\tilde{\tau}_b = A \frac{\partial u}{\partial z} \bigg|_{z=0} = \frac{A}{d} \left[ i(u_g - v_g) + j(u_g + v_g) \right],
\]  
(2.304)

using the solution (2.299). Thus, using (2.303), (2.304) and \( d^2 = 2A/f \), we see that the total frictionally induced transport in the Ekman layer is related to the
stress at the surface by \( M_E = (k \times \tilde{r}_b) / f \), reprising the result of the more general analysis, (??). From (2.304), the stress is at an angle of 45° to the left of the velocity at the surface. (However, this result is not generally true for all forms of stress.) These properties are illustrated in Fig. 2.13.

The vertical velocity at the top of the Ekman layer, \( w_E \), is obtained using (2.303) and (2.304). If \( f \) is constant we obtain

\[
\begin{align*}
  w_E &= -\nabla \cdot M_E = \frac{1}{f_0} \text{curl}_z \tilde{r}_b = V_x - U_y = \frac{d}{2} \zeta_g, \\
  & \quad \text{(2.305)}
\end{align*}
\]

where \( \zeta_g \) is the vorticity of the geostrophic flow. Thus, the vertical velocity at the top of the Ekman layer, which arises because of the frictionally-induced divergence of the cross-isobaric flow in the Ekman layer, is proportional to the geostrophic vorticity in the free fluid and is proportional to the Ekman layer height \( \sqrt{2A/f_0} \).

**Another bottom boundary condition**

In the analysis above we assumed a no slip condition at the surface, namely that the velocity tangential to the surface vanishes. This is certainly appropriate if \( A \) is a molecular velocity, but in a turbulent flow, where \( A \) is interpreted as an eddy viscosity, the flow very close to the surface may be far from zero. Then, unless we wish to explicitly calculate the flow in an additional very thin viscous boundary layer the no-slip condition may be inappropriate. An alternative, slightly more general boundary condition is to suppose that the stress at the surface is given by

\[
\tau = \rho_0 C u, \quad \text{(2.306)}
\]

where \( C \) is a constant. The surface boundary condition is then

\[
A \frac{\partial u}{\partial z} = C u. \quad \text{(2.307)}
\]

If \( C \) is infinite we recover the no-slip condition. If \( C = 0 \), we have instead a condition of no stress at the surface, also known as a free slip condition. For intermediate values of \( C \) the boundary condition is known as a ‘mixed condition’. Evaluating the solution in these cases is left as an exercise for the reader (problem 2.25).

**2.12.4 Explicit solutions. II: The upper ocean**
2.12 The Ekman Layer

Figure 2.14 An idealized Ekman spiral in a Southern Hemisphere ocean, driven by an imposed wind-stress. A Northern Hemisphere spiral would be the reflection of this about the vertical axis. Such a clean spiral is rarely observed in the real ocean. The net transport is at right angles to the wind, independent of the detailed form of the friction. The angle of the surface flow is 45° to the wind only for a Newtonian viscosity.

Boundary conditions and solution

The wind provides a stress on the upper ocean, and the Ekman layer serves to communicate this to the oceanic interior. Appropriate boundary conditions are thus:

At \( z = 0 \):
\[
\Lambda \frac{\partial u}{\partial z} = \tilde{\tau}^x, \quad \Lambda \frac{\partial v}{\partial z} = \tilde{\tau}^y
\]  \hspace{1cm} (a given surface stress)  
(2.308a)

As \( z \to -\infty \):
\[
u = v_g, \quad v = v_g
\]  \hspace{1cm} (a geostrophic interior)  
(2.308b)

where \( \tilde{\tau} \) is the given (kinematic) wind stress at the surface. Solutions to (2.294a) with (2.308) are found by the same methods as before, and are

\[
u = v_g + \sqrt{2 \over f d} e^{z/d} \left[ \tilde{\tau}^x \cos(z/d - \pi/4) - \tilde{\tau}^y \sin(z/d - \pi/4) \right],
\]  \hspace{1cm} (2.309)

and

\[
u = v_g + \sqrt{2 \over f d} e^{z/d} \left[ \tilde{\tau}^x \sin(z/d - \pi/4) + \tilde{\tau}^y \cos(z/d - \pi/4) \right].
\]  \hspace{1cm} (2.310)

Note that the boundary layer correction depends only on the imposed surface stress, and not the interior flow itself. This is a consequence of the type of boundary conditions chosen, for in the absence of an imposed stress the boundary layer correction is zero — the interior flow already satisfies the gradient boundary condition at the top surface. Similar to the bottom boundary layer the velocity vectors of the solution trace a diminishing spiral as they descend into the interior (Fig. 2.14, which is drawn for the Southern Hemisphere).

Mass flux, surface flow and vertical velocity

The mass flux induced by the surface stress is obtained by integrating (2.309) and (2.310) from the surface to \( -\infty \). We explicitly find

\[
U = \int_{-\infty}^{0} (u - u_g) \, dz = \frac{\tilde{\tau}^y}{f}, \quad V = \int_{-\infty}^{0} (v - v_g) \, dz = -\frac{\tilde{\tau}^x}{f}.
\]  \hspace{1cm} (2.311)
which indicates that the ageostrophic mass transport is perpendicular to the wind-stress, as noted previously from more general considerations.

Suppose that the surface wind is eastward. Then $\tilde{\tau}_y = 0$ and the solutions immediately give

$$u(0) - u_g = (\tilde{\tau}_x / fd) \cos(\pi / 4), \quad v(0) - v_g = (\tilde{\tau}_x / fd) \sin(\pi / 4).$$

(2.312)

Therefore the magnitudes of the frictional flow in the $x$ and $y$ directions are equal to each other, and the flow is 45° to the right (for $f > 0$) of the wind. This result is dependent on the form of the frictional parameterization chosen, but not on the size of the viscosity.

At the edge of the Ekman layer the vertical velocity is given by (2.293), and so is proportional to the curl of the wind-stress. (The second term on the right-hand side of (2.293) is the vertical velocity due to the divergence of the geostrophic flow, and is usually much smaller than the first term.) The production of a vertical velocity at the edge of the Ekman layer is one of most important effects of the layer, especially with regard to the large-scale circulation, for it provides an efficient means whereby surface fluxes are communicated to the interior flow (see Fig. 2.15).

2.12.5 Observations

Ekman layers are generally quite hard to observe, in either ocean or atmosphere, largely because of a signal-to-noise problem — the noise largely coming from inertial and gravity waves (section 2.10) and, especially in the atmosphere, the effects
of stratification and buoyancy-driven turbulence. As regards oceanography, from about 1980 onwards improved instruments have made it possible to observe the vector current with depth, and to average that current and correlate it with the overlying wind, and a number of observations consistent with Ekman dynamics have emerged. The main differences between observations and theory can be ascribed to the effects of stratification (which causes a shallowing and flattening of the spiral), and the interaction of the Ekman spiral with turbulence (and the inadequacy of the eddy-diffusivity parameterization). In spite of these differences of detail, Ekman layer theory remains a remarkable and enduring foundation of geophysical fluid dynamics.

2.12.6 Frictional parameterization

[Some readers will be reading these sections on Ekman layers after having been introduced to quasi-geostrophic theory; this section is for them. Other readers may return to this section after reading chapter 5, or take (2.313) on faith.]

Suppose that the free atmosphere is described by the quasi-geostrophic vorticity equation,

$$\frac{D\zeta_g}{Dt} = f_0 \frac{\partial w}{\partial z}, \quad (2.313)$$

where $\zeta_g$ is the geostrophic relative vorticity. Let us further model the atmosphere as a single homogeneous layer of thickness $H$ lying above an Ekman layer of thickness $d \ll H$. If the vertical velocity is negligible at the top of the layer (at $z = H + d$) the equation of motion becomes

$$\frac{D\zeta_g}{Dt} = f_0 \frac{[w(H + d) - w(d)]}{H} = -\frac{f_0 d}{2H} \zeta_g \quad (2.314)$$

using (2.305). This equation shows that the Ekman layer acts as a linear drag on the interior flow, with a drag coefficient $r$ equal to $f_0 d/2H$ and with associated timescale $T_{Ek}$ given by

$$T_{Ek} = \frac{2H}{f_0 d} = \frac{2H}{\sqrt{2f_0 A}}. \quad (2.315)$$

In the oceanic case the corresponding vorticity equation for the interior flow is

$$\frac{D\zeta_g}{Dt} = \frac{1}{H} \text{curl}_z \tau_s, \quad (2.316)$$

where $\tau_s$ is the surface stress. The surface stress thus acts as if it were a body force on the interior flow, and neither the Coriolis parameter nor the depth of the Ekman layer explicitly appear in this formula.

The Ekman layer is actually a very efficient way of communicating surface stresses to the interior. To see this, suppose that eddy mixing were the sole mechanism of transferring stress from the surface to the fluid interior, and there were no Ekman layer. Then the timescale of spindown of the fluid would be given by using

$$\frac{d\zeta}{dt} = A \frac{\partial^2 \zeta}{\partial z^2}, \quad (2.317)$$
implying a turbulent spindown time, \( T_{\text{turb}} \) of

\[
T_{\text{turb}} \sim \frac{H^2}{A},
\]

(2.318)

where \( H \) is the depth over which we require a spin-down. This is much longer than the spin-down of a fluid that has an Ekman layer, for we have

\[
\frac{T_{\text{turb}}}{T_{\text{Ek}}} = \frac{(H^2/A)}{(2H/f_0d)} = \frac{H}{d} \gg 1,
\]

(2.319)

using \( d = \sqrt{2A/f_0} \). The effects of friction are evidently enhanced because of the presence of a secondary circulation confined to the Ekman layers (as in Fig. 2.15) in which the vertical scales are much smaller than those in the fluid interior and so where viscous effects become significant, and these frictional stresses are then communicated to the fluid interior via the induced vertical velocities at the edge of the Ekman layers.

Notes

1 The distinction between Coriolis force and acceleration has not always been made in the literature. For a fluid in geostrophic balance, one might either say that there is a balance between the pressure force and the Coriolis force, with no net acceleration, or that the pressure force produces a Coriolis acceleration. The descriptions are equivalent, because of Newton’s second law, but should not be conflated.

The Coriolis forces is named after Gaspard Gustave de Coriolis (1792-1843), who introduced the force in the context of rotating mechanical systems [Coriolis, 1832, 1835]. See Persson (1998) for a historical account and interpretation.

2 Phillips (1973). See also Stommel and Moore (1989) and Gill (1982). (There are typographic errors in the second term of each of Gill’s equations (4.12.11) and (4.12.12).)

3 Phillips (1966). See White (2003) for a review. In the early days of numerical modelling these equations were the most primitive — i.e., the least filtered — equations that could practically be integrated numerically. Associated with increasing computer power there is a tendency for comprehensive numerical models to use non-hydrostatic equations of motion that do not make the shallow-fluid or traditional approximations, and it is conceivable that the meaning of the word ‘primitive’ may evolve to accommodate them.

4 The Boussinesq approximation is named for Boussinesq (1903), although similar approximations were used earlier by Oberbeck (1879, 1888). Spiegel and Veronis (1960) give a physically based derivation for an ideal gas, and Mihaljan (1962) provides an asymptotic derivation of the equations. Mahrt (1986) discusses its applicability in the atmosphere.

5 I thank W. R. Young for discussions on this point.

6 Various versions of anelastic equations exist — see Batchelor (1953a), Ogura and Phillips (1962), Gough (1969), Gilman and Glatzmaier (1981), Lipps and Hemler (1982) and Durran (1989) although not all have potential vorticity and energy conservation laws (Bannon 1995, 1996; Scinocca and Shepherd 1992). The system we derive is most similar to that of Ogura and Phillips (1962) and unpublished notes by
J. S. A. Green. The connection between the Boussinesq and anelastic equations is discussed by, among others, Lilly (1996) and Ingersoll (2005).

7 A numerical model that includes sound waves must take very small timesteps in order to maintain numerical stability, in particular to satisfy the CFL criterion. An alternative is to use an implicit time-stepping scheme that effectively lets the numerics do the filtering of the sound waves, and this approach is favoured by many numerical modellers. If we make the hydrostatic approximation then all sound waves except those that propagate horizontally are eliminated, and there is little need, as regards the numerics, to also make the anelastic approximation.

8 It is named for C.-G. Rossby (see endnote on page 243) but was also used by Kibel (1940) and is sometimes called the Kibel or Rossby-Kibel number. The notion of geostrophic balance and so, implicitly, that of a small Rossby number, predates either Rossby or Kibel.

9 After Taylor (1921b) and Proudman (1916). The Taylor-Proudman effect is sometimes called the Taylor-Proudman ‘theorem’, but it is more usefully thought of as a physical effect, with manifestations even when the conditions for its satisfaction are not precisely met.

10 Foster (1972).

11 Many numerical models of the large-scale circulation in the atmosphere and ocean do make the hydrostatic approximation. In these models convection must be parameterized; otherwise, it would simply occur at the smallest scale available, namely the size of the numerical grid, and this type of unphysical behaviour should be avoided. Of course in non-hydrostatic models convection must also be parameterized if the horizontal resolution of the model is too coarse to properly resolve the convective scales. See also problem 2.10.

12 After Ekman (1905). The problem was posed to Ekman, a student of Vilhelm Bjerknes, by Fridtjof Nansen, the polar explorer and statesman, who wanted to understand the motion of icebergs.


Further Reading

This book provides a compact introduction to a variety of topics in GFD.

A rich book, especially strong on equatorial dynamics and gravity wave motion.

A deservedly well-known textbook at the upper-division undergraduate/beginning graduate level.

A primary reference, especially for flow at low Rossby number. Although the book requires some effort, there is a handsome pay-off for those who study it closely.

White (2002) provides a clear and thorough summary of the equations of motion for meteorology, including the non-hydrostatic and primitive equations.

Concentrates on the equations of motion and the mathematical tools needed for a fundamental understanding.
Problems

2.1 For an ideal gas in hydrostatic balance, show that:

(a) The integral of the potential plus internal energy from the surface to the top of the atmosphere \((p = 0)\) is equal to its enthalpy;

(b) \(d\sigma/dz = c_p(T/\theta)d\theta/dz\), where \(\sigma = I + p\alpha + \Phi\) is the dry static energy;

(c) The following expressions for the pressure gradient force are all equal (even without hydrostatic balance):

\[-\frac{1}{\rho}\nabla p = -\frac{\theta}{\rho}\nabla \Pi = -\frac{c_p^2}{\rho\theta} \nabla (\rho\theta).\]  

(P2.1)

where \(\Pi = c_p T/\theta\) is the Exner function.

(d) Show that item (a) also holds for a gas with an arbitrary equation of state, \(p = p(\rho, T)\).

2.2 Show that, without approximation, the unforced, inviscid momentum equation may be written in the forms

\[\frac{Dv}{Dt} = T \nabla \eta - \nabla (p\alpha + I)\]  

(P2.2)

and

\[\frac{\partial v}{\partial t} + \omega \times v = T \nabla \eta - \nabla B\]  

(P2.3)

where \(\omega = \nabla \times v\), \(\eta\) is the specific entropy \((d\eta = c_p d\ln \theta)\) and \(B = I + v^2/2 + p\alpha\) where \(I\) is the internal energy per unit mass.

Hint: First show that \(T \nabla \eta = \nabla I + p \nabla \alpha\), and note also the vector identity \(v \times (\nabla \times v) = \frac{1}{2} \nabla (v \cdot v) - (v \cdot \nabla)v\).

2.3 Consider two-dimensional fluid flow in a rotating frame of reference on the \(f\)-plane. Linearize the equations about a state of rest.

(a) Ignore the pressure term and determine the general solution to the resulting equations. Show that the speed of fluid parcels is constant. Show that the trajectory of the fluid parcels is a circle with radius \(|U|/f\), where \(|U|\) is the fluid speed.

(b) What is the period of oscillation of a fluid parcel?

(c) ♦ If parcels travel in straight lines in inertial frames, why is the answer to (b) not the same as the period of rotation of the frame of reference? [To answer this fully you need to understand the dynamics underlying inertial oscillations and inertia circles. See Durran (1993), Egger (1999) and Phillips (2000).]

2.4 A fluid at rest evidently satisfies the hydrostatic relation, which says that the pressure at the surface is given by the weight of the fluid above it. Now consider a deep atmosphere on a spherical planet. A unit cross-sectional area at the planet’s surface supports a column of fluid whose cross-section increases with height, because the total area of the atmosphere increases with distance away from the center of the planet. Is the pressure at the surface still given by the hydrostatic relation, or is it greater than this because of the increased mass of fluid in the column? If it is still given by the hydrostatic relation, then the pressure at the surface, integrated over the entire area of the planet, is less than the total weight of the fluid; resolve this paradox. But if the pressure at the surface is greater than that implied by hydrostatic balance, explain how the hydrostatic relation fails.

2.5 By considering how the direction of the coordinate axes change with position [as in Holton (1992), for example] show geometrically that in spherical coordinates:

\[\frac{Di}{Dt} = u \partial i/\partial x + v \partial i/\partial y = (u/r \cos \theta)(j \sin \theta - k \cos \theta),\]  

(P2.4)

\[\frac{Dj}{Dt} = u \partial j/\partial x + v \partial j/\partial y = -i(u/r) \tan \theta - kv/a,\]  

(P2.5)
\[
\frac{Dk}{Dt} = u \frac{\partial k}{\partial x} + v \frac{\partial k}{\partial y} = i(u/r) + j(v/r).
\] (P2.6)

Then, using (2.44a) show that (2.45) results.

2.6 At what latitude is the angle between the direction of Newtonian gravity (due solely to the mass of the earth) and that of effective gravity (Newtonian gravity plus centrifugal terms) the largest? At what latitudes, if any, is this angle zero?

2.7 ♦ Write the momentum equations in true spherical coordinates, including the centrifugal and gravitational terms. Show that for reasonable values of the wind, the dominant balance in the meridional component of this equation involve a balance between centrifugal and pressure gradient terms. Can this balance be subtracted out of the equations in a sensible way, so leaving a useful horizontal momentum equation that involves the Coriolis and acceleration terms? If so, obtain a closed set of equations for the flow this way. Discuss the pros and cons of this approach versus the geometric approximation discussed in section 2.2.1.

2.8 For an ideal gas show that the expressions (2.222) and (2.226) are equivalent.

2.9 Consider an ocean at rest with known vertical profiles of potential temperature and salinity, \( \theta(z) \) and \( S(z) \). Suppose we also know the equation of state in the form \( \rho = \rho(\theta, S, p) \). Obtain an expression for the buoyancy frequency. Check your expression by substituting the equation of state for an ideal gas and recovering a known expression for the buoyancy frequency.

2.10 Convection and its parameterization

(a) Consider a Boussinesq system in which the vertical momentum equation is modified by the parameter \( \alpha \) to read

\[
\alpha^2 \frac{Dw}{Dt} = -\frac{\partial \phi}{\partial z} + b,
\] (P2.7)

and the other equations are unchanged. (If \( \alpha = 0 \) the system is hydrostatic, and if \( \alpha = 1 \) the system is the original one.) Linearise these equations about a state of rest and of constant stratification (as in section 2.10.1) and obtain the dispersion relation for the system, and plot it for various values of \( \alpha \), including 0 and 1. Show that, for \( \alpha > 1 \) the system approaches its limiting frequency more rapidly than with \( \alpha = 1 \).

(b) ♦ Argue that if \( N^2 < 0 \), convection in a system with \( \alpha > 1 \) generally occurs at a larger scale than with \( \alpha = 1 \). Show this explicitly by adding some diffusion or friction to the right-hand sides of the equations of motion and obtaining the dispersion relation. You may do this approximately.

2.11 (a) The geopotential height is the height of a given pressure level. Show that in an atmosphere with a uniform lapse rate (i.e., \( dT/dz = \Gamma = \text{constant} \)) the geopotential height at a pressure \( p \) is given by

\[
z = \frac{T_0}{T} \left[ 1 - \left(\frac{P_0}{p}\right)^{-\Gamma/g} \right]
\] (P2.8)

where \( T_0 \) is the temperature at \( z = 0 \).

(b) In an isothermal atmosphere, obtain an expression for the geopotential height as a function of pressure, and show that this is consistent with the expression (P2.8) in the appropriate limit.

2.12 Consider the simple Boussinesq equations, \( Dv/ Dt = kb + \nu \nabla^2 v \), \( \nabla \cdot v = 0 \), \( Db/ Dt = Q + \kappa \nabla^2 b \). Obtain an energy equation similar to (2.112) but now with the terms on the right-hand side that represent viscous and diabatic effects. Over a closed volume,
show that the dissipation of kinetic energy is balanced by a buoyancy source. Show also that, in a statistically steady state, the heating must occur at a lower level than the cooling if a kinetic-energy dissipating circulation is to be maintained.

2.13 ♦ Suppose a fluid is contained in a closed container, with insulating sidewalls, and heated from below and cooled from above. The heating and cooling are adjusted so that there is no net energy flux into the fluid. Let us also suppose that any viscous dissipation of kinetic energy is returned as heating, so the total energy of the fluid is exactly constant. Suppose the fluid starts out at rest and at a uniform temperature, and the heating and cooling are then turned on. A very short time afterwards, the fluid is lighter at the bottom and heavier at the top; that is, its potential energy has increased. Where has this energy come from? Discuss this paradox for both a compressible fluid (e.g., an ideal gas) and for a simple Boussinesq fluid.

2.14 Consider a rapidly rotating (i.e., in near geostrophic balance) Boussinesq fluid on the $f$-plane.

(a) Show that the pressure divided by the density scales as $\phi \sim fUL$.

(b) Show that the horizontal divergence of the geostrophic wind vanishes. Thus, argue that the scaling $W \sim UH/L$ is an overestimate for the magnitude of the vertical velocity. (Optional extra: obtain a scaling estimate for the magnitude of vertical velocity in rapidly rotating flow.)

(c) Using these results, or otherwise, discuss whether hydrostatic balance is more or less likely to hold in a rotating flow that in non-rotating flow.

2.15 Estimate the size of the zonal wind 5 km above the surface in the midlatitude atmosphere in summer and winter using (approximate) values for the meridional temperature gradient in the atmosphere. Also estimate the shear corresponding to the pole-equator temperature gradient in the ocean.

2.16 Using approximate but realistic values for the observed stratification, what is the buoyancy period for (a) the mid-latitude troposphere, (b) the stratosphere, (c) the oceanic thermocline, (d) the oceanic abyss?

2.17 Consider a dry, hydrostatic, ideal-gas atmosphere whose lapse rate is one of constant potential temperature. What is its vertical extent? That is, at what height does the density vanish? Is this a problem for the anelastic approximation discussed in the text?

2.18 Show that for an ideal gas, the expressions (2.226), (2.221), (2.222) are all equivalent, and express $N^2$ terms of the temperature lapse rate, $\partial T/\partial z$.

2.19 ♦ Calculate a reasonably accurate, albeit approximate, expression for the buoyancy equation for seawater. (Derived from notes by R. deSzoeke)

Solution (i): The buoyancy frequency is given by

$$N^2 = \frac{g}{\alpha} \left( \frac{\partial \rho_{\text{pot}}}{\partial z} \right)_{\text{env}} = \frac{g}{\alpha} \left( \frac{\partial \alpha_{\text{pot}}}{\partial z} \right)_{\text{env}} = -\frac{g^2}{\alpha^2} \left( \frac{\partial \alpha_{\text{pot}}}{\partial p} \right)_{\text{env}} \quad (P2.9)$$

where $\alpha_{\text{pot}} = \alpha(\theta, S, p_R)$ is the potential density, and $p_R$ a reference pressure. From (1.173)

$$\alpha_{\text{pot}} = \alpha_0 \left[ 1 - \frac{\alpha_0}{c_0^2} p_R + \beta_T (1 + \gamma p_R) \theta' + \frac{1}{2} \beta_T^2 \theta'^2 - \beta_S (S - S_0) \right]. \quad (P2.10)$$

Using this and (P2.9) we obtain the buoyancy frequency,

$$N^2 = -\frac{g^2}{\alpha^2} \alpha_0 \left[ \beta_T \left( 1 + \gamma p_R + \beta_T^2 \theta' \right) \left( \frac{\partial \theta}{\partial p} \right)_{\text{env}} - \beta_S \left( \frac{\partial S}{\partial p} \right)_{\text{env}} \right], \quad (P2.11)$$
although we must substitute local pressure for the reference pressure $p_R$. (Why?)

**Solution (ii):** The sound speed is given by

$$c_s^{-2} = -\frac{1}{\alpha^2} \left( \frac{\partial \alpha}{\partial p} \right)_{0S} = \frac{1}{\alpha^2} \left( \frac{\alpha^2}{c_s^2} - \gamma \alpha \theta \right)$$  \hspace{1cm} (P2.12)

and, using (P2.9) and (2.226) the square of the buoyancy frequency may be written

$$N^2 = \frac{g}{\alpha} \left( \frac{\partial \alpha}{\partial z} \right)_{env} - \frac{g^2}{\alpha^2} \left[ \left( \frac{\partial \alpha}{\partial p} \right)_{env} + \frac{\alpha^2}{c_s^2} \right]$$  \hspace{1cm} (P2.13)

Using (1.173), (P2.12) and (P2.13) we recover (P2.11), although now with $p$ explicitly in place of $p_R$.

**2.20** Begin with the mass conservation in the height-coordinates, namely $D\rho/Dt + \rho \nabla \cdot \mathbf{v} = 0$. Transform this into pressure coordinates using the chain rule (or otherwise) and derive the mass conservation equation in the form $\nabla p \cdot \mathbf{u} + \partial \omega/\partial p = 0$.

**2.21** Starting with the primitive equations in pressure coordinates, derive the form of the primitive equations of motion in sigma-pressure coordinates. In particular, show that the prognostic equation for surface pressure is,

$$\frac{\partial p_s}{\partial t} + \nabla \cdot (p_s \mathbf{u}) + p_s \frac{\partial \sigma}{\partial \sigma} = 0$$  \hspace{1cm} (P2.14)

and that hydrostatic balance may be written $\partial \Phi/\partial \sigma = -RT/\sigma$.

**2.22** Starting with the primitive equations in pressure coordinates, derive the form of the primitive equations of motion in log-pressure coordinates in which $Z = -H \ln(p/p_r)$ is the vertical coordinate. Here, $H$ is a reference height (e.g., a scale height $RT_r/g$ where $T_r$ is a typical or an average temperature) and $p_r$ is a reference pressure (e.g., 1000 mb). In particular, show that if the ‘vertical velocity’ is $W = DZ/Dt$ then

$$W = -H \omega/p$$

and obtain the mass conservation equation (2.156b). Show that this can be written in the form

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{1}{\rho_s} \frac{\partial}{\partial z} (\rho_s W) = 0,$$  \hspace{1cm} (P2.16)

where $\rho_s = \rho_r \exp(-Z/H)$.

**2.23 (a)** Prove that the argument of the square root in (2.25) is always positive.

**Solution:** The largest value or the argument occurs when $m = 0$ and $k^2 = 1/(4H^2)$. The argument is then $1 - 4H^2 N^2/c_s^2$. But $c_s^2 = \gamma RT_0 = \gamma gH$ and $N^2 = g\kappa/H$ so that $4N^2 H^2/c_s^2 = 4\kappa/\gamma \approx 0.8$.

**(b)** ♦ This argument seems to depend on the parameters in the ideal gas equation of state. Is it more general than this? Is a natural system possible for which the argument is negative, and if so what physical interpretation could one ascribe to the situation?

**2.24** Consider a wind stress imposed by a mesoscale cyclonic storm (in the atmosphere) given by

$$\mathbf{\tau} = -Ae^{-r/(\lambda^2)}(r i + x j)$$  \hspace{1cm} (P2.17)

where $r^2 = x^2 + y^2$, and $A$ and $\lambda$ are constants. Also assume constant Coriolis gradient $\beta = \partial f/\partial y$ and constant ocean depth $H$. Find (a) the Ekman transport, (b) the vertical velocity $w_E(x,y,z)$ below the Ekman layer, (c) the northward velocity $v(x,y,z)$ below the Ekman layer and (d) indicate how you would find the westward velocity $u(x,y,z)$ below the Ekman layer.
2.25 In an atmospheric Ekman layer on the $f$-plane for a fluid with $\rho = \rho_a = 1$ let us write the momentum equation as

$$
\mathbf{f} \times \mathbf{u} = -\nabla \phi + \frac{\partial \tau}{\partial z}
$$

(P2.18)

where $\tau = K \frac{\partial \mathbf{u}}{\partial z}$ and $K$ is a constant coefficient of viscosity. An independent formula for the stress at the ground is $\tau = C \mathbf{u}$, where $C$ is a constant. Assume that in the free atmosphere the wind is geostrophic and zonal, with $\mathbf{u}_g = U \mathbf{i}$.

(a) Find an expression for the wind vector at the ground. Discuss the limits $C = 0$ and $C = \infty$. Show that when $C = 0$ the frictionally-induced vertical velocity at the top of the Ekman layer is zero.

(b) Find the vertically integrated horizontal mass flux caused by the boundary layer.

(c) When the stress on the atmosphere is $\tau$, the stress on the ocean beneath is $-\tau$. Determine the direction and strength of the surface current in terms of the surface wind, the oceanic Ekman depth and the ratio $\rho_a/\rho_o$, where $\rho_o$ is the density of the seawater. How does the boundary-layer mass flux in the ocean compare to that in the atmosphere?

Partial solution: A useful trick in Ekman layer problems is to write the velocity as a complex number, $\tilde{\mathbf{u}} = \mathbf{u} + i \mathbf{v}$. The Ekman layer equation, (2.294a), may then be written as

$$
A \frac{\partial^2 \tilde{U}}{\partial z^2} = i f \tilde{U},
$$

(P2.19)

where $\tilde{U} = \tilde{\mathbf{u}} - \tilde{\mathbf{u}}_g$. The solution to this is

$$
\tilde{U} - \tilde{U}_g = (\tilde{U}(0) - \tilde{U}_g) \exp\left[-\frac{(1 + i)z}{4d}\right],
$$

(P2.20)

where the boundary condition of finiteness at infinity eliminates the exponentially growing solution. The boundary condition at $z = 0$ is $\frac{\partial \mathbf{u}_g}{\partial z} = \frac{(C/K) \mathbf{u}_g}{2K}$ which gives $(\tilde{U}(0) - \tilde{U}_g) \exp(i\pi/4) = -Cd/(\sqrt{2K})\tilde{U}(0)$, and the rest of the solution follows.

2.26 The logarithmic boundary layer

Close to ground rotational effects are unimportant and small-scale turbulence generates a mixed layer. In this layer, assume that the stress is constant and that it can be parameterized by an eddy diffusivity the size of which is proportional to the distance from the surface. Show that the velocity then varies logarithmically with height.

Solution: Write the stress as $\tau = \rho_0 u^* \frac{\partial \mathbf{u}}{\partial z}$ where the constant $u^*$ is called the frictional velocity. Using the eddy diffusivity hypothesis this stress is given by

$$
\tau = \rho_0 u^* \frac{\partial u}{\partial z} = \rho_0 A \frac{\partial u}{\partial z}
$$

(P2.21)

where $A$ is von Karman’s (‘universal’) constant (approximately equal to 0.4). From (P2.21) we have $\frac{\partial u}{\partial z} = u^*/(Az)$ which integrates to give $u = (u^*/k) \ln(z/z_0)$. The parameter $z_0$ is known as the roughness length, and is typically of order centimeters or a little larger, depending on the surface.
Another advantage of a mathematical statement is that it is so definite that it might be definitely wrong... Some verbal statements have not this merit.

L. F. Richardson (1881–1953).

CHAPTER THREE

Shallow Water Systems and Isentropic Coordinates

Conventionally, 'the' shallow water equations describe a thin layer of constant density fluid in hydrostatic balance, rotating or not, bounded from below by a rigid surface and from above by a free surface, above which we suppose is another fluid of negligible inertia. Such a configuration can be generalized to multiple layers of immiscible fluids lying one on top of each other, forming a 'stacked shallow water' system, and this class of systems is the main subject of this chapter.

The single-layer model is one of the simplest useful models in geophysical fluid dynamics, because it allows for a consideration of the effects of rotation in a simple framework without with the complicating effects of stratification. By adding layers we can then study the effects of stratification, and indeed the model with just two layers is not only a simple model of a stratified fluid, it is a surprisingly good model of many phenomena in the ocean and atmosphere. Indeed, the models are more than just pedagogical tools — we will find that there is a close physical and mathematical analogy between the shallow water equations and a description of the continuously stratified ocean or atmosphere written in isopycnal or isentropic coordinates, with a meaning beyond a coincidental similarity in the equations. We begin with the single-layer case.

3.1 DYNAMICS OF A SINGLE, SHALLOW LAYER

Shallow water dynamics apply, by definition, to a fluid layer of constant density in which the horizontal scale of the flow is much greater than the layer depth. The fluid motion is then fully determined by the momentum and mass continuity
Fig. 3.1 A shallow water system. \( h \) is the thickness of a water column, \( H \) its mean thickness, \( \eta \) the height of the free surface and \( \eta_b \) is the height of the lower, rigid, surface, above some arbitrary origin, typically chosen such that the average of \( \eta_b \) is zero. \( \Delta \eta \) is the deviation free surface height, so we have \( \eta = \eta_b + h = H + \Delta \eta \).

equations, and because of the assumed small aspect ratio the the hydrostatic approximation is well satisfied, and we invoke this from the outset. Consider, then, fluid in a container above which is another fluid of negligible density (and therefore negligible inertia) relative to the fluid of interest, as illustrated in Fig. 3.1. As usual, our notation is that \( v = u \hat{i} + v \hat{j} + w \hat{k} \) is the three dimensional velocity and \( u = u \hat{i} + v \hat{j} \) is the horizontal velocity. \( h(x, y) \) is thickness of the liquid column, \( H \) is its mean height, and \( \eta \) is the height of the free surface. In a flat-bottomed container \( \eta = h \), whereas in general \( h = \eta - \eta_b \), where \( \eta_b \) is the height of the floor of the container.

3.1.1 Momentum equations

The vertical momentum equation is just the hydrostatic equation,

\[
\frac{\partial p}{\partial z} = -\rho g, \tag{3.1}
\]

and, because density is assumed constant, we may integrate this to

\[
p(x, y, z) = -\rho g z + p_o. \tag{3.2}
\]

At the top of the fluid, \( z = \eta \), the pressure is determined by the weight of the overlying fluid and this is assumed negligible. Thus, \( p = 0 \) at \( z = \eta \) giving

\[
p(x, y, z) = \rho g (\eta(x, y) - z). \tag{3.3}
\]

The consequence of this is that the horizontal gradient of pressure is independent of height. That is

\[
\nabla_z p = \rho g \nabla z \eta, \tag{3.4}
\]

where

\[
\nabla_z = \frac{\partial}{\partial x} + \frac{\partial}{\partial y} \tag{3.5}
\]
is the gradient operator at constant $z$. (In the rest of this chapter we will drop the subscript $z$ unless that causes ambiguity. The three-dimensional gradient operator will be denoted $\nabla_3$. We will also mostly use Cartesian coordinates, but the shallow water equations may certainly be applied over a spherical planet — indeed, ‘Laplace’s tidal equations’ are essentially the shallow water equations on a sphere.)

The horizontal momentum equations therefore become

$$\frac{Du}{Dt} = -\frac{1}{\rho} \nabla p = -g \nabla \eta. \tag{3.6}$$

The right-hand side of this equation is independent of the vertical coordinate $z$. Thus, if the flow is initially independent of $z$, it must stay so. (This $z$-independence is unrelated to that arising from the rapid rotation necessary for the Taylor-Proudman effect.) The velocities $u$ and $v$ are functions only of $x, y$ and $t$ and the horizontal momentum equation is therefore

$$\frac{Du}{Dt} = \frac{\partial u}{\partial t} + u \frac{\partial u}{\partial x} + v \frac{\partial u}{\partial y} = -g \nabla \eta. \tag{3.7}$$

That the horizontal velocity is independent of $z$ is a consequence of the hydrostatic equation, which ensures that the horizontal pressure gradient is independent of height. (Another starting point would be to take this independence of the horizontal motion with height as the definition of shallow water flow. In real physical situations such independence does not hold exactly — for example, friction at the bottom may induce a vertical dependence of the flow in a boundary layer.) In the presence of rotation (3.7) easily generalizes to

$$\frac{Du}{Dt} + f \times u = -g \nabla \eta, \tag{3.8}$$

where $f = f k$. Just as with the primitive equations, $f$ may be constant or may vary with latitude, so that on a spherical planet $f = 2\Omega \sin \vartheta$ and on the $\beta$-plane $f = f_0 + \beta y$.

### 3.1.2 Mass continuity equation

*From first principles*

The mass contained in a fluid column of height $h$ and cross-sectional area $A$ is given by $\int_A \rho h \, dA$ (see Fig. [3.2]). If there is a net flux of fluid across the column boundary (by advection) then this must be balanced by a net increase in the mass in $A$, and therefore a net increase in the height of the water column. The mass convergence into the column is given by

$$F_m = \text{Mass flux in} = -\int_S \rho u \cdot ds, \tag{3.9}$$

where $S$ is the area of the vertical boundary of the column. The surface area of the column is comprised of elements of area $hn\delta l$, where $\delta l$ is a line element circumscribing the column and $n$ is a unit vector perpendicular to the boundary, pointing outwards. Thus (3.9) becomes

$$F_m = -\int \rho hu \cdot n \, dl. \tag{3.10}$$
Using the divergence theorem in two-dimensions, (3.10) simplifies to

\[ F_m = -\int_A \nabla \cdot (\rho u h) \, dA, \]  

(3.11)

where the integral is over the cross-sectional area of the fluid column (looking down from above). This is balanced by the local increase in height of the water column, given by

\[ F_m = \frac{d}{dt} \int_A \rho h \, dV = \int_A \rho h \, dA = \int_A \rho \frac{\partial h}{\partial t} \, dA. \]  

(3.12)

Because \( \rho \) is constant, the balance between (3.11) and (3.12) leads to

\[ \int_A \left( \frac{\partial h}{\partial t} + \nabla \cdot (uh) \right) \, dA = 0, \]  

(3.13)

and because the area is arbitrary the integrand itself must vanish, whence,

\[ \frac{\partial h}{\partial t} + \nabla \cdot (uh) = 0, \]  

(3.14)

or equivalently

\[ \frac{Dh}{Dt} + h \nabla \cdot u = 0. \]  

(3.15)

This derivation holds whether or not the lower surface is flat. If it is, then \( h = \eta \), and if not \( h = \eta - b \). Eqs. (3.8) and (3.14) or (3.15) form a complete set, summarized in the shaded box on the next page.

**From the 3D mass conservation equation**

Since the fluid is incompressible, the three-dimensional mass continuity equation is just \( \nabla \cdot \boldsymbol{v} = 0 \). Writing this out in component form

\[ \frac{\partial w}{\partial z} = - \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) = - \nabla \cdot \boldsymbol{u} \]  

(3.16)
The Shallow Water Equations

For a single-layer fluid, and including the Coriolis term, the inviscid shallow water equations are:

Momentum:

\[
\frac{Du}{Dt} + f \times u = -g \nabla \eta. \tag{SW.1}
\]

Mass Continuity:

\[
\frac{Dh}{Dt} + h \nabla \cdot u = 0 \quad \text{or} \quad \frac{\partial h}{\partial t} + \nabla \cdot (hu) = 0. \tag{SW.2}
\]

where \( u \) is the horizontal velocity, \( h \) is the total fluid thickness, \( \eta \) is the height of the upper free surface and \( \eta_b \) is the height of the lower surface (the bottom topography). Thus, \( h(x, y, t) = \eta(x, y, t) - \eta_b(x, y) \). The material derivative is

\[
\frac{D}{Dt} = \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y}, \tag{SW.3}
\]

with the rightmost expression holding in Cartesian coordinates.

Integrate this from the bottom of the fluid \( (z = \eta_b) \) to the top \( (z = \eta) \), noting that the right-hand side is independent of \( z \), to give

\[
w(\eta) - w(\eta_b) = -h \nabla \cdot u. \tag{3.17}
\]

At the top the vertical velocity is the material derivative of the position of a particular fluid element. But the position of the fluid at the top is just \( \eta \), and therefore (see Fig. 3.2)

\[
w(\eta) = D\eta \tag{3.18a}
\]

At the bottom of the fluid we have similarly

\[
w(\eta_b) = D\eta_b \tag{3.18b}
\]

where, absent earthquakes and the like, \( \partial \eta_b / \partial t = 0 \). Using (3.18a,b), (3.17) becomes

\[
\frac{D}{Dt} (\eta - \eta_b) + h \nabla \cdot u = 0 \tag{3.19}
\]

or, as in (3.15),

\[
\frac{Dh}{Dt} + h \nabla \cdot u = 0. \tag{3.20}
\]

3.1.3 A rigid lid

The case where the upper surface is held flat by the imposition of a rigid lid is sometimes of interest. The ocean suggests one such example, for here the bathymetry
at the bottom of the ocean provides much larger variations in fluid thickness than do the small variations in the height of the ocean surface. Suppose then the upper surface is at a constant height $H$ then, from (3.14) with $\partial h/\partial t = 0$ the mass conservation equation becomes

$$\nabla_h \cdot (uh_h) = 0.$$  

(3.21)

where $h_h = H - \eta_h$ Note that this allows us to define an incompressible mass-transport velocity, $U \equiv h_h u$.

Although the upper surface is flat, the pressure there is no longer constant because a force must be provided by the rigid lid to keep the surface flat. The horizontal momentum equation is

$$\frac{D u}{Dt} = -\frac{1}{\rho_0} \nabla p_{\text{lid}},$$  

(3.22)

where $p_{\text{lid}}$ is the pressure at the lid, and the complete equations of motion are then (3.21) and (3.22). If the lower surface is flat, the two-dimensional flow itself is divergence-free, and the equations reduce to the two-dimensional incompressible Euler equations.

### 3.1.4 Stretching and the vertical velocity

Because the horizontal velocity is depth independent, the vertical velocity plays no role in advection. However, $w$ is certainly not zero for then the free surface would be unable to move up or down, but because of the vertical independence of the horizontal flow $w$ does have a simple vertical structure; to determine this we write the mass conservation equation as

$$\frac{\partial w}{\partial z} = -\nabla \cdot u$$  

(3.23)

and integrate upwards from the bottom to give

$$w = w_h - (\nabla \cdot u)(z - \eta_h).$$  

(3.24)

Thus, the vertical velocity is a linear function of height. Eq. (3.24) can be written

$$\frac{Dz}{Dt} = \frac{D\eta_h}{Dt} - (\nabla \cdot u)(z - \eta_h),$$  

(3.25)

and at the upper surface $w = D\eta_h/DT$ so that here we have

$$\frac{D\eta}{DT} = \frac{D\eta_h}{DT} - (\nabla \cdot u)(\eta - \eta_h),$$  

(3.26)

Eliminating the divergence term from the last two equations gives

$$\frac{D}{DT}(z - \eta_h) = \frac{z - \eta_h}{\eta - \eta_h} \frac{D}{DT}(\eta - \eta_h),$$  

(3.27)

which in turn gives

$$\frac{D}{DT} \left( \frac{z - \eta_h}{h} \right) = \frac{D}{DT} \left( \frac{z - \eta_h}{h} \right) = 0.$$  

(3.28)

This means that the ratio of the height of a fluid parcel above the floor to the total depth of the column is fixed; that is, the fluid stretches uniformly in a column, and this is kinematic property of the shallow water system.
3.1.5 Analogy with compressible flow

The shallow water equations (3.8) and (3.14) are analogous to the compressible gas dynamic equations in two dimensions, namely

$$\frac{Du}{Dt} = -\frac{1}{\rho} \nabla p \quad (3.29)$$

and

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (u \rho) = 0, \quad (3.30)$$

along with an equation of state which we take to be $p = f(\rho)$. The mass conservation equations (3.14) and (3.30) are identical, with the replacement $\rho \leftrightarrow h$. If $p = C \rho^\gamma$, then (3.29) becomes

$$\frac{Du}{Dt} = -\frac{1}{\rho} \frac{dp}{d\rho} \nabla \rho = -C \gamma \rho^{\gamma-2} \nabla \rho. \quad (3.31)$$

If $\gamma = 2$ then the momentum equations (3.8) and (3.31) become equivalent, with $\rho \leftrightarrow h$ and $C \gamma \rightarrow g$. In an ideal gas $\gamma = c_p/c_v$ and values typically are in fact less than two (in air $\gamma \approx 7/5$); however, if the equations are linearized, then the analogy is exact for all values of $\gamma$, for then (3.31) becomes $\frac{\partial v'/\partial t} = -\rho_0^{-1} c_s^2 \nabla \rho'$ where $c_s^2 = dp/d\rho$, and the linearized shallow water momentum equation is $\frac{\partial u'/\partial t} = -H^{-1}(gH) \nabla h'$, so that $\rho_0 \rightarrow H$ and $c_s^2 \rightarrow gH$. The sound waves of a compressible fluid are then analogous to shallow water waves, considered in section 3.7.

3.2 REDUCED GRAVITY EQUATIONS

Consider now a single shallow moving layer of fluid on top a deep, quiescent fluid layer (Fig. 3.3), and beneath a fluid of negligible inertia. This configuration is often used a model of the upper ocean: the upper layer represents flow in perhaps the upper few hundred meters of the ocean, the lower layer the near-stagnant abyss. If we turn the model upside-down we have a model, perhaps slightly less realistic, of the atmosphere: the lower layer represents motion in the troposphere above which lies an inactive stratosphere. The equations of motion are virtually the same in both cases.

![Figure 3.3 The reduced gravity shallow water system. An active layer lies over a deep, more dense, quiescent layer. In a common variation the upper surface is held flat by a rigid lid, and $\eta_0 = 0$.](image-url)
3.2.1 Pressure gradient in the active layer
We’ll derive the equations for the oceanic case (active layer on top) in two cases, which differ slightly in the assumption made about the upper surface.

I Free upper surface
The pressure in the upper layer is given by integrating the hydrostatic equation down from the upper surface. Thus, at a height $z$ in the upper layer

$$p_1(z) = g \rho_1 (\eta_0 - z),$$  \hspace{1cm} (3.32)

where $\eta_0$ is the height of the upper surface. Hence, everywhere in the upper layer,

$$\frac{1}{\rho_1} \nabla p_1 = -g \nabla \eta_0,$$  \hspace{1cm} (3.33)

and the momentum equation is

$$\frac{D u}{Dt} + f \times u = -g \nabla \eta_0.$$  \hspace{1cm} (3.34)

In the lower layer the the pressure is also given by the weight of the fluid above it. Thus, at some level $z$ in the lower layer,

$$p_2(z) = \rho_1 g (\eta_0 - \eta_1) + \rho_2 g (\eta_1 - z).$$  \hspace{1cm} (3.35)

But if this layer is motionless the horizontal pressure gradient in it is zero and therefore

$$\rho_1 g \eta_0 = -\rho_1 g' \eta_1 + \text{constant},$$  \hspace{1cm} (3.36)

where $g' = g(\rho_2 - \rho_1)/\rho_1$ is the reduced gravity. The momentum equation becomes

$$\frac{D u}{Dt} + f \times u = g' \nabla \eta_1.$$  \hspace{1cm} (3.37)

The equations are completed by the usual mass conservation equation,

$$\frac{D h}{Dt} + h \nabla \cdot u = 0,$$  \hspace{1cm} (3.38)

where $h = \eta_0 - \eta_1$. Because $g \gg g'$, (3.36) shows that surface displacements are much smaller than the displacements at the interior interface. We see this in the real ocean where the mean interior isopycnal displacements may be several tens of meters but variations in the mean height of ocean surface are of order centimeters.

II The rigid lid approximation
The smallness of the upper surface displacement suggests that we will make little error is we impose a rigid lid at the top of the fluid. Displacements are no longer allowed, but the lid will in general impart a pressure force to the fluid. Suppose that this is $P(x, y, t)$, then the horizontal pressure gradient in the upper layer is simply

$$\nabla p_1 = \nabla P.$$  \hspace{1cm} (3.39)
The pressure in the lower layer is again given by hydrostasy, and is

\[ p_2 = -\rho_1 g \eta_1 + \rho_2 g (\eta_1 - z) + P \]

\[ = \rho_1 g h - \rho_2 g (h + z) + P, \tag{3.40} \]

so that

\[ \nabla p_2 = -g (\rho_2 - \rho_1) \nabla h + \nabla P. \tag{3.41} \]

Then if \( \nabla p_2 = 0 \) we have

\[ g (\rho_2 - \rho_1) \nabla h = \nabla P, \tag{3.42} \]

and the momentum equation for the upper layer is just

\[ \frac{D u}{D t} + f \times u = -g'_1 \nabla h. \tag{3.43} \]

where \( g' = g (\rho_2 - \rho_1) / \rho_1 \). These equations differ from the usual shallow water equations only in the use of a reduced gravity \( g' \) in place of \( g \) itself. It is the density difference between the two layers that is important. Similarly, if we take a shallow water system, with the moving layer on the bottom, and we suppose that overlying it is a stationary fluid of finite density, then we would easily find that the fluid equations for the moving layer are the same as if the fluid on top had zero inertia, except that \( g \) would be replaced by an appropriate reduced gravity (problem 3.1).

**3.3 MULTI-LAYER SHALLOW WATER EQUATIONS**

We now consider the dynamics of multiple layers of fluid stacked on top of each other. This is a crude representation of continuous stratification, but it turns out to be a powerful model of many geophysically interesting phenomena as well as being physically realizable in the laboratory. The pressure is continuous across the interface, but the density jumps discontinuously and this allows the horizontal velocity to have a corresponding discontinuity. The set up is illustrated in Fig. 3.4.

In each layer pressure is given by the hydrostatic approximation, and so anywhere in the interior we can find the pressure by integrating down from the top. Thus, at a height \( z \) in the first layer we have

\[ p_1 = \rho_1 g (\eta_0 - z), \tag{3.44} \]

and in the second layer,

\[ p_2 = \rho_1 g (\eta_0 - \eta_1) + \rho_2 g (\eta_1 - z) = \rho_1 g \eta_0 + \rho_1 g'_1 \eta_1 - \rho_2 g z, \tag{3.45} \]

where \( g'_1 = g (\rho_2 - \rho_1) / \rho_1 \), and so on. The term involving \( z \) is irrelevant for the dynamics, because only the horizontal derivative enters the equation of motion. Omitting this term, for the \( n \)'th layer the dynamical pressure is given by the sum from the top down:

\[ p_n = \rho_1 \sum_{i=0}^{n-1} g'_i \eta_i, \tag{3.46} \]
Figure 3.4 The multi-layer shallow water system. The layers are numbered from the top down. The co-ordinates of the interfaces are denoted $\eta_i$ and the layer thicknesses $h_i$, so that $h_i = \eta_i - \eta_{i-1}$.

where $g'_i = g (\rho_{i+1} - \rho_i) / \rho_1$ (but $g_0 = g$). The interface displacements may be expressed in terms of the layer thicknesses by summing from the bottom up:

$$\eta_n = \eta_b + \sum_{i=n+1}^{i=N} h_i.$$  \hspace{1cm} (3.47)

The momentum equation for each layer may then be written, in general,

$$\frac{Du_n}{Dt} + f \times u_n = -\frac{1}{\rho_n} \nabla p_n,$$  \hspace{1cm} (3.48)

where the pressure is given by (3.46) and in terms of the layer depths using (3.48). If we make the Boussinesq approximation then $\rho_n$ on the right-hand side of (3.48) is replaced by $\rho_1$.

Finally, the mass conservation equation for each layer has the same form as the single-layer case, and is

$$\frac{Dh_n}{Dt} + h_n \nabla \cdot u_n = 0.$$  \hspace{1cm} (3.49)

The two- and three-layer cases

The two-layer model is the simplest model to capture the effects of stratification. Evaluating the pressures using (3.46) and (3.47) we find:

$$p_1 = \rho_1 g \eta_0 = \rho_1 g (h_1 + h_2 + \eta_b)$$
$$p_2 = \rho_1 [g \eta_0 + g'_1 \eta_1] = \rho_1 \left[ g (h_1 + h_2 + \eta_b) + g'_1 (h_2 + \eta_b) \right].$$  \hspace{1cm} (3.50a)

(3.50b)

The momentum equations for the two layers are then

$$\frac{Du_1}{Dt} + f \times u_1 = -g \nabla \eta_0 = -g \nabla (h_1 + h_2 + \eta_b).$$  \hspace{1cm} (3.51a)
3.3 Multi-Layer Shallow Water Equations

The two layer shallow water system. A fluid of density \( \rho_1 \) lies over a denser fluid of density \( \rho_2 \). In the reduced gravity case the lower layer may be arbitrarily thick and is assumed stationary and so has no horizontal pressure gradient. In the 'rigid-lid' approximation the top surface displacement is neglected, but there is then a non-zero pressure gradient induced by the lid.

\[
\frac{Du_2}{Dt} + f \times u_2 = -\frac{\rho_1}{\rho_2} \left( g \nabla \eta_0 + g' \nabla \eta_1 \right) \\
= -\frac{\rho_1}{\rho_2} \left[ g \nabla (\eta_b + h_1 + h_2) + g' \nabla (h_2 + \eta_b) \right].
\] (3.51b)

In the Boussinesq approximation \( \rho_1/\rho_2 \) is replaced by unity.

In a three layer model the dynamical pressures are found to be

\[
p_1 = \rho_1 gh \\
p_2 = \rho_1 \left[ gh + g' \left( h_2 + h_3 + \eta_b \right) \right] \\
p_3 = \rho_1 \left[ gh + g' \left( h_2 + h_3 + \eta_b \right) + g' \left( h_3 + \eta_b \right) \right],
\] (3.52a, 3.52b, 3.52c)

where \( h = \eta_0 = \eta_b + h_1 + h_2 + h_3 \) and \( g' = g(\rho_3 - \rho_2)/\rho_1 \). More layers can obviously be added in a systematic fashion.

3.3.1 Reduced-gravity multi-layer equation

As with a single active layer, we may envision multiple layers of fluid overlying a deeper stationary layer. This is a useful model of the stratified upper ocean overlying a nearly stationary and nearly unstratified abyss. Indeed we use such a model to study the ‘ventilated thermocline’ in chapter 16 and a detailed treatment may be found there. If we suppose there is a lid at the top, then the model is almost the same as that of the previous section. However, now the horizontal pressure gradient in the lowest model layer is zero, and so we may obtain the pressures in all the active layers by integrating the hydrostatic equation upwards from this layer.
fluid velocity, out of page
fluid velocity, into page
Coriolis force
Pressure force
Free surface
Pressure force
Coriolis force

Fig. 3.6 Geostrophic flow in a shallow water system, with a positive value of the Coriolis parameter \( f \), as in the Northern hemisphere. The pressure force is directed down the gradient of the height field, and this can be balanced by the Coriolis force if the fluid velocity is at right angles to it. If \( f \) were negative, the geostrophic flow would be reversed.

Suppose we have \( N \) moving layers, then the reader may verify that the dynamic pressure in the \( n' \)th layer is given by

\[
p_n = - \sum_{i=n}^{i=N} \rho_1 \gamma_i \eta_i,
\]

where as before \( \gamma_i = g(\rho_{i+1} - \rho_i)/\rho_1 \). If we have a lid at the top, so that \( \eta_0 = 0 \), then the interface displacements are related to the layer thicknesses by

\[
\eta_n = - \sum_{i=1}^{i=n} h_i.
\]

From these expressions the momentum equation in each layer is easily constructed.

### 3.4 GEOSTROPHIC BALANCE AND THERMAL WIND

Geostrophic balance occurs in the shallow water equations, just as in the continuously stratified equations, when the Rossby number \( U/fL \) is small and the Coriolis term dominates the advective terms in the momentum equation. In the single-layer shallow water equations the geostrophic flow is:

\[
f \times u = - \nabla \eta.
\]

Thus, the geostrophic velocity is proportional to the slope of the surface, as sketched in Fig. 3.6. (For the rest of this section, we will drop the subscript \( g \), and take all velocities to be geostrophic.)

In both the single-layer and multi-layer case, the slope of an interfacial surface is directly related to the difference in pressure gradient on either side and so, by geostrophic balance, to the shear of the flow. This is the shallow water analog of the thermal wind relation. To obtain an expression for this, consider the interface,
Figure 3.7 Margules' relation: using hydrostasy, the difference in the horizontal pressure gradient between the upper and the lower layer is given by $-g' \rho_1 s$ where $s = \tan \phi = \Delta z/\Delta y$ is the interface slope and $g' = (\rho_2 - \rho_1)/\rho_1$. Geostrophic balance then gives $f \partial \eta (u_1 - u_2) = g' s$, which is a special case of (3.60).

$\eta$, between two layers labelled 1 and 2. The pressure in two layers is given by the hydrostatic relation and so,

\begin{align*}
p_1 &= A(x, y) - \rho_1 g z \quad \text{at some } z \text{ in layer 1} \quad (3.56a) \\
p_2 &= A(x, y) - \rho_1 g \eta + \rho_2 g (\eta - z) = A(x, y) + \rho_1 g' \eta - \rho_2 g z \quad \text{at some } z \text{ in layer 2} \quad (3.56b)
\end{align*}

where $A(x, y)$ is the pressure where $z = 0$. (We don’t need to specify where this is, except that it is in or at the top of the top layer). Thus we find

$$\frac{1}{\rho_1} \nabla (p_1 - p_2) = -g' \nabla \eta. \quad (3.57)$$

If the flow is geostrophically balanced and Boussinesq then, in each layer, the velocity obeys

$$f \mathbf{u}_i = \frac{1}{\rho_1} \mathbf{k} \times \nabla p_i. \quad (3.58)$$

Using (3.57) then gives

$$f (u_1 - u_2) = -\mathbf{k} \times g' \nabla \eta, \quad (3.59)$$

or in general

$$f (u_n - u_{n+1}) = -\mathbf{k} \times g' \nabla \eta. \quad (3.60)$$

This is the thermal wind equation for the shallow water system. It applies at any interface, and it implies the shear is proportional to the interface slope, a result known as the 'Margules relation' (Fig. 3.7).²

Suppose that we represent the atmosphere by two layers of fluid; a meridionally decreasing temperature may then be represented by an interface that slopes upward toward the pole. Then, in either hemisphere, we have

$$u_1 - u_2 = \frac{g_1}{f} \frac{\partial \eta}{\partial y} > 0, \quad (3.61)$$

and the temperature gradient is associated with a positive shear. (See problem 3.2.)

3.5 **FORM DRAG**

When the interface between two layers varies with position — that is, when it is wavy — the layers exert a pressure force on each other. Similarly, if the bottom of
the fluid is not flat then the topography and the bottom layer will in general exert forces on each other. These kind of forces are known as form drag, and it is an important means whereby momentum can be added to or extracted from a flow. Consider a layer confined between two interfaces, \( \eta_1(x,y) \) and \( \eta_2(x,y) \). Then over some zonal interval \( L \) the average zonal pressure force on that fluid layer is given by

\[
F_p = -\frac{1}{L} \int_{x_1}^{x_2} \int_{\eta_1}^{\eta_2} \frac{\partial p}{\partial x} \, dx \, dz.
\]  

Integrating by parts first in \( z \) and then in \( x \), and noting that by hydrostasy \( \frac{\partial p}{\partial z} \) does not depend on horizontal position within the layer, we obtain

\[
F_p = -\frac{1}{L} \int_{x_1}^{x_2} \left[ \frac{\partial p}{\partial x} \right]^{\eta_1}_{\eta_2} \, dx
= -\eta_1 \frac{\partial p_1}{\partial x} + \eta_2 \frac{\partial p_2}{\partial x} = +p_1 \frac{\partial \eta_1}{\partial x} - p_2 \frac{\partial \eta_2}{\partial x},
\]  

where \( p_1 \) is the pressure at \( \eta_1 \), and similarly for \( p_2 \), and to obtain the second line we suppose that the integral is around a closed path, such as a circle of latitude, and the average is denoted with an overbar. These terms represent the transfer of momentum from one layer to the next, and at a particular interface, \( i \), we may define the form drag, \( \tau_i \), by

\[
\tau_i \equiv p_i \frac{\partial \eta_i}{\partial x} = -\eta_i \frac{\partial p_i}{\partial x}.
\]  

The form drag is a stress, and as the layer depth shrinks to zero its vertical derivative, \( \frac{\partial \tau}{\partial z} \), is the force on the fluid. It is a particularly important mechanism for the vertical transfer of momentum and its ultimate removal in an eddying fluid, and it one of the the main mechanisms whereby the wind stress at the top of the ocean is communicated to the ocean bottom. At the fluid bottom the form drag is \( p\eta_{bx} \), where \( \eta_b \) is the bottom topography, and this is proportional to the momentum exchange with the solid earth. This is a significant mechanism for the ultimate removal of momentum in the ocean, especially in the Antarctic Circumpolar Current where it is likely to be much larger than bottom (or Ekman) drag arising from small scale turbulence and friction. In the two layer, flat-bottomed case the only form drag occurring is that at the interface, and the momentum transfer between the layers is just \( p_1 \frac{\partial \eta_1}{\partial x} \) or \( -\eta_1 \frac{\partial p}{\partial x} \); then, the force on each layer due to the other is equal and opposite, as we would expect from momentum conservation.

For flows in geostrophic balance, the form drag is related to the meridional heat flux. The pressure gradient and velocity are related by \( p f v' = \frac{\partial p'}{\partial x} \) and the interfacial displacement is proportional to the temperature perturbation, \( b' \) (in fact one may show that \( \eta' \approx -b' / (\partial \tilde{B} / \partial z) \)). Thus \( -\eta' \frac{\partial p_{\tilde{B}}}{\partial x} \propto v' B' \), a correspondence that will re-occur when we consider the Eliassen-Palm flux in chapter 7.

### 3.6 CONSERVATION PROPERTIES OF SHALLOW WATER SYSTEMS

There are two common types of conservation property in fluids: (i) material invariants and (ii) integral invariants. Material invariance occurs when a property (\( \phi \) say) is conserved on each fluid element, and so obeys the equation \( D\phi /Dt = 0 \).
An integral invariant is one that is conserved following an integration over some, usually closed, volume; energy is an example.

### 3.6.1 A material invariant: potential vorticity

The vorticity of a fluid (considered at greater length in chapter 4), denoted \( \omega \), is defined to be the curl of the velocity field, so that

\[
\omega \equiv \nabla \times \mathbf{v}.
\]  
(3.65)

Let us also define the shallow water vorticity, \( \omega^* \), as the curl of the horizontal velocity, so that

\[
\omega^* \equiv \nabla \times \mathbf{u}.
\]  
(3.66)

and, because \( \partial u/\partial z = \partial v/\partial z = 0 \), only its vertical component is non-zero and

\[
\omega^* = k \left( \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right) \equiv k \zeta.
\]  
(3.67)

Using the vector identity

\[
(u \cdot \nabla)u = \frac{1}{2} \nabla (u \cdot u) - u \times (\nabla \times u),
\]  
(3.68)

we write the momentum equation, (3.8), as

\[
\frac{\partial \mathbf{u}}{\partial t} + \omega^* \times \mathbf{u} = -\nabla (g\eta + \frac{1}{2} \mathbf{u}^2).
\]  
(3.69)

To obtain an evolution equation for the vorticity we take the curl of (3.69), and make use of the vector identity

\[
\nabla \times (\omega^* \times \mathbf{u}) = (u \cdot \nabla) \omega^* - (\omega^* \cdot \nabla) \mathbf{u} + \omega^* \nabla \cdot \mathbf{u} - \mathbf{u} \nabla \cdot \omega^*
\]  
(3.70)

using the fact that \( \nabla \cdot \omega^* \) is the divergence of a curl and therefore zero, and \( (\omega^* \cdot \nabla) \mathbf{u} = 0 \) because \( \omega^* \) is perpendicular to the surface in which \( \mathbf{u} \) varies. The curl of (3.69) is then

\[
\frac{\partial \omega^*}{\partial t} + (u \cdot \nabla) \omega^* = -\omega^* \nabla \cdot \mathbf{u},
\]  
(3.71)

or

\[
\frac{\partial \zeta}{\partial t} + (u \cdot \nabla) \zeta = -\zeta \nabla \cdot \mathbf{u}.
\]  
(3.72)

where \( \zeta = k \cdot \omega^* \). However, the mass conservation equation may be written as

\[
-\zeta \nabla \cdot \mathbf{u} = \frac{\zeta}{h} \frac{Dh}{Dt}.
\]  
(3.73)

Thus, (3.72) becomes

\[
\frac{D\zeta}{Dt} = \frac{\zeta}{h} \frac{Dh}{Dt},
\]  
(3.74)

which simplifies to

\[
\frac{D}{Dt} \left( \frac{\zeta}{h} \right) = 0.
\]  
(3.75)
The important quantity $\zeta/h$, often denoted $Q$, is known as the potential vorticity, and (3.75) is known as the potential vorticity equation. We re-derive this conservation law in a more general way in section 4.6.

Because $Q$ is conserved on parcels, then so is any function of $Q$; that is, $F(Q)$ is a material invariant, where $F$ is any function. To see this algebraically, multiply (3.75) by $F'(Q)$, the derivative of $F$ with respect to $Q$, giving

$$F'(Q) \frac{DQ}{Dt} = \frac{DF}{Dt}F(Q) = 0.$$  \hspace{1cm} (3.76)

Since $F$ is arbitrary there are an infinite number of material invariants corresponding to different choices of $F$.

**Effects of rotation**

In a rotating frame of reference, the shallow water momentum equation is

$$\frac{Du}{Dt} + f \times u = -g \nabla \eta,$$  \hspace{1cm} (3.77)

where (as before) $f = f \mathbf{k}$. This may be written in vector invariant form as

$$\frac{\partial u}{\partial t} + (\omega^* + f) \times u = -\nabla(g\eta + \frac{1}{2}u^2),$$  \hspace{1cm} (3.78)

and taking the curl of this gives the vorticity equation

$$\frac{\partial \zeta}{\partial t} + (u \cdot \nabla)(\zeta + f) = -(f + \zeta) \nabla \cdot u.$$  \hspace{1cm} (3.79)

This is the same as the shallow water vorticity equation in a non-rotating frame, save that $\zeta$ is replaced by $\zeta + f$, the reason for this being that $f$ is the vorticity that the fluid has by virtue of the background rotation. Thus, (3.79) is simply the equation of motion for the total or absolute vorticity, $\omega_a = \omega^* + f = (\zeta + f)\mathbf{k}$.

The potential vorticity equation in the rotating case follows, much as in non-rotating case, by combining (3.79) with the mass conservation equation, giving

$$\frac{D}{Dt} \left( \frac{\zeta + f}{h} \right) = 0.$$  \hspace{1cm} (3.80)

That is, $Q = (\zeta + f)/h$, the potential vorticity in a rotating shallow system, is a material invariant.

**Vorticity and circulation**

Although vorticity itself is not a material invariant, its integral over a horizontal material area is invariant. To demonstrate this in the non-rotating case, consider the integral

$$C = \int_A \zeta \, dA = \int_A Q \, h \, dA,$$  \hspace{1cm} (3.81)

over a surface $A$, the cross-sectional area of a column of height $h$ (as in Fig. 3.2). Taking the material derivative of this gives

$$\frac{DC}{Dt} = \int_A \frac{DQ}{Dt} \, h \, dA + \int_A Q \frac{D}{Dt}(h \, dA).$$  \hspace{1cm} (3.82)
The first term is zero, by (3.74); the second term is just the derivative of the volume of a column of fluid and it too is zero, by mass conservation. Thus,

\[ \frac{DC}{Dt} = \frac{D}{Dt} \int_A \zeta \, dA = 0. \tag{3.83} \]

Thus, the integral of the vorticity over a some cross-sectional area of the fluid is unchanging, although both the vorticity and area of the fluid may individually change. Using Stokes’s theorem, it may be written

\[ \frac{DC}{Dt} = \frac{D}{Dt} \oint u \cdot dl, \tag{3.84} \]

where the line integral is around the boundary of \( A \). This is an example of Kelvin’s circulation theorem, which we shall meet again in more general form in chapter 4, where we also consider the rotating case.

A slight generalization of (3.83) is possible. Consider the integral

\[ I = \int F(Q)h \, dA \]

where again \( F \) is any differentiable function of its argument. It is clear that

\[ \frac{D}{Dt} \int_A F(Q)h \, dA = 0. \tag{3.85} \]

If the area of integration in (3.69) or (3.85) is the whole domain (enclosed by frictionless walls, for example) then it is clear that the integral of \( hF(Q) \) is a constant, including as a special case the integral of \( \zeta \).

### 3.6.2 Energy conservation — an integral invariant

Since we have made various simplifications in deriving the shallow water system, it is not self-evident that energy should be conserved, or indeed what form the energy takes. The kinetic energy density, that is the kinetic energy per unit area, is \( \rho h u^2 / 2 \).

The potential energy density of the fluid is

\[ PE = \int_0^h \rho_0 g z \, dz = \frac{1}{2} \rho_0 gh^2. \tag{3.86} \]

The factor \( \rho_0 \) appears in both kinetic and potential energies and, because it is a constant, we will omit it.

Using the mass conservation equation (3.15) we obtain an equation for the evolution of potential energy density:

\[ \frac{D}{Dt} \left( \frac{gh^2}{2} + gh^2 \nabla \cdot u \right) = 0 \tag{3.87a} \]

or

\[ \frac{\partial}{\partial t} \left( \frac{gh^2}{2} \right) + \nabla \cdot \left( u \frac{gh^2}{2} \right) + \frac{gh^2}{2} \nabla \cdot u = 0. \tag{3.87b} \]

From the momentum and mass continuity equations we obtain an equation for the evolution of kinetic energy density, namely

\[ \frac{D}{Dt} \left( \frac{hu^2}{2} + u^2h \nabla \cdot u \right) = -g u \cdot \nabla h^2 / 2. \tag{3.88a} \]
or
\[ \frac{\partial}{\partial t} \left( \frac{hu^2}{2} \right) + \nabla \cdot \left( u \frac{hu^2}{2} \right) + g u \cdot \nabla \frac{h^2}{2} = 0. \] (3.88b)

Adding (3.87b) and (3.88b) we obtain
\[ \frac{\partial}{\partial t} \left( \frac{1}{2} (hu^2 + gh^2) \right) + \nabla \cdot \left[ \frac{1}{2} u \left( gh^2 + hu^2 + gh^2 \right) \right] = 0, \] (3.89)

or
\[ \frac{\partial E}{\partial t} + \nabla \cdot F = 0. \] (3.90)

where \( E = KE + PE = (hu^2 + gh^2)/2 \) is the density of the total energy and \( F = u(hu^2 + gh^2 + gh^2)/2 \) is the energy flux. If the fluid is confined to a domain bounded by rigid walls, on which the normal component of velocity vanishes, then on integrating (3.89) over that area and using Gauss's theorem, the total energy is seen to be conserved; that is
\[ \frac{d\hat{E}}{dt} = \frac{1}{2} \frac{d}{dt} \int_A (hu^2 + gh^2) dA = 0. \] (3.91)

Such an energy principle also holds in the case with bottom topography. Note that, as we found in the case for a compressible fluid in chapter 2, the energy flux in (3.90) is not just the energy density multiplied by the velocity but it contains an additional term \( g u h^2 / 2 \), and this represents the energy transfer occurring when the fluid does work against the pressure force (see problem 3.3).

3.7 SHALLOW WATER WAVES

Let us now look at the gravity waves that occur in shallow water. To isolate the essence of the phenomena, we will consider waves in a single fluid layer, with a flat bottom and a free upper surface, in which gravity provides the sole restoring force.

3.7.1 Non-rotating shallow water waves

Given a flat bottom the fluid thickness is equal to the free surface displacement (Fig. 3.1), and we let
\[ h(x, y, t) = H + h'(x, y, t) = H + \eta'(x, y, t), \] (3.92a)
\[ u(x, y, t) = u'(x, y, t). \] (3.92b)

The mass conservation equation, (3.15), then becomes
\[ \frac{\partial \eta'}{\partial t} + (H + \eta') \nabla \cdot u' + u' \cdot \nabla \eta' = 0, \] (3.93)

and neglecting squares of small quantities this yields the linear equation
\[ \frac{\partial \eta'}{\partial t} + H \nabla \cdot u' = 0. \] (3.94)
Similarly, linearizing the momentum equation, (3.8) with \( f = 0 \), yields

\[
\frac{\partial u'}{\partial t} = -g \nabla \eta' .
\] (3.95)

Eliminating velocity by differentiating (3.94) with respect to time and taking the divergence of (3.95) leads to

\[
\frac{\partial^2 \eta'}{\partial t^2} - g H \nabla^2 \eta' = 0,
\] (3.96)

which may be recognized as a wave equation. We can find the dispersion relationship for this by substituting the trial solution

\[
\eta' = \text{Re} \hat{\eta} e^{i(k \cdot x - \omega t)}
\] (3.97)

where \( \hat{\eta} \) is a complex constant, \( k = ik + jl \) is the horizontal wavenumber, and \( \text{Re} \) indicates that the real part of the solution should be taken. If for simplicity we restrict attention for the moment to the one-dimensional problem, with no variation in the \( y \)-direction, then substituting into (3.96) leads to the dispersion relationship

\[
\omega = \pm ck,
\] (3.98)

where \( c = \sqrt{gH} \). That is, the wave speed is proportional to the square root of the mean fluid depth and is independent of the wavenumber — that is, the waves are dispersionless. The general solution is a superposition of all such waves, with the amplitudes of each wave (or Fourier component) being determined by the Fourier decomposition of the initial conditions.

Because the waves are dispersionless, the general solution can be written

\[
\eta'(x,t) = \frac{1}{2} [F(x - ct) + F(x + ct)],
\] (3.99)

where \( F(x) \) is the height field at \( t = 0 \). From this, it is easy to see that the shape of an initial disturbance is preserved as it propagates both to the right and to the left at speed \( c \). (See also problem 3.7.)

### 3.7.2 Rotating shallow water (Poincaré) waves

We now consider the effects of rotation. Linearizing the rotating, flat-bottomed \( f \)-plane shallow water equations [i.e., (SW.1) and (SW.2) on page 127] about a state of rest we obtain

\[
\frac{\partial u'}{\partial t} - f_0 v' = -g \frac{\partial \eta'}{\partial x}, \quad \frac{\partial v'}{\partial t} + f_0 u' = -g \frac{\partial \eta'}{\partial y}, \quad \frac{\partial \eta'}{\partial t} + H \left( \frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} \right) = 0.
\] (3.100a,b,c)

It is convenient to nondimensionalize these equations and we write

\[
(x, y) = L(\hat{x}, \hat{y}), \quad (u', v') = U(u, v), \quad t = \frac{L}{U} \hat{t}, \quad f_0 = \frac{f_0}{T}, \quad \eta' = H \hat{\eta}.
\] (3.101)
Eq. (3.100) then becomes
\[
\frac{\partial \hat{u}}{\partial t} \hat{f}_0 + \hat{f}_0 \hat{u} = - \hat{c}^2 \frac{\partial \hat{\eta}}{\partial \hat{x}}, \quad \frac{\partial \hat{v}}{\partial t} + \hat{f}_0 \hat{v} = - \hat{c}^2 \frac{\partial \hat{\eta}}{\partial \hat{y}}, \quad \frac{\partial \hat{\eta}}{\partial t} + \left( \frac{\partial \hat{u}}{\partial \hat{x}} + \frac{\partial \hat{v}}{\partial \hat{y}} \right) = 0.
\]
\[\text{(3.102a,b,c)}\]
where \(\hat{c} = \sqrt{gH/U}\) is the nondimensional speed of nonrotating shallow-water waves. (It is also the inverse of the Froude number \(U/\sqrt{gH}\).) To obtain a dispersion relationship we let
\[
(\hat{u}, \hat{v}, \hat{\eta}) = (\tilde{u}, \tilde{v}, \tilde{\eta}) e^{i(\hat{k} \cdot \hat{x} - \hat{\omega} \hat{t})},
\]
where \(\hat{k} = \hat{k}_i + \hat{l}_j\) and \(\hat{\omega}\) is the nondimensional frequency, and substitute into (3.102), giving
\[
\begin{pmatrix}
  i\hat{\omega} - \hat{f}_0 & i\hat{c}^2 \hat{k} \\
  \hat{f}_0 & i\hat{\omega} \\
  i\hat{k} & i\hat{l} & i\hat{\omega}
\end{pmatrix}
\begin{pmatrix}
  \tilde{u} \\
  \tilde{v} \\
  \tilde{\eta}
\end{pmatrix} = 0.
\]
\[\text{(3.104)}\]
This homogeneous equation has nontrivial solutions only if the determinant of the matrix vanishes. This condition gives
\[
\hat{\omega} \left( \hat{\omega}^2 - \hat{f}_0^2 - \hat{c}^2 \hat{k}^2 \right) = 0.
\]
\[\text{(3.105)}\]
where \(\hat{K}^2 = \hat{k}^2 + \hat{l}^2\). There are two classes of solution to (3.105). The first is simply \(\hat{\omega} = 0\), time-independent flow corresponding to geostrophic balance in (3.100). (Because geostrophic balance gives a divergence-free velocity field for constant Coriolis parameter the equations are satisfied by a time-independent solution.) The second set of solutions satisfies the dispersion relation
\[
\hat{\omega}^2 = \hat{f}_0^2 + \hat{c}^2(\hat{k}^2 + \hat{l}^2).
\]
\[\text{(3.106)}\]
which in dimensional form is:
\[
\omega^2 = f_0^2 + gH(k^2 + l^2).
\]
\[\text{(3.107)}\]
The corresponding waves are known as Poincaré waves, and the dispersion relationship is illustrated in Fig. 3.8. Note that the frequency is always greater than the Coriolis frequency \(f_0\). There are two interesting limits:

(i) The short waves limit: If
\[
K^2 \gg \frac{f_0^2}{gH},
\]
where \(K^2 = k^2 + l^2\), then the dispersion relationship reduces to that of the nonrotating case (3.98). This condition is equivalent to requiring that the wavelength be much shorter than the deformation radius, \(L_d\). Specifically, if \(l = 0\) and \(\lambda = 2\pi/k\) is the wavelength, the condition is
\[
\lambda^2 \ll L_d^2(2\pi)^2
\]
\[\text{(3.109)}\]
The numerical factor of \((2\pi)^2\) is more than an order of magnitude, so care must be taken when deciding if the condition is satisfied in particular cases. Furthermore, the wavelength must still be longer than the depth of the fluid, else the shallow water condition is not met.
3.7 Shallow Water Waves

(ii) The long wave limit: If
\[ K^2 \ll \frac{f_0^2}{gh}, \]  
that is if the wavelength is much longer than the deformation radius \( L_d \), then the dispersion relationship is
\[ \omega = f_0. \]  

These are known as inertial oscillations. The equations of motion giving rise to them are
\[ \frac{\partial u'}{\partial t} - f_0 v' = 0, \quad \frac{\partial v'}{\partial t} + f_0 u' = 0, \]  
which are equivalent to material equations for free particles in a rotating frame, unconstrained by pressure forces, namely
\[ \frac{d^2 x}{dt^2} - f_0 v = 0, \quad \frac{d^2 y}{dt^2} + f_0 u = 0. \]

See also problem 3.9.

3.7.3 Kelvin waves

The Kelvin wave is a particular type of gravity wave that exists in the presence of both rotation and a lateral boundary. Suppose there is a solid boundary at \( y = 0 \); clearly harmonic solutions in the \( y \)-direction are not allowable, as these would not satisfy the condition of no-normal flow at the boundary. Do any wavelike solutions exist? The affirmative answer to this question was provided by Kelvin and the associated waves are now eponymously known as Kelvin waves.\(^5\) We begin with the linearized shallow water equations, namely
\[ \frac{\partial u'}{\partial t} - f_0 v' = -g \frac{\partial \eta'}{\partial x}, \quad \frac{\partial v'}{\partial t} + f_0 u' = -g \frac{\partial \eta'}{\partial y}, \quad \frac{\partial \eta'}{\partial t} + H \left( \frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} \right) = 0. \]  
\( (3.114a,b,c) \)
The fact that \( v' = 0 \) at \( x = 0 \) suggests that we look for a solution with \( v' = 0 \) everywhere, whence these equations become

\[
\frac{\partial u'}{\partial t} = -g \frac{\partial \eta'}{\partial x}, \quad f_0 u' = -g \frac{\partial \eta'}{\partial y'}, \quad \frac{\partial \eta'}{\partial t} + H \frac{\partial u'}{\partial x} = 0.
\]

Equations (3.115a,b,c) lead to the standard wave equation

\[
\frac{\partial^2 u'}{\partial t^2} = c^2 \frac{\partial^2 u'}{\partial x^2},
\]

where \( c = \sqrt{gH} \), the usual wave speed of shallow water waves. The solution of (3.116) is

\[
u' = F_1(x + ct, y') + F_2(x - ct, y'),
\]

with corresponding surface displacement

\[
\eta' = \sqrt{H/g} \left[ -F_1(x + ct, y') + F_2(x - ct, y') \right].
\]

The solution represents the superposition of two waves, one \( F_1 \) travelling in the negative \( x \)-direction, and the other in the positive \( x \)-direction. To obtain the \( y \)-dependence of these functions we use (3.115b) which gives

\[
\frac{\partial F_1}{\partial y'} = \frac{f_0}{\sqrt{gH}} F_1, \quad \frac{\partial F_2}{\partial y'} = -\frac{f_0}{\sqrt{gH}} F_2,
\]

with solutions

\[
F_1 = F(x + ct) e^{y'/L_d} \quad F_2 = G(x - ct) e^{-y'/L_d},
\]

where \( L_d = \sqrt{gH/f_0} \) is the radius of deformation. The solution \( F_1 \) grows exponentially away from the wall, and so fails to satisfy the condition of boundedness at infinity. It must be thus eliminated, leaving the general solution

\[
\begin{align*}
u' &= e^{-y'/L_d} G(x - ct), \quad v' = 0, \\
\eta' &= \sqrt{H/g} e^{-y'/L_d} G(x - ct).
\end{align*}
\]

These are Kelvin waves, and they decay exponentially away from the boundary. If \( f_0 \) is positive, as in the Northern Hemisphere, the boundary is to the right of an observer moving with the wave. Given a constant Coriolis parameter, we could equally well have obtained a solution on a meridional wall, in which case we would find that the wave again moves such that the wall is to the right of the wave direction. (This is obvious once it is realized that \( f \)-plane dynamics are isotropic in \( x \) and \( y \).) Thus, in the Northern Hemisphere the wave moves anticlockwise round a basin, and conversely in the Southern Hemisphere, and in both hemispheres the direction is cyclonic.

### 3.8 GEOSTROPHIC ADJUSTMENT

We noted in chapter 2 that the large-scale, extra-tropical circulation of the atmosphere is in near-geostrophic balance. Why is this? Why should the Rossby number
be small? Arguably, the magnitude of the velocity in the atmosphere and ocean is ultimately given by the strength of the forcing, and so ultimately by the differential heating between pole and equator (although even this argument is not satisfactory, since the forcing mainly determines the energy throughput, not directly the energy itself, and the forcing is itself dependent on the atmosphere's response). But even supposing that the velocity magnitudes are given, there is no a priori guarantee that the forcing or the dynamics will produce length-scales that are such that the Rossby number is small. However, there is in fact a powerful and ubiquitous process whereby a fluid in an initially unbalanced state naturally evolves toward a state of geostrophic balance, namely geostrophic adjustment. This process occurs quite generally in rotating fluids, stratified or not. To pose the problem in a simple form we will consider the free evolution of a single shallow layer of fluid whose initial state is manifestly unbalanced, and we will suppose that surface displacements are small so that the evolution of the system is described by the linearized shallow equations of motion. These are

\[
\frac{\partial u}{\partial t} + f \times u = -g \nabla \eta, \quad \frac{\partial \eta}{\partial t} + H \nabla \cdot u = 0,
\]

where \(\eta\) is the free surface displacement and \(H\) is the mean fluid depth, and we omit the primes on the linearized variables.

3.8.1 Non-rotating flow

We consider first the non-rotating problem set, with little loss of generality, in one dimension. We suppose that initially the fluid is at rest but with a simple discontinuity in the height field so that

\[
\eta(x, t = 0) = \begin{cases} 
+\eta_0 & x < 0 \\
-\eta_0 & x > 0
\end{cases}
\]

and \(u(x, t = 0) = 0\) everywhere. We can physically realize these initial conditions by separating two fluid masses of different depths by a thin dividing wall, and then quickly removing the wall. What is the subsequent evolution of the fluid? The general solution to the linear problem is given by (3.99) where the functional form is determined by the initial conditions so that here

\[
F(x) = \eta(x, t = 0) = -\eta_0 \text{sgn}(x).
\]

Eq. (3.99) states that this initial pattern is propagated to the right and to the left. That is, two discontinuities in fluid height simply propagate to the right and left at a speed \(c = \sqrt{gH}\). Specifically, the solution is

\[
\eta(x, t) = \frac{1}{2} \eta_0 [\text{sgn}(x + ct) + \text{sgn}(x - ct)]
\]

The initial conditions may be much more complex than a simple front, but, because the waves are dispersionless, the solution is still simply a sum of the translation of those initial conditions to the right and to the left at speed \(c\). The velocity field in this class of problem is obtained from

\[
\frac{\partial u}{\partial t} = -g \frac{\partial \eta}{\partial x}.
\]
Chapter 3. Shallow Water Systems and Isentropic Coordinates

Fig. 3.9 The time development of an initial ‘top hat’ height disturbance, with zero initial velocity. Fronts propagate in both directions, and the velocity is non-zero between fronts, but ultimately the velocity and height disturbance are radiated away to infinity.

which gives, using (3.99),

$$u = -\frac{\eta}{2c} [F(x + ct) - F(x - ct)].$$

(3.127)

Consider the case with initial conditions given by (3.123). At a given location, away from the initial disturbance, the fluid remains at rest and undisturbed until the front arrives. After the front has passed, the fluid surface is again undisturbed and the velocity is uniform and non zero. Specifically:

$$\eta = \begin{cases} -\eta_0 \text{sgn}(x) & |x| > ct \\ 0 & |x| < ct \end{cases}$$

$$u = \begin{cases} 0 & |x| > ct \\ \eta_0 \beta/c & |x| < ct \end{cases}.$$  

(3.128)

The solution with ‘top-hat’ initial conditions in the height field, and zero initial velocity, is a superposition two discontinuities similar to (3.128) and is illustrated in Fig. 3.9. Two fronts propagate in either direction from each discontinuity and, in this case, the final velocity, as well as the fluid displacement, is zero after all the fronts have passed. That is, the disturbance is radiated completely away.

3.8.2 Rotating flow

Rotation makes a profound difference to the adjustment problem of the shallow water system, because a steady, adjusted, solution does exist with nonzero gradients in the height field — the associated pressure gradients being balanced by the Coriolis force — and potential vorticity conservation provides a powerful constraint on the fluid evolution.$^6$ In a rotating shallow fluid that conservation is represented by

$$\frac{\partial Q}{\partial t} + u \cdot \nabla Q = 0,$$

(3.129)

where $Q = (\zeta + f)/H$. In the linear case with constant Coriolis parameter (3.129) becomes

$$\frac{\partial q}{\partial t} = 0, \quad q = \left(\zeta + \beta \frac{\eta}{H}\right).$$

(3.130)
This equation may be obtained either from the linearized velocity and mass conservation equations, (3.122), or from (3.129) directly. In the latter case, we write

$$Q = \frac{\zeta + f_0}{H + \eta} \approx \frac{1}{H} (\zeta + f_0) \left(1 - \frac{\eta}{H}\right) \approx \frac{1}{H} \left(f_0 + \zeta - f_0 \frac{\eta}{H}\right) = \frac{f_0}{H} + \frac{q}{H}$$  \hspace{1cm} (3.131)

having used $f_0 \gg |\zeta|$ and $H \gg |\eta|$. The term $f_0/H$ is a constant and so dynamically unimportant, as is the $H^{-1}$ factor multiplying $q$. Further, the advective term $u \cdot \nabla Q$ becomes $u \cdot \nabla q$, and this is second order in perturbed quantities and so is neglected. Thus, making these approximations, (3.129) reduces to (3.130). The potential vorticity field is therefore fixed in space! Of course, this was also true in the nonrotating case where the fluid is initially at rest. Then $q = \zeta = 0$ and the fluid remains irrotational throughout the subsequent evolution of the flow. However, this is rather a weak constraint on the subsequent evolution of the fluid; it does nothing, for example, to prevent the conversion of all the potential energy to kinetic energy. In the rotating case the potential vorticity is non-zero, and potential vorticity conservation and geostrophic balance are all we need to infer the final steady state, assuming it exists, without solving for the details of the flow evolution, as we now see.

With an initial condition for the height field given by (3.123), the initial potential vorticity is given by

$$q(x, y) = \begin{cases} -\frac{f_0 \eta_0}{H} & x < 0 \\ \frac{f_0 \eta_0}{H} & x > 0 \end{cases}$$  \hspace{1cm} (3.132)

and this remains unchanged throughout the adjustment process. The final steady state is then the solution of the equations

$$\zeta - f_0 \frac{\eta}{H} = q(x, y), \hspace{0.5cm} f_0 u = -g \frac{\partial \eta}{\partial y}, \hspace{0.5cm} f_0 v = g \frac{\partial \eta}{\partial x},$$  \hspace{1cm} (3.133a,b,c)

where $\zeta = \partial v/\partial x - \partial u/\partial y$. Because the Coriolis parameter is constant, the velocity field is horizontally non-divergent and we may define a streamfunction $\psi = g \eta / f_0$. Equations (3.133) then reduce to

$$\left(\nabla^2 - \frac{1}{L_d^2}\right) \psi = q(x, y),$$  \hspace{1cm} (3.134)

where $L_d = \sqrt{gH/f_0}$ is known as the Rossby radius of deformation or often just the ‘deformation radius’ or the ‘Rossby radius’. It is a naturally occurring length-scale in problems involving both rotation and gravity, and arises in slightly different form in stratified fluids.

The initial conditions (3.132) admit of a nice analytic solution, for the flow will remain uniform in $y$, and (3.134) reduces to

$$\frac{\partial^2 \psi}{\partial x^2} - \frac{1}{L_d^2} \psi = -\frac{f_0 \eta_0}{H} \text{sgn}(x).$$  \hspace{1cm} (3.135)

We solve this separately for $x > 0$ and $x < 0$ and then match the solutions and their
Fig. 3.10 Solutions of a linear geostrophic adjustment problem. Top panel: the initial height field, given by (3.123) with $\eta_0 = 1$. Second panel: equilibrium (final) height field, $\eta$ given by (3.136) and $\eta = f_0 \psi / g$. Third panel: equilibrium geostrophic velocity (normal to the gradient of height field), given by (3.137). Bottom panel: potential vorticity, given by (3.132), and this does not evolve. The distance, $x$ is non-dimensionalized by the deformation radius $L_d = \sqrt{gH/f_0}$, and the velocity by $\eta_0 (g/f_0 L_d)$. Changes to the initial state occur only within $O(L_d)$ of the initial discontinuity; and as $x \to \pm \infty$ the initial state is unaltered.

First derivatives at $x = 0$, imposing also the condition that the streamfunction decay to zero as $x \to \pm \infty$. The solution is

$$
\psi = \begin{cases} 
-(g\eta_0/f_0)(1 - e^{-x/L_d}) & x > 0 \\
+(g\eta_0/f_0)(1 - e^{x/L_d}) & x < 0 
\end{cases} \tag{3.136}
$$

The velocity field associated with this is obtained from (3.133c), and is

$$
u = 0, \quad u = -\frac{g\eta_0}{f_0 L_d} e^{-|x|/L_d}. \tag{3.137}$$

The velocity is perpendicular to the slope of the free surface, and a jet forms along the initial discontinuity, as illustrated in Fig. 3.10.

The important point of this problem is that the variations in the height and field are not radiated away to infinity, as in the non-rotating problem. Rather, potential...
vorticity conservation constrains the influence of the adjustment to within a deformation radius (we see now why this name is appropriate) of the initial disturbance. This property is a general one in geostrophic adjustment — it also arises if the initial condition consists of a velocity jump, as considered in problem 3.11.

### 3.8.3 Energetics of adjustment

How much of the initial potential energy of the flow is lost to infinity by gravity wave radiation, and how much is converted to kinetic energy? The linear equations (3.122) lead to

$$\frac{1}{2} \frac{\partial}{\partial t} (Hu^2 + g\eta^2) + gH \nabla \cdot (u\eta) = 0,$$

so that energy conservation holds in the form

$$E = \frac{1}{2} \int (Hu^2 + g\eta^2) \, dx, \quad \frac{dE}{dt} = 0,$$

provided the integral of the divergence term vanishes, as it normally will in a closed domain. The fluid has a non-zero potential energy, \(1/2 \int_{-\infty}^{\infty} g\eta^2 \, dx\), if there are variations in fluid height, and with the initial conditions (3.123) the initial potential energy is

$$PE_I = \int_{0}^{\infty} g\eta^2_0 \, dx.$$

This is nominally infinite if the fluid has no boundaries, and the initial potential energy density is \(g\eta^2_0/2\) everywhere.

In the non-rotating case, and with initial conditions (3.123), after the front has passed, the potential energy density is zero and the kinetic energy density is \(Hu^2/2 = g\eta^2_0/2\), using (3.128) and \(c^2 = gH\). Thus, all the potential energy is locally converted to kinetic energy as the front passes, and eventually the kinetic energy is distributed uniformly along the line. In the case illustrated in Fig. 3.9 the potential energy and kinetic energy are both radiated away from the initial disturbance. (Note that although we can superpose the solutions from different initial conditions, we cannot superpose their potential and kinetic energies.) The general point is that the evolution of the disturbance is not confined to its initial location.

In contrast, in the rotating case the conversion from potential to kinetic energy is largely confined to within a deformation radius of the initial disturbance, and at locations far from the initial disturbance the initial state is essentially unaltered. The conservation of potential vorticity has prevented the complete conversion of potential energy to kinetic energy, a result that is not sensitive to the precise form of the initial conditions (see also problem 3.10).

In fact, in the rotating case, some of the initial potential energy is converted to kinetic energy, some remains as potential energy and some is lost to infinity; let us calculate these amounts. The final potential energy, after adjustment, is, using 3.136

$$PE_F = \frac{1}{2} g\eta^2_0 \left[ \int_{0}^{\infty} (1 - e^{-x/L_d})^2 \, dx + \int_{-\infty}^{0} (1 - e^{x/L_d})^2 \, dx \right].$$

(3.141)
This is nominally infinite, but the change in potential energy is finite and is given by
\[ PE_I - PE_F = g \eta_0^2 \int_0^\infty (2 e^{-x/L_d} - e^{-2x/L_d}) \, dx = \frac{3}{2} g \eta_0^2 L_d. \] (3.142)

The initial kinetic energy is zero, because the fluid is at rest, and its final value is, using (3.137),
\[ KE_F = \frac{1}{2} H \int u^2 \, dx = H \left( \frac{g \eta_0}{f L_d} \right)^2 \int_0^\infty e^{-2x/L_d} \, dx = \frac{g \eta_0^2 L_d}{2}. \] (3.143)

Thus one-third of the difference between the initial and final potential energies is converted to kinetic energy, and this is trapped within a distance of order a deformation radius of the disturbance; the remainder, an amount \( g L_d \eta_0^2 \) is radiated away to infinity. In any finite region surrounding the initial discontinuity the final energy is less than the initial energy.

### 3.8.4 General initial conditions

Because of the linearity of the (linear) adjustment problem a spectral viewpoint is useful, in which the fields are represented as the sum or integral of non-interacting Fourier modes. For example, suppose that the height field of the initial disturbance is a two-dimensional field given by
\[ \eta(0) = \int \int \tilde{\eta}_{k,l}(0) e^{i(kx+ly)} \, dk \, dl \] (3.144)
where the Fourier coefficients \( \tilde{\eta}_{k,l}(0) \) are given, and the initial velocity field is zero. Then the initial (and final) potential vorticity field is given by
\[ q = -f_0 \frac{H}{\eta} \int \int \tilde{\eta}_{k,l}(0) e^{i(kx+ly)} \, dk \, dl. \] (3.145)

To obtain an expression for the final height and velocity fields, we express the potential vorticity field as
\[ q = \int \tilde{q}_{k,l} \, dk \, dl. \] (3.146)

The potential vorticity field does not evolve, and it is related to the initial height field by
\[ \tilde{q}_{k,l} = -f_0 \frac{H}{\eta} \eta_{k,l}(0). \] (3.147)

In the final, geostrophically balanced, state, the potential vorticity is related to the height field by
\[ q = \frac{\partial}{f_0} \nabla^2 \eta - f_0 \frac{H}{\eta} \eta \quad \text{and} \quad \tilde{q}_{k,l} = \left( -\frac{\partial}{f_0} K^2 - f_0 \frac{H}{\eta} \right) \tilde{\eta}_{k,l}, \] (3.148a,b)

where \( K^2 = k^2 + l^2 \). Using (3.147) and (3.148), the Fourier components of the final height field satisfy
\[ \left( -\frac{\partial}{f_0} K^2 - f_0 \frac{H}{\eta} \right) \tilde{\eta}_{k,l} = -f_0 \frac{H}{\eta} \tilde{\eta}_{k,l}(0) \] (3.149)
3.8 Geostrophic Adjustment

or

\[ \tilde{\eta}_{k,l} = \frac{\tilde{\eta}_k(0)}{K^2L_d^2 + 1}. \]  

(3.150)

In physical space the final height field is just the spectral integral of this, namely

\[ \eta = \iint \tilde{\eta}_{k,l} \, dk \, dl = \iint \frac{\tilde{\eta}_k(0)}{K^2L_d^2 + 1} \, dk \, dl. \]  

(3.151)

We see that at large scales \((K^2L_d^2 \ll 1)\) \(\eta_{k,l}\) is almost unchanged from its initial state; the velocity field, which is then determined by geostrophic balance, thus adjusts to the pre-existing height field. At large scales most of the energy in geostrophically balanced flow is potential energy; thus, it is energetically easier for the velocity to change to come into balance with the height field than vice versa. At small scales, however, the final height field has much less variability than it did initially.

Conversely, at small scales the height field adjusts to the velocity field. To see this, let us suppose that the initial conditions contain vorticity but have zero height displacement. Specifically, if the initial vorticity is \(\nabla^2 \psi(0)\), where \(\psi(0)\) is the initial streamfunction, then it is straightforward to show that the final streamfunction is given by

\[ \psi = \iint \tilde{\psi}_{k,l} \, dk \, dl = \iint \frac{K^2L_d^2 \tilde{\psi}_k(0)}{K^2L_d^2 + 1} \, dk \, dl. \]  

(3.152)

The final height field then obtained from this, via geostrophic balance, by \(\eta = (f_0/g)\psi\). Evidently, for small scales \((K^2L_d^2 \gg 1)\) the streamfunction, and hence the vortical component of the velocity field, are almost unaltered from their initial values. On the other hand, at large scales the final streamfunction has much less variability than it does initially, and so the height field will be largely governed by whatever variation it (and not the velocity field) had initially. In general, the final state is a superposition of the states given by (3.151) and (3.152). The divergent component of the initial velocity field does not affect the final state because it has no potential vorticity, and so all of the associated energy is lost to infinity.

Finally, we remark that just as in the problem with a discontinuous initial height profile the change in total energy during adjustment is negative — this can be seen from the form of the integrals above, although we leave the specifics as a problem to the reader. That is, some of the initial potential and kinetic energy is lost to infinity, but some is trapped by the potential vorticity constraint.

3.8.5 A variational perspective

In the non-rotating problem, all of the initial potential energy is eventually radiated away to infinity. In the rotating problem, the final state contains both potential and kinetic energy. Why is the energy not all radiated away to infinity? It is because potential vorticity conservation on parcels prevents all of the energy being dispersed. This suggests that it may be informative to think of the geostrophic adjustment problem as a variational problem: we seek to minimize the energy consistent with the conservation of potential vorticity. We stay in the linear approximation in which, because the advection of potential vorticity is neglected, potential vorticity remains constant at each point.
The energy of the flow is given by the sum of potential and kinetic energies, namely
\[
\text{Energy} = \int H(u^2 + g\eta^2) \, dA, \tag{3.153}
\]
where \( dA = dx \, dy \) and the potential vorticity field is
\[
q = \zeta - f \frac{\eta}{H}. \tag{3.154}
\]

The problem is then to extremize the energy subject to potential vorticity conservation. This is a constrained problem in the calculus of variations, sometimes called an isoperimetric problem because of its origins in maximizing the area of a surface for a given perimeter.\(^7\) The mathematical problem is to extremize the integral
\[
I = \int \left\{ H(u^2 + v^2) + g\eta^2 + \lambda(x, y)\left[(v_x - u_y) - f_0\eta/H\right]\right\} \, dA. \tag{3.155}
\]
where \( \lambda(x, y) \) is a Lagrange multiplier, undetermined at this stage. It is a function of space: if it were a constant, the integral would merely extremize energy subject to a given integral of potential vorticity, and rearrangements of potential vorticity (which here we wish to disallow) would leave the integral unaltered.

As there are three independent variables there are three Euler-Lagrange equations that must be solved in order to minimize \( I \). These are
\[
\frac{\partial L}{\partial h} - \frac{\partial}{\partial x} \frac{\partial L}{\partial h_x} - \frac{\partial}{\partial y} \frac{\partial L}{\partial h_y} = 0, \tag{3.156}
\]
where \( L \) is the integrand on the right-hand side of (3.155). Substituting this into (3.156) gives, after a little algebra,
\[
2g\eta - \lambda f_0 = 0, \quad 2u + \frac{\partial \lambda}{\partial y} = 0, \quad 2v - \frac{\partial \lambda}{\partial x} = 0, \tag{3.157}
\]
and then eliminating \( \lambda \) gives the simple relationships
\[
u = \frac{g}{f_0} \frac{\partial \eta}{\partial y}, \quad v = \frac{g}{f_0} \frac{\partial \eta}{\partial x}, \tag{3.158}
\]
which are the equations of geostrophic balance. Thus, in the linear approximation, geostrophic balance is a minimum energy state for a given field of potential vorticity.

### 3.9 ISENTROPIC COORDINATES

We now return to the continuously stratified primitive equations, and consider the use of potential density as a vertical coordinate. In practice this means using potential temperature in the atmosphere or buoyancy (density) in the ocean; such coordinate systems are generically called isentropic coordinates, and sometimes isopycnal coordinates if density is used. This may seem an odd thing to do but for adiabatic flow in particular the resulting equations of motion have an attractive form that aids
the interpretation of large-scale flow. The thermodynamic equation then becomes a statement for the conservation of the mass of fluid with a given value of potential density and, because the the flow of both the atmosphere and ocean is largely along isentropic surfaces, the momentum and vorticity equations have quasi-two-dimensional form.

The particular choice of vertical coordinate is determined by the form of the thermodynamic equation in the equation-set at hand; thus, if the thermodynamic equation is \( \frac{D\theta}{Dt} = \dot{\theta} \), we transform the equations from \((x, y, z)\) coordinates to \((x, y, \theta)\) coordinates. The material derivative in this coordinate system is

\[
\frac{D}{Dt} = \frac{\partial}{\partial t} + u \left( \frac{\partial}{\partial x} \right)_\theta + v \left( \frac{\partial}{\partial y} \right)_\theta + \frac{D\theta}{Dt} \frac{\partial}{\partial \theta},
\]

where the last term on the right-hand side is zero for adiabatic flow, and the two-dimensional velocity \( \mathbf{u} = (u, v) \) is parallel to the isentropes.

### 3.9.1 A hydrostatic Boussinesq fluid

In the simple Boussinesq equations (see the table on page 72), the buoyancy is the relevant thermodynamic variable. With hydrostatic balance the horizontal and vertical momentum equations are, in height coordinates,

\[
\frac{Du}{Dt} + f \times u = -\nabla b, \quad b = \frac{\partial \phi}{\partial z},
\]

where \( b \) is the buoyancy, the variable analogous to the potential temperature \( \theta \) of an ideal gas. The thermodynamic equation is

\[
\frac{Db}{Dt} = \dot{b},
\]

and because \( b = -g\delta \rho / \rho_0 \), isentropic coordinates are the same as isopycnal coordinates.

Using (2.143) the horizontal pressure gradient may be transformed to isentropic coordinates

\[
\left( \frac{\partial \phi}{\partial x} \right)_z = \left( \frac{\partial \phi}{\partial x} \right)_b - \left( \frac{\partial z}{\partial x} \right)_b \frac{\partial \phi}{\partial z} = \left( \frac{\partial \phi}{\partial x} \right)_b - b \left( \frac{\partial z}{\partial x} \right)_b = \left( \frac{\partial M}{\partial x} \right)_b,
\]

where

\[
M \equiv \phi - zb.
\]

Thus, the horizontal momentum equation becomes

\[
\frac{Du}{Dt} + f \times u = -\nabla b M.
\]

where the material derivative is given by (3.159), with \( b \) replacing \( \theta \). Using (3.163) the hydrostatic equation becomes

\[
\frac{\partial M}{\partial b} = -z.
\]
The mass continuity equation may be derived by noting that the mass element may be written as
\[ \delta m = \frac{\partial z}{\partial b} \delta b \delta x \delta y. \] (3.166)

The mass continuity equation, \( \frac{D}{Dt} \delta m = 0 \), becomes
\[ \frac{D}{Dt} \frac{\partial z}{\partial b} + \frac{\partial z}{\partial b} \nabla_3 \cdot \bm{v} = 0, \] (3.167)
where \( \nabla_3 \cdot \bm{v} = \nabla_b \cdot \bm{u} + \frac{\partial b}{\partial b} \delta \) is the three dimensional derivative of the velocity in isentropic coordinates. Eq. (3.167) may thus be written
\[ \frac{D}{Dt} \sigma + \sigma \nabla_b \cdot \bm{u} = -\sigma \frac{\partial b}{\partial b}, \] (3.168)

where \( \sigma = \frac{\partial z}{\partial b} \) is a measure of the thickness between two isentropic surfaces. Equations (3.164), (3.165) and (3.168) comprise a closed set, with dependent variables \( \bm{u}, M \) and \( z \) in the space of independent variables \( x, y \) and \( b \).

### 3.9.2 A hydrostatic ideal gas

Deriving the equations of motion for this system requires a little more work than in the Boussinesq case but the idea is the same. For an ideal gas in hydrostatic balance we have, using (1.112),
\[ \frac{\delta \theta}{\theta} = \frac{\delta T}{T} - \kappa \frac{\delta p}{p} = \frac{\delta T}{T} + \frac{\delta \Phi}{c_p T} = \frac{1}{c_p T} \delta M, \] (3.169)

where \( M = c_p T + \Phi \) is the 'Montgomery potential', equal to the dry static energy. (We use some of the same symbols as in the Boussinesq case to facilitate comparison, but their meanings are slightly different.) From this
\[ \frac{\partial M}{\partial \theta} = \Pi, \] (3.170)

where \( \Pi = \frac{c_p T}{\theta} = \frac{c_p (p/p_R)}{R/c_p} \) is the 'Exner function'. Equation (3.170) represents the hydrostatic relation in isentropic coordinates. Note also that \( M = \theta \Pi + \Phi \).

To obtain the an appropriate form for the horizontal pressure gradient first note that, in the usual height coordinates, it is given by
\[ \frac{1}{\rho} \nabla z p = \theta \nabla_\theta \Pi, \] (3.171)

where \( \Pi = c_p T/\theta \). Using (2.143) gives
\[ \theta \nabla_\theta \Pi = \theta \nabla_\theta \Pi - \frac{\theta \partial \Pi}{\theta} \nabla_\theta \Phi. \] (3.172)

Then, using the definition of \( \Pi \) and the hydrostatic approximation to help evaluate the vertical derivative, we obtain
\[ \frac{1}{\rho} \nabla z p = c_p \nabla_\theta T + \nabla_\theta \Phi = \nabla_\theta M. \] (3.173)
Thus, the horizontal momentum equation is
\[ \frac{Dv}{Dt} + f \times u = -\nabla \phi M. \quad (3.174) \]

As in the Boussinesq case the mass continuity equation may be derived by noting that the mass element may be written as
\[ \delta m = \frac{\partial p}{\partial \theta} \delta \theta \delta x \delta y. \quad (3.175) \]
The mass continuity equation, \( D\delta m/Dt = 0 \), becomes
\[ \frac{D}{Dt} \frac{\partial p}{\partial \theta} + \frac{\partial p}{\partial \theta} \nabla \cdot v = 0 \quad (3.176) \]
or
\[ \frac{D\sigma}{Dt} + \sigma \nabla \cdot u = -\sigma \frac{\partial \phi}{\partial \theta}, \quad (3.177) \]
where now \( \sigma \equiv \frac{\partial p}{\partial \theta} \) is a measure of the (pressure) thickness between two isentropic surfaces. Equations (3.170), (3.174) and (3.177) form a closed set, analogous to (3.165), (3.164) and (3.168).

### 3.9.3 Analogy to shallow water equations

The equations of motion in isentropic coordinates have an obvious analogy with the shallow water equations, and we may think of the shallow water equations to be a finite-difference representation of the primitive equations written in isentropic coordinates, or think of the latter as the continuous limit of the shallow water equations as the number of layers increases. For example, consider a two-isentropic-level representation of (3.170), (3.174) and (3.177), in which the lower boundary is an isentrope. A natural finite differencing gives
\[ -M_1 = -z_0 \Delta \theta_0 \quad (3.178a) \]
\[ M_1 - M_2 = -z_1 \Delta \theta_1, \quad (3.178b) \]
and the momentum equations for each layer become
\[ \frac{Du_1}{Dt} + f \times u_1 = -\Delta \theta_0 \nabla z_0 \quad (3.179a) \]
\[ \frac{Du_2}{Dt} + f \times u_2 = -\Delta \theta_0 \nabla z_0 - \Delta \theta_1 \nabla z_1. \quad (3.179b) \]
Together with the mass continuity equation for each level these are just like the two-layer shallow water equations (3.51). This means that results that one might easily derive for the shallow water equations will often have a continuous analog.

### 3.10 AVAILABLE POTENTIAL ENERGY

We now revisit the issue of the internal and potential energy in stratified flow, motivated by the following remarks. In adiabatic, inviscid flow the total amount of
If a stably stratified initial state with sloping isentropes (left) is adiabatically re-arranged then the state of minimum potential energy has flat isentropes, as on the right, but the amount of fluid contained between each isentropic surface is unchanged. The difference between the potential energies of the two states is the available potential energy. 

Fig. 3.11 If a stably stratified initial state with sloping isentropes (left) is adiabatically re-arranged then the state of minimum potential energy has flat isentropes, as on the right, but the amount of fluid contained between each isentropic surface is unchanged. The difference between the potential energies of the two states is the available potential energy.

energy is conserved, and there are conversions between internal energy, potential energy and kinetic energy. In an ideal gas the potential energy and the internal energy of a column extending throughout the atmosphere are in a constant ratio to each other — their sum is called the total potential energy. In a simple Boussinesq fluid, energetic conversions involve only the potential and kinetic energy, and not the internal energy. Yet, plainly, in neither a Boussinesq fluid nor an ideal gas can all the total potential energy in a fluid be converted to kinetic energy, for then all of the fluid would be adjacent to the ground and the fluid would have no thickness, which intuitively seems impossible. Given a state of the atmosphere or ocean, how much of its total potential energy is available for conversion to kinetic energy? In particular, because total energy is conserved only in adiabatic flow, we may usefully ask: how much potential energy is available for conversion to kinetic energy under an adiabatic re-arrangement of fluid parcels?

Suppose that at any given time the flow is stably stratified, but that the isentropes (or more generally the surfaces of constant potential density) are sloping, as in Fig. 3.11. The potential energy of the system would be reduced if the isentropes were flattened, for then heavier fluid would be moved to lower altitudes, with lighter fluid replacing it at higher altitudes. In an adiabatic re-arrangement the amount of fluid between the isentropes would remain constant, and a state with flat isentropes (meaning parallel to the geopotential surfaces) evidently constitutes a state of minimum total potential energy. The difference between the total potential energy of the fluid and the total potential energy after an adiabatic re-arrangement to a state in which the isentropic surfaces are flat is called the available potential energy, or APE. 

8
3.10 Available Potential Energy

3.10.1 A Boussinesq fluid

The potential energy of a column of Boussinesq fluid of unit area is given by

\[ P = \int_0^H b z \, dz = \int_0^H \frac{b}{2} \, dz^2. \]  

(3.180)

and the potential energy of the entire fluid is given by the horizontal integral of this. The minimum potential energy of the fluid arises after an adiabatic re-arrangement in which the isopycnals are flattened, and the resulting buoyancy is only a function of \( z \). The available potential energy is then the difference between the energy of the initial state and of this minimum state, and to obtain an approximate expression for this we first integrate (3.180) by parts to give

\[ P = -\int_0^{b_m} z^2 \, db, \]  

(3.181)

where \( b_m \) is the maximum value of \( b \) in the column. (We omit a constant of integration that cancels when the state of minimum potential energy is subtracted. Alternatively, take the upper limit of the \( z \)-integral to be \( z = 0 \) and at the lower limit, at \( z = -H \) say, take \( b = 0 \).) The minimum potential energy state arises when \( z \) is a function only of \( b \), \( z = Z(b) \) say. Because mass is conserved in the re-arrangement, \( Z \) is equal to the horizontally averaged value of \( z \) on a given isopycnal surface, \( \bar{z} \), and the surfaces \( z \) and \( \bar{z} \) thus define each other completely. The average available potential energy, per unit area, is then given by

\[ APE = \int_0^{b_m} (\bar{z}^2 - z^2) \, db = \int_0^{b_m} \bar{z}^2 \, db, \]  

(3.182)

where \( z = \bar{z} + z' \); that is \( z' \) is the height variation of an isopycnal surface. The available potential energy is thus proportional to the integral of the variance of the altitude of such a surface, and it is a positive-definite quantity. To obtain an expression in \( z \)-coordinates, we express the height variations on an isopycnal surface in terms of buoyancy variations on a constant-height surface by Taylor-expanding the height about its value on the isopycnal surface. Referring to Fig. 3.12 this gives

\[ z'(\bar{b}) = \bar{z} + \frac{\partial z}{\partial b} \bigg|_{b = \bar{b}} [\bar{b} - b(\bar{z})] \approx \bar{z} - \frac{\partial z}{\partial b} \bigg|_{b = \bar{b}} b', \]  

(3.183)

where \( b' = b(\bar{z}) - \bar{b} \) is the corresponding buoyancy perturbation on the \( z \) surface and \( \bar{b} \) is the average value of \( b \) on the \( \bar{z} \) surface. Furthermore, \( \partial z / \partial b \bigg|_{z = \bar{z}} \approx \partial z / \partial b \approx (\partial \bar{b} / \partial z)^{-1} \), and (3.183) thus becomes

\[ z' = z(\bar{b}) - \bar{z} \approx -b' \left( \frac{\partial \bar{z}}{\partial b} \right) \approx \frac{-b'}{\partial \bar{b} / \partial z}. \]  

(3.184)

where \( z' = z(b) - \bar{z} \) is the height perturbation of the isopycnal surface, from its average value. Using (3.184) in (3.182) we obtain an expression for the APE per unit area, to wit

\[ APE \approx \int_0^H \left( \frac{\bar{b}^2}{\partial \bar{b} / \partial z} \right) \, dz. \]  

(3.185)
The total APE of the fluid is the horizontal integral of this, and is thus proportional to the variance of the buoyancy on a height surface. We emphasize that APE is not defined for single column of fluid, for it depends on the variations of buoyancy over a horizontal surface. Note too that this derivation neglects the effects of topography; this, and the use of a basic state stratification, effectively restrict the use of (3.185) to a single ocean basin, and even for that the approximations used limit the accuracy of the expressions.

### 3.10.2 An ideal gas

The expression for the APE for an ideal gas is obtained, *mutatis mutandis*, in the same way as it was for a Boussinesq fluid and the trusting reader may skip directly to (3.193). The internal energy of an ideal gas column of unit area is given by

$$ I = \int_0^\infty c_v T \rho \, dz = \int_0^{p_s} \frac{c_v}{g} T \, dp, \quad (3.186) $$

where $p_s$ is the surface pressure, and the corresponding potential energy is given by

$$ P = \int_0^\infty \rho g z \, dz = \int_0^{p_s} z \, dp = \int_0^\infty p \, dz = \int_0^{p_s} \frac{R}{g} T \, dp. \quad (3.187) $$

In (3.186) we use hydrostasy, and in (3.187) the equalities make successive use of hydrostasy, an integration by parts, and hydrostasy and the ideal gas relation. Thus, the total potential energy is given by

$$ TPE \equiv I + P = \frac{c_p}{g} \int_0^{p_s} T \, dp. \quad (3.188) $$
Using the ideal gas equation of state we can write this as

\[
TPE = \frac{c_p}{\gamma} \int_0^{p_s} \left( \frac{p}{p_s} \right)^\kappa \gamma d\theta = \frac{c_p p_s}{\gamma(1 + \kappa)} \int_0^{\infty} \left( \frac{p}{p_s} \right)^{\kappa+1} d\theta, \tag{3.189}
\]

after an integration by parts. (We omit a term proportional to \(p_s^{\kappa+1}\) \(\gamma d\theta \) that arises in the integration by parts, because it plays no role in what follows.) The total potential energy of the entire fluid is equal to a horizontal integral of (3.189). The minimum total potential energy arises when the pressure in (3.189) a function only of \(\gamma\), \(p = P(\theta)\), where by conservation of mass \(P\) is the average value of the original pressure on the isentropic surface, \(P = \bar{p}\). The average available potential energy per unit area is then given by the difference between the initial state and this minimum, namely

\[
APE = \frac{c_p p_s}{\gamma(1 + \kappa)} \int_0^{\infty} \left[ \left( \frac{p}{p_s} \right)^{\kappa+1} - \left( \frac{\bar{p}}{p_s} \right)^{\kappa+1} \right] d\theta, \tag{3.190}
\]

which is a positive definite quantity. A useful approximation to this is obtained by expressing the right-hand side in terms of the variance of the potential temperature on a pressure surface. We first use the binomial expansion to expand \(p^{\kappa+1} = (\bar{p} + p')^{\kappa+1}\). Neglecting third and higher order terms (3.190) becomes

\[
APE = \frac{R p_s}{2 \gamma} \int_0^{\infty} \left( \frac{\bar{p}}{p_s} \right)^{\kappa+1} \left( \frac{p'}{\bar{p}} \right)^2 d\theta. \tag{3.191}
\]

The variable \(p' = p(\theta) - \bar{p}\) is a pressure perturbation on an isentropic surface, and is related to the potential temperature perturbation on an isobaric surface by [c.f., (3.184)]

\[
p' \approx -\theta' \frac{\partial \bar{p}}{\partial \theta} \approx -\theta' \frac{\partial \bar{p}}{\partial p}. \tag{3.192}
\]

where \(\theta' = \theta(p) - \theta(\bar{p})\) is the potential temperature perturbation on the \(\bar{p}\) surface. Using (3.192) in (3.191) we finally obtain

\[
APE = \frac{R p_s^{\kappa+1}}{2} \int_0^{p_s} p^{\kappa-1} \left(-\theta' \frac{\partial \bar{p}}{\partial p}\right)^{-1} \bar{p}^2 dp. \tag{3.193}
\]

The \(APE\) is thus proportional to the variance of the potential temperature on the pressure surface or, from (3.191), proportional to the variance of the pressure on an isentropic surface.

### 3.10.3 Use, interpretation, and the atmosphere and ocean

The potential energy of a fluid is reduced when the dynamics act to flatten the isentropes. Consider, for example, the earth’s atmosphere, with isentropes sloping upward toward the pole (Fig. 3.11 with the pole on the right). Flattening these isentropes amounts to a sinking of dense air and a rising of light air, and this reduction of potential energy leads to a corresponding production of kinetic energy. Thus, if
the dynamics is such as to reduce the temperature gradient between equator and pole by flattening the isentropes then APE is converted to KE by that process. A statistically steady state is achieved because the heating from the sun continually acts to restore the horizontal temperature gradient between equator and pole, so replenishing the pool of APE, and to this extent the large-scale atmospheric circulation acts like a heat engine.

It is a useful exercise to calculate the total potential energy of the atmosphere and ocean, the available potential energy and the kinetic energy (problem 3.15). One finds

\[ TPE \gg APE > KE \]  \hspace{1cm} (3.194)

with, very approximately, \( TPE \sim 100 APE \) and \( APE \sim 10 KE \). The first inequality should not surprise us (for it was this that lead us to define \( APE \) in the first instance), but the second is not obvious (and in fact the ratio is larger in the ocean). It is related to the fact that the instabilities of the atmosphere and ocean occur at a scale smaller than the size of the domain, and are unable to release all the potential energy that might be available. Understanding this more fully is the topic of chapters 6 and 9.

**Notes**

1. The algorithm to numerically solve these equations differs from that of the free-surface shallow water equations because the mass conservation equation can no longer be stepped forward in time. Rather, an elliptic equation for \( p_{lid} \) must be derived by eliminating time derivatives from from (3.22) using (3.21), and this then solved at each timestep.

2. After Margules (1903). Margules sought to relate the energy of fronts to their slope. In this same paper the notion of available potential energy arose.

3. The expression ‘form drag’ is also commonly used in aerodynamics, and the two usages are related. In aerodynamics, form drag is the force due to pressure difference between the front and rear of an object, or any other ‘form’, moving through a fluid. Aerodynamic form drag may include frictional effects between the wind and the surface itself, but this effect is omitted in most geophysical uses.

4. (Jules) Henri Poincaré (1854–1912) was a prodigious French mathematician, physicist and philosopher, regarded as one the greatest mathematicians living at the turn of the 20th century. He is remembered for his original work in algebra and topology, and in dynamical systems and celestial mechanics, obtaining many results in what would be called non-linear dynamics and chaos when these fields re-emerged some 60 years later — the notion of ‘sensitive dependence on initial conditions’, for example, is present in his work. He also obtained a number of the results of special relativity independently of Einstein, and worked on the theory of rotating fluids — hence the Poincaré waves of this chapter. He wrote extensively and successfully for the general public on the meaning, importance and philosophy of science. Among other things he discussed whether scientific knowledge was an arbitrary convention, a notion that remains discussed and controversial to this day. (His answer: ‘convention’, in part, yes; ‘arbitrary’, no.) He was a proponent of the role of intuition in mathematical and scientific progress, and did not believe that mathematics could ever be wholly reduced to pure logic.

6. As considered by Rossby (1938).

7. An introduction to variational problems may be found in Weinstock (1952) and a number of other textbooks. Applications to many traditional problems in mechanics are discussed by Lanczos (1970).

8. Margules (1903) introduced the concept of potential energy that is available for conversion to kinetic energy. Lorenz (1955) clarified its meaning and derived useful, approximate formulae for its computation. Shepherd (1993) showed that the APE is just the non-kinetic part of the pseudo-energy, an interpretation that naturally leads to a number of extensions of the concept. There are a host of other papers on the subject, including that of Huang (1998) who looked at some of the limitations of the approximate expressions in an oceanic context.

Further Reading
This remains a good reference for geostrophic adjustment and gravity waves. The time-dependent geostrophic adjustment problem is discussed in section 7.3.

Problems

3.1 Derive the appropriate shallow water equations for a single moving layer of fluid of density \( \rho_1 \) above a rigid floor, and where above the moving fluid is a stationary fluid of density \( \rho_0 \), where \( \rho_0 < \rho_1 \). Show that as \( \rho_0 / \rho_1 \rightarrow 0 \) the usual shallow water equations emerge.

3.2 (a) Model the atmosphere as two immiscible, ‘shallow water’ fluids of different density stacked one above the other. Using reasonable values for the values of any needed physical parameters, estimate the displacement of the interfacial surface associated with a pole–equator temperature gradient of 60 K.
(b) Similarly estimate an interfacial displacement in the ocean associated with a temperature gradient of 20 K over a distance of 4000 km. (This is a crude representation of the main oceanic thermocline.)

3.3 ♦ For a shallow water fluid the energy equation, (3.90), has the form \( \partial E / \partial t + \nabla \cdot (\mathbf{v} (E + gh^2/2)) = 0 \). But for a compressible fluid, the corresponding energy equation, (1.190), has the form \( \partial E / \partial t + \nabla \cdot (\mathbf{v} (E + p)) = 0 \). In a shallow water fluid, \( p \neq gh^2/2 \) at a point so these equations are superficially different. Explain this and reconcile the two forms. (Hint: What is the average pressure in a fluid column?)

3.4 ♦ Can the shallow water equations for an incompressible fluid be derived by way of an asymptotic expansion in the aspect ratio? If so, do it. That is, without assuming hydrostasy ab initio, expand the Euler equations with a free surface in small parameter equal to the ratio of the depth of the fluid to the horizontal scale of the motion, and so obtain the shallow water equations.

3.5 ♦ The inviscid shallow water equations, rotating or not, can support gravity waves of arbitrarily short wavelengths. For sufficiently high wavenumber, the wavelength will be shorter than the depth of the fluid. Is this consistent with an asymptotic nature of the shallow water equations? Discuss.

3.6 Show that the vertical velocity within a shallow-water system is given by

\[
\mathbf{w} = \frac{z - \eta_b}{h} \frac{Dh}{Dt} + \frac{D\eta_b}{Dt}. \tag{P3.1}
\]
Interpret this result, showing that it gives sensible answers at the top and bottom of the fluid layer.
3.7 What is the appropriate generalization of (3.99) to two-dimensions? Suppose that at
time $t = 0$ the height field is given by a Gaussian distribution $h' = A e^{-r^2/\sigma^2}$ where
$r^2 = x^2 + y^2$. What is the subsequent evolution of this, in the linear approximation?
Show that the distribution remains Gaussian, and that its width increases at speed $\sqrt{\frac{gH}{2}}$, where $H$ is the mean depth of the fluid.

3.8 In an adiabatic shallow water fluid in a rotating reference frame show that the potential
vorticity conservation law is
\[
\frac{D}{Dt} \zeta + f \eta - h_b = 0 \tag{P3.2}
\]
where $\eta$ is the height of the free surface and $h_b$ is the height of the bottom topogra-
phy, both referenced to the same flat surface.

(a) A cylindrical column of air at 30° latitude with radius 100 km expands horizontally
to twice its original radius. If the air is initially at rest, what is the mean tangential
velocity at the perimeter after the expansion.

(b) An air column at 60° N with zero relative vorticity ($\zeta = 0$) stretches from the
surface to the tropopause, which we assume is a rigid lid, at 10 km. The air
column moves zonally onto a plateau 2.5 km high. What is its relative vorticity?
Suppose it then moves southward to 30° N. What is its vorticity? (Assume density
is constant.)

3.9 ♦ In the long-wave limit of Poincaré waves, fluid parcels behave as free-agents; that is,
like free solid particles moving in a rotating frame unencumbered by pressure forces.
Why then, is their frequency given by $\omega = f = 2\Omega$ where $\Omega$ is the rotation rate of
the coordinate system, and not by $\Omega$ itself? Do particles that are stationary or move
in a straight line in the inertial frame of reference satisfy the dispersion relationship
for Poincaré waves in this limit? Explain. (See also Durran [1993], Egger [1999], Phillips
[2000].)

3.10 Linearize the $f$-plane shallow-water system about a state of rest. Suppose that there
is an initial disturbance that is given in the general form
\[
\eta = \int \hat{\eta}_{k,l} e^{i(kx + ly)} \, dk \, dl \tag{P3.3}
\]
where $\eta$ is the deviation surface height and the Fourier coefficients $\hat{\eta}_{k,l}$ are given,
and that the initial velocity is zero.

(a) Obtain the geopotential field at the completion of geostrophic adjustment, and
show that the deformation scale is a natural length scale in the problem.

(b) Show that the change in total energy during the adjustment is always less than or
equal to zero. Neglect any initial divergence.
N.B. Because the problem is linear, the Fourier modes do not interact.

3.11 Geostrophic adjustment of a velocity jump
Consider the evolution of the linearized $f$-plane shallow water equations in an infinite
domain. Suppose that initially the fluid surface is flat, the zonal velocity is zero and
the meridional velocity is given by
\[
v(x) = v_0 \text{sgn}(x) \tag{P3.4}
\]

(a) Find the equilibrium height and velocity fields at $t = \infty$.

(b) What are the initial and final kinetic and potential energies?
Partial Solution:
The potential vorticity is $q = \zeta - f_0 \eta / H$, so that the initial and final state is
\[
q = 2v_0 \delta(x). \tag{P3.5}
\]
The final state streamfunction is thus given by \((\partial^2/\partial x^2 - L_d^{-2}) \psi = q\), with solution \(\psi = \psi_0 \exp(x/L_d)\) and \(\psi = \psi_0 \exp(-x/L_d)\) for \(x < 0\) and \(x > 0\), where \(\psi_0 = L_d \nu_0\) (why?), and \(\eta = f_0 \psi / g\). The energy is \(E = \int (Hv^2 + g\eta^2)/2 \, dx\). The initial KE is infinite, the initial PE is zero, and the final state has \(PE = KE = gL_d \eta_0^2/4\) — that is, the energy is equipartitioned between kinetic and potential.

3.12 In the shallow water equations show that, if the flow is approximately geostrophically balanced, the energy at large scales is predominantly potential energy and that energy at small scales is predominantly kinetic energy. Define precisely what ‘large scale’ and ‘small scale’ mean in this context.

3.13 In the shallow-water geostrophic adjustment problem, show that at large scales the velocity adjusts to the height field, and that at small scales the height field adjusts to the velocity field.

3.14 ♦ Consider the problem of minimizing the full energy [i.e., \(\int (H u^2 + g \eta^2) \, dx\)], given the potential vorticity field \(q(x,y) = (\zeta + f)/h\). Show that the balance relations analogous to (3.8.5) are \(uh = -\partial (Bq^{-1})/\partial y\) and \(vh = \partial (Bq^{-1})/\partial x\) where \(B\) is the Bernoulli function \(B = g\eta + u^2/2\). Show that steady flow does not necessarily satisfy these equations. Discuss.

3.15 Using realistic values for temperature, velocity etc., calculate approximate values for the total potential energy, the available potential energy, and the kinetic energy, of either a hemisphere in the atmosphere or an ocean basin.
Vorticity and Potential Vorticity

VORTICITY AND POTENTIAL VORTICITY both play a central role in geophysical fluid dynamics — indeed, we shall find that the large scale circulation of the ocean and atmosphere is in large-part governed by the evolution of the latter. In this chapter we define and discuss these quantities and deduce some of their dynamical properties. Along the way we will come across Kelvin’s circulation theorem, one of the most fundamental conservation laws in all of fluid mechanics, and we will find that the conservation of potential vorticity is intimately tied to this.

4.1 VORTICITY AND CIRCULATION

4.1.1 Preliminaries

Vorticity, $\omega$, is defined to be the curl of velocity and so is given by

$$\omega \equiv \nabla \times v.$$  \hfill (4.1)

Circulation, $C$, is defined to be the integral of velocity around a closed fluid loop and so is given by

$$C \equiv \oint v \cdot dl = \int_S \omega \cdot dS$$ \hfill (4.2)

where the second expression uses Stokes’ theorem, where $S$ is any surface bounded by the loop. The circulation around some path is equal to the integral of the normal component of vorticity over any surface bounded by that path. The circulation is not a field like vorticity and velocity; rather, we think of the circulation around a particular material line of finite length, and so its value generally depends on the path chosen. If $\delta S$ is an infinitesimal surface element whose normal points in the
direction of the unit vector $\hat{n}$, then
\[ \hat{n} \cdot (\nabla \times \mathbf{v}) = \frac{1}{\delta S} \oint_{\delta l} \mathbf{v} \cdot d\mathbf{l} \]  
(4.3)

where the line integral is around the infinitesimal area. Thus at a point the component of vorticity in the direction of $\mathbf{n}$ is proportional to the circulation around the surrounding infinitesimal fluid element, divided by the elemental area bounded by the path of the integral. A heuristic test for the presence of vorticity is to imagine a small paddle wheel in the flow; the paddle wheel acts as a ‘circulation-meter’, and rotates if the vorticity is non-zero.

### 4.1.2 Simple axisymmetric examples
Consider axisymmetric motion in two dimensions, so that the flow is confined to a plane. We use cylindrical coordinates $(r, \phi, z)$ where $z$ is the direction perpendicular to the plane, with velocity components $(u^r, u^\phi, u^z)$. For axisymmetric flow $u^z = u^r = 0$ but $u^\phi \neq 0$.

**Rigid Body Motion**
The velocity distribution is given by
\[ u^\phi = \Omega r \]  
(4.4)

where $\Omega$ is the angular velocity of the fluid. Associated with this is the vorticity
\[ \omega = \nabla \times \mathbf{v} = \omega^z \mathbf{k}, \]  
(4.5)

where
\[ \omega^z = \frac{1}{r} \frac{\partial}{\partial r} (ru^\phi) = \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 \Omega) = 2\Omega. \]  
(4.6)

The vorticity of a fluid in solid body rotation is thus twice the angular velocity of the fluid, and is pointed in a direction orthogonal to the plane of rotation.

**The ‘vr’ vortex**
This vortex is so-called because the tangential velocity (historically denoted ‘$v$’ in this context) is such that the product $vr$ is constant. In our notation we would have
\[ u^\phi = \frac{K}{r}, \]  
(4.7)

where $K$ is a constant determining the vortex strength. Evaluating the $z$-component of vorticity gives
\[ \omega^z = \frac{1}{r} \frac{\partial}{\partial r} (ru^\phi) = \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{K}{r} \right) = 0, \]  
(4.8)

except where $r = 0$, at which the expression is singular and the vorticity is infinite. Our paddle wheel rotates when placed at the vortex center, but, less obviously, does not if placed elsewhere.

The circulation around a circle that encloses the origin is given by
\[ C = \oint \frac{K}{r} \, r \, d\phi = 2\pi K. \]  
(4.9)
4.2 The Vorticity Equation

This does not depend on the radius, and so it is true if the radius is infinitesimal. Since the vorticity is the circulation divided by the area, the vorticity at the origin must be infinite. Consider now an integration path that does not enclose the origin, for example the contour A–B–C–D–A in Fig. 4.1. Over the segments A–B and C–D the velocity is orthogonal to the contour, and so the contribution is zero. Over B–C and D–A we have

\[ C_{BC} = \frac{K}{r_2} \theta r_2 = K \phi, \quad C_{DA} = -\frac{K}{r_1} \theta r_1 = -K \phi. \] (4.10)

Adding these two expressions we see that net circulation around the contour \( C_{ABCD} \) is zero. If we shrink the integration path to an infinitesimal size then, within the path, by Stokes’ theorem, the vorticity is zero. We can of course place the path anywhere we wish, except surrounding the origin, and obtain this result. Thus the vorticity is everywhere zero, except at the origin.

4.2 THE VORTICITY EQUATION

Using the vector identity

\[ \nabla \times (a \times b) = (b \cdot \nabla) a - (a \cdot \nabla) b + a \nabla \cdot b - b \nabla \cdot a, \] (4.13)

implies that the second term on the left hand side of (4.12) may be written

\[ \nabla \times (\omega \times v) = (v \cdot \nabla) \omega - (\omega \cdot \nabla) v + \omega \nabla \cdot v - v \nabla \cdot \omega. \] (4.14)
Because vorticity is the curl of velocity its divergence vanishes and so (4.12) becomes

$$\frac{\partial \omega}{\partial t} + (\mathbf{v} \cdot \nabla) \omega = (\omega \cdot \nabla) \mathbf{v} - \omega \nabla \cdot \mathbf{v} + \frac{1}{\rho^2} (\nabla \rho \times \nabla p) + \nabla \times \mathbf{F}. \quad (4.15)$$

The divergence term may be eliminated with the aid of the mass-conservation equation to give

$$\frac{D\tilde{\omega}}{Dt} = (\tilde{\omega} \cdot \nabla) \mathbf{v} + \frac{1}{\rho^3} (\nabla \rho \times \nabla p) + \frac{1}{\rho} \nabla \times \mathbf{F}. \quad (4.16)$$

where $\tilde{\omega} \equiv \omega / \rho$. We will set $F = 0$ in most of what follows.

The third term on the right-hand side of (4.15), as well as the second term on the right-hand side of (4.16), is variously called the baroclinic term, the non-homentropic term, or the solenoidal term. The solenoidal vector is defined by

$$S \equiv \frac{1}{\rho^2} \nabla \rho \times \nabla p = -\nabla \alpha \times \nabla p \quad (4.17)$$

A solenoid is a tube directed perpendicular to both $\nabla \alpha$ and $\nabla p$, with elements of length proportional to $\nabla p \times \nabla \alpha$. If the isolines of $p$ and $\alpha$ are parallel to each other, then solenoids do not exist. This occurs when the density is a function only of pressure for then

$$\nabla \rho \times \nabla p = \nabla \rho \times \nabla \rho \frac{dp}{d\rho} = 0. \quad (4.18)$$

The solenoidal vector may also be written

$$S = -\nabla \eta \times \nabla T. \quad (4.19)$$

This follows most easily by first writing the momentum equation in the form $\partial \mathbf{v} / \partial t + \omega \times \mathbf{v} = T \nabla \eta - \nabla B$, and taking its curl (see problem 2.2). Evidently the solenoidal term vanishes if: (i) isolines of pressure and density are parallel; (ii) isolines of temperature and entropy are parallel; (ii) density or entropy or temperature or pressure are constant. A barotropic fluid has by definition $p = p(\rho)$ and therefore no solenoids. A baroclinic fluid is one for which $\nabla p$ is not parallel to $\nabla \rho$. From (4.16) we see that the baroclinic term must be balanced by terms involving velocity or its tendency and therefore, in general, a baroclinic fluid is a moving fluid, even in the presence of viscosity.

For a barotropic fluid the vorticity equation takes the simple form,

$$\frac{D\tilde{\omega}}{Dt} = (\tilde{\omega} \cdot \nabla) \mathbf{v}. \quad (4.20)$$

If the fluid is also incompressible, meaning that $\nabla \cdot \mathbf{v} = 0$, then we have the even simpler form,

$$\frac{D\omega}{Dt} = (\omega \cdot \nabla) \mathbf{v}. \quad (4.21)$$

The terms on the right-hand side of (4.20) or (4.21) are conventionally divided into ‘stretching’ and ‘tipping’ (or ‘tilting’) terms, and we return to these in section 4.3.1.
An integral conservation property

Consider a single Cartesian component in (4.15). Then, using superscripts to denote
components,
\[
\frac{\partial \omega_x}{\partial t} = -\nabla \cdot \mathbf{v} - \omega \nabla \cdot \mathbf{v} + (\mathbf{\omega} \cdot \nabla) v_x + S_x
\]
(4.22)
where \(S_x\) is the (x-component of) the solenoidal term. Eq. (4.22) may be written
as
\[
\frac{\partial \omega_x}{\partial t} + \nabla \cdot (\mathbf{v} \omega_x - \omega v_x) = S_x,
\]
(4.23)
and this implies the Cartesian tensor form of the vorticity equation, namely
\[
\frac{\partial \omega_i}{\partial t} + \frac{\partial}{\partial x_j} (v_j \omega_i - v_i \omega_j) = S_i,
\]
(4.24)
with summation over repeated indices. The tendency of vorticity is given by the
solenoidal term plus the divergence of a vector field, and thus if the former vanishes
the volume integrated vorticity can only be altered by boundary effects. In both
atmosphere and ocean the solenoidal term is important, but we will see in section
4.5 that a useful conservation law for a scalar quantity can still be obtained.

4.2.1 Two-dimensional flow

In two-dimensional flow the fluid is confined to a surface, and independent of the
third dimension normal to that surface. Let us initially stay in a Cartesian geometry,
and then two-dimensional flow is flow on a plane, and the velocity normal to the
plane, and the rate of change of any quantity normal to that plane, are zero. Let
the normal direction be the \(z\)-direction and then the velocity in the plane, denoted
by \(u\), is
\[
\mathbf{v} = u = u_i \mathbf{i} + v_j \mathbf{j}, \quad w = 0.
\]
(4.25)
Only one component of vorticity non-zero and this is given by
\[
\omega = k \left( \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right).
\]
(4.26)
That is, in two-dimensional flow the vorticity is perpendicular to the velocity. We
let \(\zeta = \omega z = \omega \cdot \mathbf{k}\). Both the stretching and tilting terms vanish in two-dimensional
flow, and the two-dimensional vorticity equation becomes, for incompressible flow,
\[
\frac{D\zeta}{Dt} = 0,
\]
(4.27)
where \(D\zeta/Dt = \partial \zeta/\partial t + \mathbf{u} \cdot \nabla \zeta\). That is, in two-dimensional flow vorticity is conserved following the fluid elements; each material parcel of fluid keeps its value of
vorticity even as it is being advected around. Furthermore, specification of the vor-
ticity completely determines the flow field. To see this, we use the incompressibility
condition to define a streamfunction \(\psi\) such that
\[
\mathbf{u} = -\frac{\partial \psi}{\partial y}, \quad \mathbf{v} = \frac{\partial \psi}{\partial x}, \quad \zeta = \nabla^2 \psi.
\]
(4.28a,b,c)
Given the vorticity, the Poisson equation (4.28c) can be solved for the streamfunction and the velocity fields obtained through (4.28a,b), and this process is called ‘inverting the vorticity’.

Numerical integration of (4.27) is then a process of time-stepping plus inversion. The vorticity equation may then be written as an advection equation for vorticity,

\[ \frac{\partial \zeta}{\partial t} + \mathbf{u} \cdot \nabla \zeta = 0 \]  

(4.29)

in conjunction with (4.28). The vorticity is stepped forward one time-step using a finite-difference representation of (4.29), and the vorticity inverted to obtain a velocity using (4.28). (The notion that complete or nearly complete information about the flow may be obtained by inverting one field plays an important role in geophysical fluid dynamics, as we will see later on.)

Two-dimensional flow is not restricted to a single, Cartesian plane, and we may certainly envision two-dimensional flow on the surface of a sphere; in this case the velocity normal to the spherical surface (the ‘vertical velocity’) vanishes, and the equations are naturally expressed in spherical coordinates. Nevertheless, vorticity (absolute vorticity if the sphere is rotating) is still conserved on parcels as they move over the spherical surface.

### 4.3 VORTICITY AND CIRCULATION THEOREMS

#### 4.3.1 The ‘frozen-in’ property of vorticity

Let us first consider some simple topological properties of the vorticity field and its evolution.\(^1\) We define a vortex-line to be a line drawn through the fluid which is everywhere in the direction of the local vorticity. This definition is analogous to that of a streamline, which is everywhere in the direction of the local velocity. A vortex tube is formed by the collection of vortex lines passing through a closed curve (Fig. 4.2). A material-line is just a line that connects material fluid elements. Suppose we draw a vortex line through the fluid; such a line obviously connects fluid elements and therefore defines a co-incident material line. As the fluid moves the material line deforms, and the vortex line also evolves in a manner determined by the equations of motion. A remarkable property of vorticity is that, for an unforced and inviscid barotropic fluid, the flow evolution is such that a vortex line remains co-incident with the same material line with which it was initially associated. Put another way, a vortex line always contains the same material elements — the vorticity is ‘frozen’ or ‘glued’ to the material fluid.

To prove this we consider how an infinitesimal material line element \( \delta l \) evolves, \( \delta l \) being the infinitesimal material element connecting \( l \) with \( l + \delta l \). The rate of change of \( \delta l \) following the flow is given by

\[ \frac{D\delta l}{Dt} = \frac{1}{\delta t} (\delta l(t + \delta t) - \delta l(t)), \]  

(4.30)

which follows from the definition of the material derivative in the limit \( \delta t \to 0 \). From the Taylor expansion of \( \delta l(t) \) and the definition of velocity it is also apparent that

\[ \delta l(t + \delta t) = l(t) + \delta l(t) + (\mathbf{v} + \delta \mathbf{v})\delta t - (l(t) + \mathbf{v}\delta t) = \delta l + \delta \mathbf{v}\delta t, \]  

(4.31)
as illustrated in Fig. 4.2. Substituting into (4.30) gives

\[ \frac{D\delta l}{Dt} = \delta v \]  

(4.32)

But since \( \delta v = (\delta l \cdot \nabla)v \) we have that

\[ \frac{D\delta l}{Dt} = (\delta l \cdot \nabla)v \]  

(4.33)

Comparing this with (4.16), we see that vorticity evolves in the same way as a line element. To see what this means, at some initial time we can define an infinitesimal material line element parallel to the vorticity at that location, that is,

\[ \delta l(x, t = 0) = A\omega(x, t = 0) \]  

(4.34)

where \( A \) is a constant. Then, for all subsequent times the magnitude of the vorticity of that fluid element, even as it moves to a new location \( x' \), remains proportional to the length of the fluid element at that point and is oriented in the same way; that is \( \omega(x', t) = A^{-1}\delta l(x', t) \).

To see a similar result in a slightly different way note that a vortex line element

\[ l + \delta l + \delta t (v + \delta v) \]

\[ v + \delta v \]

\[ l + \delta l + \delta t (v + \delta v) \]

\[ l + \delta l + \delta t \]

\[ v \]

\[ l + \delta l \]

\[ \delta l \]

\[ v \]

\[ l \]

**Fig. 4.3** Evolution of an infinitesimal material line \( \delta l \) from time \( t \) to time \( t + \delta t \). It can be seen from the diagram that \( D\delta l/Dt = \delta v \).
is determined by the condition \( \delta l = A \omega \) or, because \( A \) is just an arbitrary scaling factor, \( \omega \times \delta l = 0 \). Now, for any line element we have that

\[
\frac{D}{Dt} (\omega \times \delta l) = \frac{D\omega}{Dt} \times \delta l - \frac{D\delta l}{Dt} \times \omega. \tag{4.35}
\]

We also have that

\[
\frac{D\delta l}{Dt} = \delta v = \delta l \cdot \nabla v \tag{4.36a}
\]

and

\[
\frac{D\omega}{Dt} = \omega \cdot \nabla v. \tag{4.36b}
\]

If the line element is initially a vortex line element then, at \( t = 0 \), \( \delta l = A \omega \) and, using (4.36), the right hand side of (4.35) vanishes. Thus, the tendency of \( \omega \times \delta l \) is zero, and the vortex line continues to be a material line.

**Stretching and tilting**

The terms on the right-hand side of (4.20) or (4.21) may be interpreted in terms of ‘stretching’ and ‘tipping’ (or ‘tilting’). Consider a single Cartesian component of (4.21),

\[
\frac{D\omega^x}{Dt} = \omega^x \frac{\partial u}{\partial x} + \omega^y \frac{\partial u}{\partial y} + \omega^z \frac{\partial u}{\partial z}. \tag{4.37}
\]

The second and third terms of this are the tilting or tipping terms because they involve changes in the orientation of the vorticity vector. They tell us that vorticity in \( x \)-direction may be generated from vorticity in the \( y \)- and \( z \)-directions if the advection acts to tilt the material lines. Because vorticity is tied to these lines, vorticity oriented in one direction becomes oriented in another, as in Fig. 4.4.

The first term on the right-hand side of (4.37) is the stretching term, and it acts to intensify the \( x \)-component of vorticity if the velocity is increasing in the \( x \)-direction — that is, if the material lines are being stretched (Fig. 4.5). This effect arises because a vortex line is tied to a material line, and therefore vorticity is amplified in proportion to the stretching of material line aligned with it. This effect is important in tornadoes, to give one example. If the fluid is incompressible stretching of a fluid mass in one direction must be accompanied by convergence in another, and this leads to the conservation of circulation, as we now discuss.

### 4.3.2 Kelvin’s Circulation Theorem

Kelvin’s circulation theorem states that under certain circumstances the circulation around a material fluid parcel is conserved; that is, the circulation is conserved ‘following the flow’.² The primary restrictions are that body forces are conservative (i.e., they are representable as potential forces, and therefore that the flow be inviscid), and that the fluid is baroropic \( [i.e., p = p(\rho)] \). Of these, the latter is the more restrictive for geophysical fluids. The circulation in the theorem is defined with respect to an inertial frame of reference; specifically, the velocity in (4.41) is the velocity relative to an inertial frame. To prove the theorem, we begin with the inviscid momentum equation,

\[
\frac{Dv}{Dt} = -\frac{1}{\rho} \nabla p - \nabla \Phi, \tag{4.38}
\]
Figure 4.4 The tilting of vorticity. Suppose that vorticity, \( \omega \), is initially directed horizontally, as in the lower figure, so that \( \omega^z \), its vertical component, is zero. The material lines, and therefore the vortex lines also, are tilted by the positive vertical velocity \( W \), so creating a vertically oriented vorticity. This mechanism is important in creating vertical vorticity in the atmospheric boundary layer (and, one may show, the \( \beta \)-effect in large-scale flow).

Fig. 4.5 Stretching of material lines distorts the cylinder of fluid as shown. Vorticity is tied to material lines, and so is amplified in the direction of the stretching. However, because the volume of fluid is conserved, the end surfaces shrink, the material lines through the cylinder ends converge, and the integral of vorticity over a material surface (the circulation) remains constant, as discussed in section 4.3.2.
where $\nabla \Phi$ represents the conservative body forces on the system. Applying the material derivative to the circulation, (4.2), gives

$$\frac{DC}{Dt} = \frac{D}{Dt} \int v \cdot dr = \int \left( \frac{Dv}{Dt} \cdot dr + v \cdot dv \right)$$

$$= \int \left[ \left( -\frac{1}{\rho} \nabla p - \nabla \Phi \right) \cdot dr + v \cdot dv \right]$$

$$= \int -\frac{1}{\rho} \nabla p \cdot dr$$

(4.39)

using (4.38) and $D(\delta l)/Dt = \delta v$. The line integration is over a closed, material, circuit. The second and third terms on the second line vanish separately, because they are exact differentials integrated around a closed loop. The term on the last line vanishes if density is constant or, more generally, if pressure is a function of density alone in which case $\nabla p$ is parallel to $\nabla \rho$. To see this, note that

$$\int \frac{1}{\rho} \nabla p \cdot dr = \int_S \nabla \times \left( \frac{\nabla p}{\rho} \right) \cdot dA = \int_A -\nabla \rho \times \nabla p \rho^2 \cdot dA,$$

(4.40)

using Stokes's theorem where $A$ is any surface bounded by the path of the line integral, and this evidently vanishes identically if $p$ is a function of $\rho$ alone. The last term is the integral of the solenoidal vector, and if it is zero (4.39) becomes

$$\frac{DC}{Dt} \int v \cdot dr = 0.$$

(4.41)

This is Kelvin’s circulation theorem. In words, the circulation around a material loop is invariant for a barotropic fluid that is subject only to conservative forces. Using Stokes’ theorem, the circulation theorem may also be written

$$\frac{DC}{Dt} \int \omega \cdot dS = 0.$$

(4.42)

That is, the area-integral of the normal component of vorticity across any material surface is constant, under the same conditions. This form is both natural and useful, and it arises because of the way vorticity is tied to material fluid elements.

**Stretching and circulation**

Let us informally consider how vortex stretching and mass conservation work together to give the circulation theorem. Let the fluid be incompressible so that the volume of a fluid mass is constant, and consider a surface normal to a vortex tube, as in Fig. 4.5. Let the volume of a small material box around the surface be $\delta V$, the length of the material lines be $\delta l$ and the surface area be $\delta A$. Then

$$\delta V = \delta l \delta A.$$

(4.43)

Because of the frozen-in property, vorticity passing through the surface is proportional to the length of the material lines. That is $\omega \propto \delta l$, and

$$\delta V \propto \omega \delta A.$$

(4.44)
4.3 Vorticity and Circulation Theorems

The right-hand side is just the circulation around the surface. Now, if the corresponding material tube is stretched \( \delta l \) increases, but the volume, \( \delta V \), remains constant by mass conservation. Thus, the circulation given by the right-hand side of (4.44) also remains constant. In other words, because of the frozen-in property vorticity is amplified by the stretching, but the vortex lines get closer together in such a way that the product \( \omega \delta S \) remains constant and circulation is conserved.

4.3.3 Baroclinic flow and the solenoidal term

In baroclinic flow, the circulation is not generally conserved. and from (4.39) we have

\[
\frac{D C}{D t} = - \oint \nabla p \cdot d\ell = - \oint \frac{dp}{\rho},
\]  

(4.45)

and this is called the baroclinic circulation theorem.\(^3\) Noting the fundamental thermodynamic relation \( T \, d \eta = dI + p \, d\alpha \) we have

\[
\alpha \, dp = d(p\alpha) - T \, d\eta + dI,
\]  

(4.46)

so that the solenoidal term on the right-hand side of (4.45) may be written as

\[
S_o = - \oint \alpha \, dp = \oint T \, d\eta = - \oint \eta \, dT = -R \oint T \, d \log p,
\]  

(4.47)

where the last equality holds only for an ideal gas. Using Stokes's theorem, \( S_o \) can also be written as

\[
S_o = - \int_A \nabla \alpha \times \nabla p \cdot dA = - \int_A \left( \frac{\partial \alpha}{\partial T} \right)_p \nabla T \times \nabla p \cdot dA = \oint_A \nabla T \times \nabla \eta \cdot dA.
\]  

(4.48)

The rate of change of the circulation across a surface depends on the existence of this solenoidal term (Fig. 4.6 and, for an example, problem 4.6).

However, even if the solenoidal vector is in general non-zero, circulation is conserved if the material path is in a surface of constant entropy, \( \eta \), and if \( \frac{D \eta}{D t} = 0 \). In this case the solenoidal term vanishes and, because \( \frac{D \eta}{D t} = 0 \), entropy remains constant on that same material loop as it evolves. This result gives rise to the conservation of potential vorticity, discussed in section 4.5.
4.3.4 Circulation in a rotating frame

The absolute and relative velocities are related by \( v_a = v_r + \Omega \times r \) so that in a rotating frame the rate of change of circulation is given by

\[
\frac{D}{Dt} \oint (v_r + \Omega \times r) \cdot dr = \oint \left[ \left( \frac{Dv_r}{Dt} + \Omega \times v_r \right) \cdot dr + (v_r + \Omega \times r) \cdot dv_r \right]. \tag{4.49}
\]

But \( \oint v_r \cdot dv_r = 0 \) and, integrating by parts,

\[
\oint (\Omega \times r) \cdot dv_r = \oint \left[ d \left[ (\Omega \times r) \cdot v_r \right] - (\Omega \times dv_r) \cdot v_r \right] = \oint \left[ d \left[ (\Omega \times r) \cdot v_r \right] + (\Omega \times v_r) \cdot dr \right]. \tag{4.50}
\]

The first term is on the right-hand side is zero and so (4.49) becomes

\[
\frac{D}{Dt} \oint (v_r + \Omega \times r) \cdot dr = \oint \left( \frac{Dv_r}{Dt} + 2\Omega \times v_r \right) \cdot dr = -\oint \frac{dp}{\rho}, \tag{4.51}
\]

where the second equality uses the momentum equation. The term on the last line vanishes if the fluid is barotropic, and if so the circulation theorem is, unsurprisingly,

\[
\frac{D}{Dt} \int (\omega_r + 2\Omega) \cdot dS = 0, \tag{4.52a,b}
\]

where the second equation uses Stokes’s theorem and we have used \( \nabla \times (\Omega \times r) = 2\Omega \), and where \( \omega_r = \nabla \times v_r \) is the relative vorticity.4

4.3.5 The circulation theorem for hydrostatic flow

Kelvin’s circulation theorem holds for hydrostatic flow, with a slightly different form. For simplicity we restrict attention to the \( f \)-plane, and start with the hydrostatic momentum equations,

\[
\frac{Du_r}{Dt} + 2\Omega \times u_r = -\frac{1}{\rho} \nabla_z p, \quad 0 = -\frac{1}{\rho} \frac{\partial p}{\partial z} - \nabla \Phi, \tag{4.53a,b}
\]

where \( \Phi = gz \) is the gravitational potential and \( \Omega = \Omega \mathbf{k} \). The advecting field is three-dimensional, and in particular we still have \( D\delta r/Dt = \delta v = (\delta r \cdot \nabla) \mathbf{v} \). Thus, using (4.53) we have

\[
\frac{D}{Dt} \oint (u_r + \Omega \times r) \cdot dr = \oint \left[ \left( \frac{Du_r}{Dt} + \Omega \times v_r \right) \cdot dr + (u_r + \Omega \times r) \cdot dv_r \right] = \oint \left( \frac{Du_r}{Dt} + 2\Omega \times u_r \right) \cdot dr = \oint \left( \frac{1}{\rho} \nabla p - \nabla \Phi \right) \cdot dr, \tag{4.54}
\]

as with (4.51), having used \( \Omega \times v_r = \Omega \times u_r \), and where the gradient operator \( \nabla \) is three-dimensional. The last term on the right-hand side vanishes because it is the
integral of the gradient of a potential around a closed path. The first term vanishes if the fluid is barotropic, so that the circulation theorem is

$$\frac{D}{Dt} \oint (u_r + \Omega \times r) \cdot dr = 0, \quad (4.55)$$

Using Stokes’s theorem we have the equivalent form

$$\frac{D}{Dt} \int (\omega_{hy} + 2\Omega) \cdot dS = 0, \quad (4.56)$$

where the subscript ‘hy’ denotes hydrostatic and, in Cartesian coordinates,

$$\omega_{hy} = \nabla \times u = -i \frac{\partial v}{\partial z} + j \frac{\partial u}{\partial z} + k \left( \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right). \quad (4.57)$$

### 4.4 VORTICITY EQUATION IN A ROTATING FRAME

Perhaps the easiest way to derive the vorticity equation appropriate for a rotating reference frame is to begin with the momentum equation in the form

$$\frac{\partial v_r}{\partial t} + (2\Omega + \omega_r) \times v_r = -\frac{1}{\rho} \nabla p + \nabla \left( \Phi - \frac{1}{2} v_r^2 \right), \quad (4.58)$$

where the potential \( \Phi \) contains the gravitational and centrifugal forces. Take the curl of this and use the identity (4.13), which here implies

$$\nabla \times [(2\Omega + \omega_r) \times v_r] = (2\Omega + \omega_r) \nabla \cdot v_r + (v_r \cdot \nabla) (2\Omega + \omega_r) - (2\Omega + \omega_r) \cdot \nabla v_r, \quad (4.59)$$

(noting that \( \nabla \cdot (2\Omega + \omega) = 0 \)), to give the vorticity equation

$$\frac{D\omega_r}{Dt} = [(2\Omega + \omega) \cdot \nabla] v - (2\Omega + \omega_r) \nabla \cdot v_r + \frac{1}{\rho^2} \nabla (\nabla \rho \times \nabla p). \quad (4.60)$$

Note that because \( \Omega \) is a constant, \( D\omega_r/Dt = D\omega_a/Dt \) where \( \omega_a = 2\Omega + \omega_r \) is the absolute vorticity. The only difference between the vorticity equation in the rotating and inertial frames of reference is in the presence of the solid-body vorticity \( 2\Omega \) on the right-hand side. The second term on the right-hand side may be folded in to the material derivative using mass continuity, and after a little manipulation (4.60) becomes

$$\frac{D}{Dt} \left( \frac{\omega_a}{\rho} \right) = \frac{1}{\rho} (2\Omega + \omega_r) \cdot \nabla v_r + \frac{1}{\rho^3} (\nabla \rho \times \nabla p). \quad (4.61)$$

However, note that it is the absolute vorticity, \( \omega_a \), that now appears on the left-hand side. If \( \rho \) is constant, \( \omega_a \) may be replaced by \( \omega_r \).

#### 4.4.1 The circulation theorem and vortex tilting

What are the implications of the circulation theorem on a rotating, spherical planet? Let us define relative circulation over some material loop as

$$C_r \equiv \oint v_r \cdot dl, \quad (4.62)$$
The projection of a material circuit on to the equatorial plane. If a fluid element moves poleward, keeping its orientation to the local vertical fixed (e.g., it stays horizontal) then the area of its projection on to the equatorial plane increases. If its total (absolute) circulation is to be maintained, then the vertical component of the relative vorticity must diminish. That is, \[ \int_A (\omega + 2\Omega) \cdot dA = 2\Omega A_\perp \]

Therefore, the \( \beta \) term in
\[ \frac{D}{Dt}(\zeta + f) = \frac{D\zeta}{Dt} + \beta v = 0 \]
ultimately arises from the tilting of a parcel relative to the axis of rotation as it moves meridionally.

and because \( v_r = v_a - 2\Omega \times r \) we have
\[ C_r = C_a - \int 2\Omega \cdot dS = C_a - 2\Omega A_\perp \]
(4.63)

where \( C_a \) is the total or absolute circulation and \( A_\perp \) is the area enclosed by the projection of the material circuit onto the plane normal to the rotation vector; that is, onto the equatorial plane (see Fig. 4.7). If the solenoidal term is zero then the circulation theorem, (4.52), may be written as
\[ \frac{D}{Dt}(C_r + 2\Omega A_\perp) = 0. \]
(4.64)

This equation tells us that the relative circulation around a circuit will change if the orientation of the plane changes; that is, if the area of its projection on to the equatorial plane changes. In large scale dynamics the most common cause of this is when a fluid parcel changes its latitude. For example, consider the two-dimensional flow of an infinitesimal, horizontal, homentropic fluid surface at a latitude \( \vartheta \) with area \( \delta S \), so that the projection of its area on the equatorial plane is \( \delta S \sin \vartheta \). If the fluid surface moves, but remains horizontal, then directly from (4.64) the relative vorticity changes as
\[ \frac{D\zeta_r}{Dt} = -2\Omega \frac{D}{Dt} \sin \vartheta = -v_r \frac{2\Omega \cos \vartheta}{a} = -\beta v_r \]
(4.65)

where
\[ \beta \equiv \frac{df}{dy} = \frac{2\Omega}{a} \cos \vartheta, \]
(4.66)
The means by which the relative vorticity of a parcel changes by virtue of its latitudinal displacement is known as the beta effect, or the \( \beta \) effect. It is a manifestation of the tilting term in the vorticity equation, and it is often the most important means by which relative vorticity does change in large-scale flow. The \( \beta \) effect arises in the full vorticity equation, as we now see.

### 4.4.2 The vertical component of the vorticity equation

In large-scale dynamics, the most important, although not the largest, component of the vorticity is often the vertical one, because this contains much of the information about the horizontal flow. We can obtain an explicit expression for its evolution by taking the vertical component of (4.60), although care must be taken because the unit vectors \((i, j, k)\) are functions of position (see problem 2.5.)

An alternative derivation begins with the horizontal momentum equations

\[
\frac{\partial u}{\partial t} - v(\zeta + f) + w \frac{\partial u}{\partial z} = -\frac{1}{\rho} \frac{\partial p}{\partial x} - \frac{1}{2} \frac{\partial}{\partial x}(u^2 + v^2) + F_x
\]

\[
\frac{\partial v}{\partial t} + u(\zeta + f) + w \frac{\partial v}{\partial z} = -\frac{1}{\rho} \frac{\partial p}{\partial y} - \frac{1}{2} \frac{\partial}{\partial y}(u^2 + v^2) + F_y.
\]

where in this section we again drop the subscript \(r\) on variables measured in the rotating frame. Cross-differentiating gives, after a little algebra,

\[
\frac{D}{Dt}(\zeta + f) = -(\zeta + f) \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right) + \left( \frac{\partial u}{\partial z} \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \frac{\partial w}{\partial x} \right) + \frac{1}{\rho^2} \left( \frac{\partial \rho}{\partial x} \frac{\partial p}{\partial y} - \frac{\partial \rho}{\partial y} \frac{\partial p}{\partial x} \right) + \left( \frac{\partial F_y}{\partial x} - \frac{\partial F_x}{\partial y} \right).
\]

We interpret the various terms as follows:

\[\frac{D\zeta}{Dt} = \frac{\partial \zeta}{\partial t} + v \cdot \nabla \zeta: \] The material derivative of the vertical component of the vorticity.

\[\frac{Df}{Dt} = v \frac{\partial f}{\partial y} = v \beta: \] The beta-effect. The vorticity is affected by the meridional motion of the fluid, so that, apart from the terms on the right-hand side, \((\zeta + f)\) is conserved on parcels. Because the Coriolis parameter changes with latitude this is like saying that the system has differential rotation. This effect is precisely that due to the change in orientation of fluid surfaces with latitude, as given above in section [4.4.1](#) and Fig. [4.7](#).

\[-(\zeta + f)(\partial u/\partial x + \partial v/\partial y): \] The divergence term, which gives rise to vortex stretching. In an incompressible fluid this may be written \((\zeta + f)\partial w/\partial z\), so that vorticity is amplified if the vertical velocity increases with height, so stretching the material lines and the vorticity.

\[(\partial u/\partial z)(\partial w/\partial y) - (\partial v/\partial z)(\partial w/\partial x): \] The tilting term, whereby a vertical component of vorticity may be generated by a vertical velocity acting on a horizontal vorticity. See Fig. [4.4](#).

\[\rho^{-2} \left[ (\partial \rho/\partial x)(\partial p/\partial y) - (\partial \rho/\partial y)(\partial p/\partial x) \right] = \rho^{-2} J(\rho, p): \] The solenoidal term, also called the non-homentropic or baroclinic term, arising when isosurfaces of pressure and density are not parallel.
The forcing and friction term. If the only contribution to this is from molecular viscosity then this term is $\nu \nabla^2 \zeta$.

**Two-dimensional and shallow water vorticity equations**

In inviscid two-dimensional incompressible flow, all of the terms on the right-hand side of (4.68) vanish and we have the simple equation

$$\frac{D(\zeta + f)}{Dt} = 0,$$

(4.69)

implying that the absolute vorticity, $\zeta_a = \zeta + f$, is materially conserved. If $f$ is a constant, then (4.69) reduces to (4.29), and background rotation plays no role. If $f$ varies linearly with $y$, so that $f = f_0 + \beta y$, then (4.69) becomes

$$\frac{\partial \zeta}{\partial t} + u \cdot \nabla \zeta + \beta v = 0,$$

(4.70)

which is known as the two-dimensional $\beta$-plane vorticity equation.

For inviscid shallow water flow, we can show that (see chapter 3)

$$\frac{D(\zeta + f)}{Dt} = -(\zeta + f) \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \right).$$

(4.71)

In this equation the vanishing of the tilting term is perhaps the only aspect which is perhaps not immediately apparent, but this succumbs to a little thought.

**4.5 POTENTIAL VORTICITY CONSERVATION**

Too much of a good thing is wonderful.

Mae West (1892–1990).

Although Kelvin’s circulation theorem is a general statement about vorticity conservation, in its original form it is not always a practically useful statement for two reasons. First, it is a not a statement about a field, such as vorticity itself. Second, it is not satisfied for baroclinic flow, such as is found in the atmosphere and ocean. (Of course non-conservative forces such as viscosity also lead to circulation non-conservation, but this applies to virtually all conservation laws and does not diminish them.) It turns out that it is possible to derive a beautiful conservation law that overcomes both of these failings and one, furthermore, that is extraordinarily useful in geophysical fluid dynamics. This is the conservation of potential vorticity introduced first by Rossby and then in a more general form by Ertel. The idea is that we can use a scalar field that is being advected by the flow to keep track of, or to take care of, the evolution of fluid elements. For a baroclinic fluid this scalar field must be chosen in a special way (it must be a function of the density and pressure alone), but there is no restriction to barotropic fluid. Then using the scalar evolution equation in conjunction with the vorticity equation gives us a scalar conservation equation. In the next few subsections we derive the equation for potential vorticity conservation in a number of superficially different ways.
4.5 Potential Vorticity Conservation

4.5.1 PV conservation from the circulation theorem

**Barotropic fluids**

Let us begin with the simple case of a barotropic fluid. For an infinitesimal volume we write Kelvin’s theorem as:

$$\frac{D}{Dt} [(\mathbf{\omega}_a \cdot \mathbf{n}) \delta A] = 0$$  \hspace{1cm} (4.72)

where \(\mathbf{n}\) is a unit vector normal to an infinitesimal surface \(\delta A\). Now consider a volume bounded by two isosurfaces of values \(\chi\) and \(\chi + \delta \chi\), where \(\chi\) is any materially conserved tracer, so satisfying \(D\chi/Dt = 0\), so that \(\delta A\) initially lies in an isosurface of \(\chi\) (see Fig. 4.8). Since \(\mathbf{n} = \nabla\chi/|\nabla\chi|\) and the infinitesimal volume \(\delta V = \delta h \delta A\), where \(\delta h\) is the separation between the two surfaces, we have

$$\omega_a \cdot \mathbf{n} \delta A = \omega_a \cdot \frac{\nabla\chi \delta V}{|\nabla\chi| \delta h}.  \hspace{1cm} (4.73)$$

Now, the separation between the two surfaces, \(\delta h\) may be obtained from

$$\delta \chi = \delta x \cdot \nabla \chi = \delta h |\nabla \chi|,  \hspace{1cm} (4.74)$$

and using this in (4.72) we obtain

$$\frac{D}{Dt} \left[ \frac{(\omega_a \cdot \nabla \chi) \delta V}{\delta \chi} \right] = 0.  \hspace{1cm} (4.75)$$

Now, as \(\chi\) is conserved on material elements, then so is \(\delta \chi\), and it may be taken out of the differentiation. The mass of the volume element \(\rho \delta V\) is also conserved, so that (4.75) becomes

$$\frac{\rho \delta V}{\delta \chi} \frac{D}{Dt} \left( \frac{\omega_a}{\rho} \cdot \nabla \chi \right) = 0$$  \hspace{1cm} (4.76)

or

$$\frac{D}{Dt} (\tilde{\omega}_a \cdot \nabla \chi) = 0$$  \hspace{1cm} (4.77)

where \(\tilde{\omega}_a = \omega_a / \rho\). Eq. (4.77) is a statement of potential vorticity conservation for a barotropic fluid. The field \(\chi\) may be chosen arbitrarily, provided that it be materially conserved.
The general case

For a baroclinic fluid the above derivation fails simply because the statement of the conservation of circulation, \( (4.72) \), is not, in general, true: there are solenoidal terms on the right-hand side and from \((4.45)\) and \((4.47)\) we have

\[
\frac{D}{Dt} \left( (\omega_a \cdot n) \delta A \right) = S \cdot n \delta A, \quad S = -\nabla \alpha \times \nabla p = -\nabla \eta \times \nabla T. \tag{4.78a,b}
\]

However, the right-hand side of \((4.78a)\) may be annihilated by choosing the circuit around which we evaluate the circulation to be such that the solenoidal term is identically zero. Given the form of \(S\), this occurs if the values of any of \(p, \rho, \eta, T\) are constant on that circuit; that is, if \(\chi = p, \rho, \eta\) or \(T\). But the derivation also demands that \(\chi\) be a materially conserved quantity, which usually restricts the choice of \(\chi\) to be \(\eta\) (or potential temperature), or to be \(\rho\) itself if the thermodynamic equation is \(D\rho/Dt = 0\). Thus, the conservation of potential vorticity for inviscid, adiabatic flow is

\[
\frac{D}{Dt} \left( \tilde{\omega} a \cdot \nabla \theta \right) = 0 \tag{4.79}
\]

where \(D\theta/Dt = 0\). For diabatic flow source terms appear on the right-hand side, and we derive these later on. A summary of this derivation provided Fig. 4.9.

4.5.2 PV conservation from the frozen-in property

In this section we show that potential vorticity conservation is a consequence of the frozen-in property of vorticity. This is not surprising, because the circulation theorem itself has a similar origin. Thus, this derivation is not independent of the derivation in the previous section, just a minor re-expression of it. We first consider the case in which the solenoidal term vanishes from the outset.

Barotropic fluids

If \(\chi\) is a materially conserved tracer then the difference in \(\chi\) between two infinitesimally close fluid elements is also conserved and

\[
\frac{D}{Dt} (\chi_1 - \chi_2) = \frac{D\chi}{Dt} = 0. \tag{4.80}
\]

But \(\delta \chi = \nabla \chi \cdot \delta l\) where \(\delta l\) is the infinitesimal vector connecting the two fluid elements. Thus

\[
\frac{D}{Dt} (\nabla \chi \cdot \delta l) = 0 \tag{4.81}
\]

But since the line element and the vorticity (divided by density) obey the same equation, we can replace the line element by vorticity (divided by density) in \((4.81)\) to obtain again

\[
\frac{D}{Dt} \left( \frac{\nabla \chi \cdot \omega_a}{\rho} \right) = 0. \tag{4.82}
\]

That is, the potential vorticity, \(Q = (\tilde{\omega} a \cdot \nabla \chi)\) is a material invariant, where \(\chi\) is any scalar quantity that satisfies \(D\chi/Dt = 0\).
Mass: \( \rho \delta A \delta h = \text{constant} \)

Entropy: \( |\nabla \theta| \delta h = \text{constant} \)

Fig. 4.9 Geometry of potential vorticity conservation. The circulation equation is
\[
\frac{D[(\omega_a \cdot n) \delta A]}{Dt} = S \cdot n \delta A \quad \text{where} \quad S \propto \nabla \theta \times \nabla T.
\]
We choose \( n = \nabla \theta/|\nabla \theta| \), where \( \theta \) is materially conserved, to annihilate the solenoidal term on the right-hand side, and we note that \( \delta A = \delta V/\delta h \), where \( \delta V \) is the volume of the cylinder, and that \( \delta h = \delta \theta/|\nabla \theta| \). The circulation is
\[
C \equiv \omega_a \cdot n \delta A = \omega_a \cdot (\nabla \theta/|\nabla \theta|)(\delta V/\delta h) = [\rho^{-1} \omega_a \cdot \nabla \theta](\delta M/\delta \theta) = \rho \delta V \text{ is the mass of the cylinder.}
\]
As \( \delta M \) and \( \delta \theta \) are materially conserved, so is the potential vorticity \( \rho^{-1} \omega_a \cdot \nabla \theta \).

Baroclinic fluids

In baroclinic fluids we cannot casually substitute the vorticity for that of a line element in (4.81) because of the presence of the solenoidal term, and in any case a little more care would not be amiss. From (4.81) we obtain
\[
\delta l \cdot \frac{D\nabla \chi}{Dt} + \nabla \chi \cdot \frac{D\delta l}{Dt} = 0
\]
(4.83)
or, using (4.33),
\[
\delta l \cdot \frac{D\nabla \chi}{Dt} + \nabla \chi \cdot [(\delta l \cdot \nabla) \nu] = 0.
\]
(4.84)
Now, let us choose \( \delta l \) to correspond to a vortex line, so that at the initial time \( \delta l = \epsilon \vec{\omega}_a \). (Note that in this case the association of \( \delta l \) with a vortex line can only be made instantaneously, and we cannot set \( D\delta l/Dt \propto D\omega_a/Dt \).) Then,
\[
\vec{\omega}_a \cdot \frac{D\nabla \chi}{Dt} + \nabla \chi \cdot [(\omega_a \cdot \nabla) \nu] = 0,
\]
(4.85)
or, using the vorticity equation (4.16),
\[ \mathbf{\omega}_a \cdot \frac{D}{Dt} \nabla \chi + \nabla \chi \cdot \left[ \frac{D\mathbf{\omega}_a}{Dt} - \frac{1}{\rho^3} \nabla \rho \times \nabla p \right] = 0. \] (4.86)

This may be written
\[ \frac{D}{Dt} \mathbf{\omega}_a \cdot \nabla \chi = \frac{1}{\rho^3} \nabla \chi \cdot (\nabla \rho \times \nabla p). \] (4.87)

The term on the right-hand side is in general non-zero for an arbitrary choice of scalar, but it will evidently vanish if \( \nabla p, \nabla \rho \) and \( \nabla \chi \) are coplanar. If \( \chi \) is any function of \( p \) and \( \rho \) this will be satisfied, but \( \chi \) must also be a materially conserved scalar. If, as for an ideal gas, \( \rho = \rho(\eta, p) \) (or \( \eta = \eta(p, \rho) \)) where \( \eta \) is the entropy (which is materially conserved), and if \( \chi \) is a function of entropy \( \eta \) alone, then \( \chi \) satisfies both conditions. Explicitly, the solenoidal term vanishes because
\[ \nabla \chi \cdot (\nabla \rho \times \nabla p) = \frac{d\chi}{d\eta} \nabla \eta \cdot \left[ \left( \frac{\partial \rho}{\partial p} \nabla p + \frac{\partial \rho}{\partial \eta} \nabla \eta \right) \times \nabla p \right] = 0. \] (4.88)

Thus, provided \( \chi \) satisfies the two conditions
\[ \frac{D\chi}{Dt} = 0 \quad \text{and} \quad \chi = \chi(p, \rho), \] (4.89)
then (4.87) becomes
\[ \frac{D}{Dt} \left( \frac{\mathbf{\omega}_a \cdot \nabla \chi}{\rho} \right) = 0. \] (4.90)

The natural choice for \( \chi \) is potential temperature, whence
\[ \frac{D}{Dt} \left( \frac{\mathbf{\omega}_a \cdot \nabla \theta}{\rho} \right) = 0. \] (4.91)

The presence of a density term in the denominator is not necessary for incompressible flows (i.e., if \( \nabla \cdot \mathbf{v} = 0 \)).

### 4.5.3 PV conservation — an algebraic derivation

Finally, we give a algebraic derivation of potential vorticity conservation. We will take the opportunity to include frictional and diabatic processes, although these may also be included in the derivations above. We begin with the frictional vorticity equation in the form
\[ \frac{D\mathbf{\omega}_a}{Dt} = (\mathbf{\omega}_a \cdot \nabla) \mathbf{v} + \frac{1}{\rho^3} (\nabla \rho \times \nabla p) + \frac{1}{\rho} (\nabla \times F). \] (4.92)

where \( F \) represents any nonconservative force term on the right-hand side of the momentum equation (i.e., \( \frac{D\mathbf{v}}{Dt} = -\rho^{-1} \nabla p + F \)). We have also the equation for our materially conserved scalar \( \chi \),
\[ \frac{D\chi}{Dt} = \dot{\chi}. \] (4.93)
where $\dot{\chi}$ represents any sources and sinks of $\chi$. Now

$$\frac{D}{Dt} (\vec{\omega}_a \cdot \nabla) = \frac{D}{Dt} \vec{\omega}_a \cdot \nabla \chi + [(\vec{\omega}_a \cdot \nabla) v] \cdot \nabla \chi. \quad (4.94)$$

which may be obtained just by expanding the left-hand side. Thus, using (4.93),

$$\frac{D}{Dt} \vec{\omega}_a \cdot \nabla \chi = (\vec{\omega}_a \cdot \nabla) \dot{\chi} - [(\vec{\omega}_a \cdot \nabla) v] \cdot \nabla \chi. \quad (4.95)$$

Now take the dot product of (4.92) with $\nabla \chi$:

$$\nabla \chi \cdot \frac{D}{Dt} \vec{\omega}_a = \nabla \chi \cdot [(\vec{\omega}_a \cdot \nabla) v] + \nabla \chi \cdot \left[ \frac{1}{\rho^3} (\nabla \rho \times \nabla p) \right] + \nabla \chi \cdot \left[ \frac{1}{\rho} (\nabla \times F) \right]. \quad (4.96)$$

The sum of the last two equations yields

$$\frac{D}{Dt} (\vec{\omega}_a \cdot \chi) = \vec{\omega}_a \cdot \chi + \nabla \chi \cdot \left[ \frac{1}{\rho^3} (\nabla \rho \times \nabla p) \right] + \frac{\nabla \chi}{\rho} \cdot (\nabla \times F). \quad (4.97)$$

This equation reprises (4.87), but with the addition of frictional and diabatic terms. As before, the solenoidal term is annihilated if we choose $\chi = \theta(p, \rho)$, so giving the evolution equation for potential vorticity in the presence of forcing and diabatic terms, namely

$$\frac{D}{Dt} (\vec{\omega}_a \cdot \nabla \theta) = \vec{\omega}_a \cdot \chi + \frac{\nabla \theta}{\rho} \cdot (\nabla \times F). \quad (4.98)$$

### 4.5.4 Effects of salinity and moisture

For seawater the equation of state may be written as

$$\theta = \theta(\rho, p, S) \quad (4.99)$$

where $\theta$ is potential temperature and $S$ is salinity. In the absence of diabatic terms (which include saline diffusion) potential temperature is a materially conserved quantity. However, because of the presence of salinity, $\theta$ cannot be used to annihilate the solenoidal term; that is

$$\nabla \theta \cdot (\nabla \rho \times \nabla p) = \left( \frac{\partial \theta}{\partial S} \right)_{p, \rho} \nabla S \cdot (\nabla \rho \times \nabla p) \neq 0. \quad (4.100)$$

Strictly speaking then, there is no potential vorticity conservation principle for seawater. However, such a blunt statement rather overemphasizes the importance of salinity in the ocean, and the nonconservation of potential vorticity because of this effect is rather small.

In a moist atmosphere in which condensational heating occurs there is no ‘moist potential vorticity’ that is generally materially conserved.\(^7\) We may choose to define a moist PV ($Q_e$ say) based on moist equivalent potential temperature but it does not always obey $DQ_e/Dt = 0$.\(^7\)
4.5.5 Effects of rotation, and summary remarks

In a rotating frame the potential vorticity conservation equation is obtained simply by replacing \( \mathbf{\omega}_a \) by \( \mathbf{\omega} + 2\mathbf{\Omega} \), where \( \mathbf{\Omega} \) is the rotation rate of the rotating frame. The operator \( \frac{D}{Dt} \) is reference-frame invariant, and so may be evaluated using the usual formulae with velocities measured in the rotating frame.

In the above derivations, we have generally referred to the quantity \( \mathbf{\omega}_a \cdot \nabla \theta / \rho \) as potential vorticity; however, it is clear that this form is not unique. If \( \theta \) is a materially conserved variable, then so is \( g(\theta) \) where \( g \) is any function, so that \( \mathbf{\omega}_a \cdot \nabla g(\theta) / \rho \) is also a potential vorticity, although when such a non-standard definition is used we qualify the expression ‘potential vorticity’ with some adjective.

The conservation of potential vorticity has profound consequences in fluid dynamics, especially in a rotating, stratified fluid. The nonconservative terms are often small, and large-scale flow in both the ocean and atmosphere is characterized by conservation of potential vorticity. Such conservation is a very powerful constraint on the flow, and indeed it turns out that potential vorticity is a much more useful quantity for baroclinic, or nonhomentropic fluids than for barotropic fluids, because the required use of a special conserved scalar imparts additional information. A large fraction of the remainder of this book explores, in one way or another, the consequences of potential vorticity conservation.

4.6 * POTENTIAL VORTICITY IN THE SHALLOW WATER SYSTEM

In chapter 3 we derived potential vorticity conservation by direct manipulation of the shallow water equations. We now show that shallow water potential vorticity is also derivable from the conservation of circulation. Specifically, we will begin with the three-dimensional form of Kelvin’s theorem, and then make the small aspect ratio assumption (which is the key assumption underlying shallow water dynamics), and thereby recover shallow water potential vorticity conservation. In the following two subsections we give two variants of such derivation (see also Fig. 4.10).

4.6.1 Using Kelvin’s theorem

We begin with

\[
\frac{D}{Dt} (\mathbf{\omega}_3 \cdot \delta S') = 0, \tag{4.101}
\]

where \( \mathbf{\omega}_3 \) is the curl of the three-dimensional velocity and \( \delta S' = n \delta S \) is an arbitrary infinitesimal vector surface element, with \( n \) a unit vector pointing in the direction normal to the surface. If we separate the vorticity and surface element into vertical and horizontal components we can write (4.101) as

\[
\frac{D}{Dt} \left[ (\zeta + f) \delta A + \mathbf{\omega}_h \cdot \delta S_h \right] = 0 \tag{4.102}
\]

where \( \mathbf{\omega}_h \) and \( \delta S_h \) are the horizontally-directed components of the vorticity and the surface element, and \( \delta A = k \delta S \) is the area of a horizontal cross-section of a fluid column. In Cartesian form the horizontal component of the vorticity is

\[
\mathbf{\omega}_h = i \left( \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \right) - j \left( \frac{\partial w}{\partial x} - \frac{\partial v}{\partial z} \right) = i \frac{\partial w}{\partial y} - j \frac{\partial w}{\partial x}. \tag{4.103}
\]
4.6 Potential Vorticity in the Shallow Water System

The mass of a column of fluid, $hA$, is conserved in the shallow water system. Furthermore, the vorticity is tied to material lines so that $\zeta A$ is also a material invariant, where $\zeta = \omega \cdot k$ is the vertical component of the vorticity. From this, $\zeta / h$ must be materially conserved; that is $D(\zeta / h)/Dt = 0$, which is the conservation of potential vorticity in a shallow water system. In a rotating system this generalizes to $D[(\zeta + f)/h]/Dt = 0$.

where vertical derivatives of the horizontal velocity are zero by virtue of the nature of the shallow water system. Now, the vertical velocity in the shallow water system is smaller than the horizontal velocity by the order of the aspect ratio — the ratio of the fluid depth to the horizontal scale of motion. Furthermore, the size of the horizontally-directed surface element is also an aspect-ratio smaller than the vertically-directed component. That is,

$$|\omega_h| \sim \alpha |\zeta| \quad \text{and} \quad |\delta S_h| \sim \alpha |\delta A|,$$  \hspace{1cm} (4.104)

where $\alpha = H/L$ is the aspect ratio. Thus $\omega_h \cdot \delta S_h$ is an aspect-number squared smaller than the term $\zeta \delta A$ and in the small aspect ratio approximation should be neglected. Kelvin’s circulation theorem, (4.102) becomes

$$\frac{D}{Dt}[(\zeta + f)\delta A] = 0 \quad \text{or} \quad \frac{D}{Dt} \left[ \frac{(\zeta + f)}{h} h \delta A \right] = 0,$$  \hspace{1cm} (4.105a,b)

where $h$ is the depth of the fluid column. But $h \delta A$ is the volume of the fluid column, and this is constant. Thus, (4.105b) gives, as in (3.80),

$$\frac{D}{Dt} \left( \frac{\zeta + f}{h} \right) = 0,$$  \hspace{1cm} (4.106)

where, because horizontal velocities are independent of the vertical coordinate, the advection is purely horizontal.

4.6.2 Using an appropriate scalar field

In a constant density fluid we can write potential vorticity conservation as

$$\frac{D}{Dt}(\omega_3 \cdot \nabla \chi) = 0,$$  \hspace{1cm} (4.107)
where \( \chi \) is any materially-conserved scalar [c.f. (4.77) or (4.82)]. In the flat-bottomed shallow water system, a useful choice of scalar is the ratio \( z/h \), where \( h \) is the local thickness of the fluid column because, from (3.28),

\[
\frac{D}{Dt} \left( \frac{z}{h} \right) = 0,
\]

if (for simplicity) the fluid is flat-bottomed. With this choice of scalar, potential vorticity conservation becomes

\[
\frac{D}{Dt} \left[ \omega \cdot \nabla \left( \frac{z}{h} \right) \right] = 0,
\]

where \( \omega \) and \( D/Dt \) are fully three dimensional. Expanding the dot product gives

\[
\frac{D}{Dt} \left[ \zeta + \frac{f}{h} - \frac{z}{h^2} \omega_h \cdot \nabla z \right] = 0.
\]

For an order-unity Rossby number, the ratio of the size of the two terms in this equation is

\[
\frac{|\zeta|}{|(z/h)\omega_h \cdot \nabla_h h|} \sim \frac{[U/L]}{[WH/L^2]} = \frac{UL}{WH} = \alpha^2 \ll 1.
\]

Thus, the second term in (4.110) is an aspect-ratio squared smaller than the first and, upon its neglect, (4.106) is recovered.

### 4.7 Potential Vorticity in Approximate, Stratified Models

If approximate models of stratified flow — Boussinesq, hydrostatic and so on — are to be useful then they should conserve an appropriate form potential vorticity, and we consider a few such cases here.

#### 4.7.1 The Boussinesq equations

A Boussinesq fluid is incompressible (that is, the volume of a fluid element is conserved, and \( \nabla \cdot \mathbf{v} = 0 \)) and the equation for vorticity itself is isomorphic to that for a line element. However, the Boussinesq equations are not barotropic — \( \nabla \rho \) is not parallel to \( \nabla p \) — and although the pressure gradient term \( \nabla \phi \) disappears on taking its curl (or equivalently disappears on integration around a closed path) the buoyancy term \( \mathbf{k} b \) does not, and it is this that prevents Kelvin’s circulation theorem from holding. Specifically, the evolution of circulation in the Boussinesq equations obeys

\[
\frac{D}{Dt} [(\omega_a \cdot \mathbf{n})\delta A] = (\nabla \times \mathbf{b} k) \cdot \mathbf{n} \delta A,
\]

where here, as in (4.72), \( \mathbf{n} \) is a unit vector orthogonal to an infinitesimal surface element of area \( \delta A \). The right-hand side is annihilated if we choose \( \mathbf{n} \) to be parallel to \( \nabla b \), because \( \nabla b \cdot \nabla \times (b k) = 0 \). In the simple Boussinesq equations the thermodynamic equation is

\[
\frac{Db}{Dt} = 0,
\]

(4.113)
and potential vorticity conservation is therefore (with $\omega_\alpha = \omega + 2\Omega$)

$$\frac{DQ}{Dt} = 0, \quad Q = (\omega + 2\Omega) \cdot \nabla b.$$  \hfill (4.114a,b)

Expanding (4.114b) in Cartesian coordinates with $\Omega = f k$ we obtain:

$$Q = (v_x - u_y) b_z + (w_y - v_z) b_x + (u_z - w_x) b_y + f b_z.$$  \hfill (4.115)

In the general Boussinesq equations $b$ itself is not materially conserved. We cannot expect to obtain a conservation law if salinity is present, but if the equation of state and the thermodynamic equation are:

$$b = b(\theta, z), \quad \frac{D\theta}{Dt} = 0,$$  \hfill (4.116)

then potential vorticity conservation follows, because taking $n$ to be parallel to $\nabla \theta$ will cause the right-hand side of (4.112) to vanish. That is,\footnote{\textbf{Note:} This equation is repeated in the text.}

$$\nabla \theta \cdot \nabla \times (b k) = \left( \frac{\partial \theta}{\partial z} \nabla z + \frac{\partial \theta}{\partial b} \nabla b \right) \cdot \nabla \times (b k) = 0.$$  \hfill (4.117)

The materially conserved potential vorticity is thus

$$Q = \omega_\alpha \cdot \nabla \theta.$$  \hfill (4.118)

Note that if the equation of state is $b = b(\theta, \phi)$, where $\phi$ is the pressure, then potential vorticity is not conserved because then, in general, $\nabla \phi \cdot \nabla \times (b k) \neq 0$.

4.7.2 The hydrostatic equations

Making the hydrostatic approximation has no effect on the satisfaction of the circulation theorem. Thus, in a baroclinic hydrostatic fluid we have

$$\frac{D}{Dt} \left( (\omega_{hy} + 2\Omega) \cdot \nabla b \right) = -\nabla \times \nabla \alpha \cdot \nabla p \cdot \nabla b,$$  \hfill (4.119)

where, from \textbf{4.5}, $\omega_{hy} = \nabla \times \mathbf{u} = -iv_z + ju_z + \mathbf{k}(v_x - u_y)$, but the gradient operator and material derivative are fully three-dimensional. Derivation of potential vorticity conservation then proceeds, as in section 4.5.1, by choosing the circuit over which the circulation is calculated to be such that the right-hand side vanishes; that is, to be such that the solenoidal term is annihilated. Precisely as before, this occurs if the circuit is barotropic and without further ado we write

$$\frac{DQ_{hy}}{Dt} = \frac{D}{Dt} \left( \frac{(\omega_{hy} + 2\Omega) \cdot \nabla \theta}{\rho} \right) = 0.$$  \hfill (4.120)

Expanding this gives in Cartesian coordinates

$$Q_{hy} = \frac{1}{\rho} \left[ (v_x - u_y) \theta z - v_x \theta_x + u_z \theta_y + 2\Omega \theta_\rho \right].$$  \hfill (4.121)

In spherical coordinates the hydrostatic approximation is usually accompanied by the traditional approximation and the expanded expression for a conserved potential vorticity is more complicated. It can still be derived from Kelvin’s theorem, but this is left as an exercise for the reader (problem 4.4).
4.7.3 Potential Vorticity on isentropic surfaces

If we begin with the primitive equations in isentropic coordinates then potential vorticity conservation follows quite simply. Cross differentiating the horizontal momentum equations \( (3.164) \) gives the vorticity equation \( \text{c.f.} \( (3.72) \) \)

\[
\frac{D}{Dt}(\zeta + f) + (\zeta + f) \nabla_\theta \cdot \mathbf{u} = 0. \tag{4.122}
\]

where \( \frac{D}{Dt} = \partial / \partial t + \mathbf{u} \cdot \nabla_\theta \). The thermodynamic equation is

\[
\frac{D\sigma}{Dt} + \sigma \nabla \cdot \mathbf{u} = 0, \tag{4.123}
\]

where \( \sigma = \partial z / \partial b \) (Boussinesq) or \( \partial p / \partial \theta \) (ideal gas) is the thickness of an isopycnal layer. Eliminating the divergence between \( (4.122) \) and \( (4.123) \) gives

\[
\frac{D}{Dt} \left( \frac{\zeta + f}{\sigma} \right) = 0. \tag{4.124}
\]

The derivation, and the result, are precisely the same as with the shallow water equations (sections \( 3.6.1 \) and \( 4.6 \)).

A connection between isentropic and height coordinates

The hydrostatic potential vorticity written in height coordinates may be transformed into a form that reveals its intimate connection with isentropic surfaces. Let us make the Boussinesq approximation for which the potential vorticity is

\[
Q_{\text{hy}} = (v_x - u_y) b_z - v_z b_x + u_z b_y, \tag{4.125}
\]

where \( b \) is the buoyancy. We can write this as

\[
Q_{\text{hy}} = b_z \left[ (v_x - v_z b_x) - (u_y - u_z b_y) \right]. \tag{4.126}
\]

But the terms in the inner brackets are just the horizontal velocity derivatives at constant \( b \). To see this, note that

\[
\left( \frac{\partial v}{\partial x} \right)_b = \left( \frac{\partial v}{\partial x} \right)_z + \frac{\partial v}{\partial z} \left( \frac{\partial z}{\partial x} \right)_b = \left( \frac{\partial v}{\partial x} \right)_z - \frac{\partial v}{\partial z} \left( \frac{\partial b}{\partial x} \right)_z / \frac{\partial b}{\partial z}, \tag{4.127}
\]

with a similar expression for \( (\partial u / \partial y)_b \). (These relationships follow from standard rules of partial differentiation. Derivatives with respect to \( z \) are taken at constant \( x \) and \( y \).) Thus, we obtain

\[
Q_{\text{hy}} = \frac{\partial b}{\partial z} \left[ \left( \frac{\partial v}{\partial x} \right)_b - \left( \frac{\partial u}{\partial y} \right)_b \right] = \frac{\partial b}{\partial z} \zeta b. \tag{4.128}
\]

Thus, potential vorticity is simply the horizontal vorticity evaluated on a surface of constant buoyancy, multiplied by the vertical derivative of buoyancy, a measure of static stability. An analogous derivation, with a similar result, proceeds for the ideal gas equations, with potential temperature replacing buoyancy.
4.8 THE IMPERMEABILITY OF ISENTROPES TO POTENTIAL VORTICITY

A kinematical result is a result valid forever.

An interesting property of isentropic surfaces is that they are ‘impermeable’ to potential vorticity, meaning that the mass integral of potential vorticity ($\int Q\rho\,dV$) over a volume bounded by an isentropic surface remains constant, even in the presence of diabatic sources, provided the surfaces do not intersect a non-isentropic surface like the ground. This may seem surprising, especially because unlike most conservation laws the result does not require adiabatic flow, and for that reason it leads to interesting interpretations of a number of phenomena. However, at the same time impermeability is a consequence of the definition of potential vorticity rather than the equations of motion, and in that sense is a kinematic property.

To derive the result we define $s = \rho Q = \nabla \cdot (\theta \omega_a)$ and integrate over some volume $V$ to give

$$I = \int_V s\,dV = \int_V \nabla \cdot (\theta \omega_a)\,dV = \int_S \theta \omega_a \cdot dS,$$

using the divergence theorem, where $S$ is the surface surrounding the volume $V$. If this is an isentropic surface then we have

$$I = \theta \int_S \omega_a \cdot dS = \theta \int_V \nabla \cdot \omega_a\,dV = 0,$$

again using the divergence theorem. That is, over a volume wholly enclosed by a single isentropic surface the integral of $s$ vanishes. If the volume is bounded by more than one isentropic surface neither of which intersect the surface, for example by concentric spheres of different radii as in Fig. 4.11a, the result still holds. The quantity $s$ is called ‘potential vorticity concentration’, or ‘PV concentration’. The integral of $s$ over a volume is akin to the total amount of a conserved material property, like salt content, and so may be called ‘PV substance’. That is, the PV concentration is the amount of potential vorticity substance per unit volume (following the meaning for concentration introduced in section 1.2.3 on page 11) and

$$\text{PV substance} = \int s\,dV = \int \rho Q\,dV.$$  

(4.131)

Suppose now that fluid volume is enclosed by an isentrope that intersects the ground, as in Fig. 4.11b. Let $A$ denote the isentropic surface, $B$ denote the ground, $\theta_A$ the constant value of $\theta$ on the isentrope, and $\theta_B(x, y, t)$ the non-constant value of $\theta$ on the ground. The integral of $s$ over the volume is then

$$I = \int_V \nabla \cdot (\theta \omega_a)\,dV = \theta_A \int_A \omega_a \cdot dS + \int_B \theta_B \omega_a \cdot dS$$

$$= \theta_A \int_{A+B} \omega_a \cdot dS + \int_B (\theta_B - \theta_A) \omega_a \cdot dS$$  

(4.132)

$$= \int_B (\theta_B - \theta_A) \omega_a \cdot dS.$$

The first term on the second line vanishes after using the divergence theorem. Thus, the value of $I$, and so its rate of change, is a function only of an integral over the surface $B$, and the PV flux there must be calculated using the full equations of motion.
However, we do not need to be concerned with a flux of PV concentration through the isentropic surface. Put another way, the PV substance in a volume can change only when isentropes enclosing the volume intersect a boundary such as the earth’s surface.

4.8.1 Interpretation and application

Motion of the isentropic surface

How can the above results hold in the presence of heating? The isentropic surfaces must move in such a way that the total amount of PV concentration contained between them nevertheless stays fixed, and we now demonstrate this explicitly. The potential vorticity equation may be written

$$\frac{\partial Q}{\partial t} + \mathbf{v} \cdot \nabla Q = S_Q,$$ (4.133)

where, from (4.98), $S_Q = (\omega_a/\rho) \cdot \nabla \theta + \nabla \theta \cdot (\nabla \times F)/\rho$. Using mass continuity this may be written as

$$\frac{\partial s}{\partial t} + \nabla \cdot J = 0,$$ (4.134)

where $J = \rho \mathbf{v}Q + N$ and $\nabla \cdot N = -\rho S_Q$. Written this way, the quantity $J/s$ is a notional velocity, $\mathbf{v}_Q$ say, and $s$ satisfies

$$\frac{\partial s}{\partial t} + \nabla \cdot (\mathbf{v}_Q s) = 0.$$ (4.135)

That is, $s$ evolves as if it were being fluxed by the velocity $\mathbf{v}_Q$. The concentration of a chemical tracer $\chi$ (i.e., $\chi$ is the amount of tracer per unit volume) obeys a similar equation, to wit

$$\frac{\partial \chi}{\partial t} + \nabla \cdot (\mathbf{v}_Q \chi) = 0.$$ (4.136)

However, whereas (4.136) implies that $D(\chi/\rho)/Dt = 0$, (4.135) does not imply that $\partial Q/\partial t + \mathbf{v}_Q \cdot \nabla Q = 0$ because $\partial \rho/\partial t + \nabla \cdot (\rho \mathbf{v}_Q) \neq 0$. 

---

Fig. 4.11 (a) Two isentropic surfaces that do not intersect the ground. The integral of PV concentration over the volume between them, $V$, is zero, even if there is heating and the contours move. (b) An isentropic surface, $A$, intersects the ground, $B$, so enclosing a volume $V$. The rate of change of PV concentration over the volume is given by an integral over $B$. 

However, whereas (4.136) implies that $D(\chi/\rho)/Dt = 0$, (4.135) does not imply that $\partial Q/\partial t + \mathbf{v}_Q \cdot \nabla Q = 0$ because $\partial \rho/\partial t + \nabla \cdot (\rho \mathbf{v}_Q) \neq 0$. 

---

**Chapter 4. Vorticity and Potential Vorticity**
Now, the impermeability result tells us that there can be no notional velocity across an isentropic surface. How can this be satisfied by the equations of motion? We write the right-hand side of (4.133) as
\[
\rho S_Q = \nabla \cdot (\dot{\omega}_a + \partial \nabla \times F) = \nabla \cdot (\dot{\omega}_a + F \times \nabla \theta).
\]
(4.137)
Thus, \(N = -\dot{\omega}_a - F \times \nabla \theta\) and we may write the \(J\) vector as
\[
J = \rho v_Q - \dot{\omega}_a - F \times \nabla \theta = \rho Q (v_\perp + v_\parallel) - \dot{\omega}_a - F \times \nabla \theta,
\]
(4.138)
where, making use of the thermodynamic equation,
\[
v_\parallel = v - \frac{v \cdot \nabla \theta}{|\nabla \theta|^2} \nabla \theta, \quad v_\perp = -\frac{\partial \theta}{\partial t} \frac{1}{|\nabla \theta|^2} \nabla \theta.
\]
(4.139a)
\[
\omega_\parallel = \omega_a - \frac{\omega_a \cdot \nabla \theta}{|\nabla \theta|^2} \nabla \theta = \omega_a - \frac{Q}{|\nabla \theta|^2} \nabla \theta.
\]
(4.139b)
The subscripts ‘\(\perp\)’ and ‘\(\parallel\)’ denote components perpendicular and parallel to the local isentropic surface, and \(v_\perp\) is the velocity of the isentropic surface normal to itself.

Eq. (4.138) may be verified by using (4.139) and \(D\theta/Dt = \dot{\theta}\).

The ‘parallel’ terms in (4.138) are all vectors parallel to the local isentropic surface, and therefore do not lead to any flux of PV concentration across that surface. Furthermore, the term \(\rho Q v_\perp\) is \(\rho Q\) multiplied by the normal velocity of the surface. That is to say, the notional velocity associated with the flux normal to the isentropic surface is equal to the normal velocity of the isentropic surface itself, and so it too provides no flux of PV concentration across that surface (even through there may well be a mass flux across the surface). Put simply, the isentropic surface always moves in such a way as to ensure that there is no flux of PV concentration across it. In our proof of the impermeability result in the previous section we used the fact that the potential vorticity multiplied by density is the divergence of something. In the demonstration above we used the fact that the terms \(\text{forcing potential vorticity}\) are the divergence of something.

* Dynamical choices of PV flux and a connection to Bernoulli’s theorem

If we add a non-divergent vector to the flux, \(J\), then it has no effect on the evolution of \(s\). This gauge invariance means that the notional velocity, \(v_Q = J/(\rho Q)\) is similarly non-unique, although it does not mean that there are not dynamical choices for it that are more appropriate in given circumstances. To explore this, let us obtain a general expression for \(J\) by starting with the definition of \(s\), so that
\[
\frac{\partial s}{\partial t} = \nabla \theta \cdot \frac{\partial \omega_a}{\partial t} + \omega_a \cdot \nabla \frac{\partial \theta}{\partial t}
\]
\[
= \nabla \theta \cdot \nabla \times \frac{\partial \mathbf{v}}{\partial t} + \nabla \cdot \left( \omega_a \frac{\partial \theta}{\partial t} \right) = -\nabla \cdot J',
\]
(4.140)
where
\[
J' = \nabla \theta \times \frac{\partial \mathbf{v}}{\partial t} - \frac{\partial \theta}{\partial t} \omega_a + \nabla \phi \times \nabla \chi.
\]
(4.141)
The last term in this expression is an arbitrary divergence-free vector. If we choose \(\phi = \theta\) and \(\chi = B\), where \(B\) is the Bernoulli function given by \(B = I + \mathbf{v}^2/2 + p/\rho\) where \(I\) is the internal energy per unit mass, then
\[
J' = \nabla \theta \times \left( \nabla B + \frac{\partial \mathbf{v}}{\partial t} \right) - \omega_a (\partial - \mathbf{v} \cdot \nabla \theta),
\]
(4.142)
having used the thermodynamic equation $D\theta/Dt = \dot{\theta}$. Now, the momentum equation may be written, without approximation, in the form (see problems 2.1 and 2.2)

$$\frac{\partial \mathbf{v}}{\partial t} = -\omega_a \times \mathbf{v} + T\nabla \eta + \mathbf{F} - \nabla B,$$  \hspace{1cm} (4.143)

where $\eta$ is the specific entropy ($d\eta = c_p d \ln \theta$). Using (4.142) and (4.143) gives

$$J' = \rho Q \mathbf{v} - \dot{\theta} \omega_a + \nabla \eta \times \mathbf{F}.$$  \hspace{1cm} (4.144)

which is the same as (4.138). Furthermore, using (4.141) for steady flow,

$$J = \nabla \theta \times \nabla B.$$  \hspace{1cm} (4.145)

That is, the flux of potential vorticity (in this gauge) is aligned with the intersection of $\theta$- and $B$-surfaces. For steady inviscid and adiabatic flow the Bernoulli function is constant along streamlines; that is, surfaces of constant Bernoulli function are aligned with streamlines, and, because $\theta$ is materially conserved, streamlines are formed at intersecting $\theta$- and $B$-surfaces, as in (1.193). In the presence of forcing, this property is replaced by (4.145), that the flux of PV concentration is along such intersections.

This choice of gauge leading to (4.144) is physical in that it reduces to the true advective flux $\mathbf{v}\rho Q$ for unforced, adiabatic flow, but it is not a unique choice, nor mandated by the dynamics. Choosing $\chi = 0$ leads to the flux

$$J_1 = \rho Q \mathbf{v} - \dot{\theta} \omega_a + \nabla \theta \times (\mathbf{F} - \nabla B),$$  \hspace{1cm} (4.146)

and using (4.141) this vanishes for steady flow, a potentially useful property.

### 4.8.2 Summary Remarks

The impermeability result has a number of consequences, some obvious with hindsight, and it also provides an interesting point of view and diagnostic tool. Here, we will just remark:

* There can be no net transport of potential vorticity across an isentropic surface, and the total amount of potential vorticity in a volume wholly enclosed by isentropic surfaces is zero.

* Thus, and trivially, the amount of potential vorticity contained between two isentropes isolated from the earth’s surface in the northern hemisphere is the negative of the corresponding amount in the southern hemisphere.

* Potential vorticity flux lines (i.e., lines everywhere parallel to $J$) can either close in on themselves or begin and end at boundaries (e.g., the ground, the ocean surface). However, $J$ may change its character. Thus, for example, at the base of the oceanic mixed layer $J$ may change from from being a diabatic flux above to an adiabatic advective flux below. There may be a similar change in character at the atmospheric tropopause.

* The flux vector $J$ is defined only to within the curl of a vector. Thus the vector $J' = J + \nabla \times A$, where $A$ is an arbitrary vector, is as valid as is $J$ in the above derivations and diagnostics.
Notes

1. The frozen-in property — that vortex lines are material lines — was derived by Helmholtz (1858) and is sometimes called Helmholtz’s theorem.

2. The theorem originates with Thomson (1869), who later became later Lord Kelvin.

3. Silberstein (1896) proved that ‘the necessary and sufficient condition for the generation of vortical flow...influenced only by conservative forces...is that the surface of constant pressure and surface of constant density...intersect’, as we derived in section 4.2 and this leads to (4.45). Bjerknes (1898a,b) explicitly put this into the form of a circulation theorem and applied it to problems of meteorological and oceanographic importance (see Thorpe et al. 2003), and the theorem is sometimes called the Bjerknes theorem or the Bjerknes-Silberstein theorem. It is occasionally stated as the evolution of circulation around a circuit is determined by the number of solenoids passing through any surface bounded by that circuit, but the meaning is that of (4.45).

Vilhelm Bjerknes (1862–1951) was a physicist and hydrodynamicist who in 1917 moved to the University of Bergen as founding head of the Bergen Geophysical Institute. Here he did what is probably his most influential work in meteorology, setting up and contributing to the ‘Bergen School of Meteorology’. Among other things he and his colleagues were the first to consider, as a practical proposition, the use of numerical methods — initial data in conjunction with the fluid equations of motion — to forecast the state of the atmosphere, based on earlier work describing how that task might be done. Bjerknes (1904). Innaccurate initial velocity fields compounded with the shear complexity of the effort ultimately defeated them, but the effort was continued (also unsuccessfully) by L. F. Richardson (Richardson 1922), before J. Charney, R. Fjortoft and J. Von Neumann eventually made what may be regarded as the first successful numerical forecast (Charney et al. 1950). Their success can be attributed to the used of a simplified, filtered, set of equations and the use of an electronic computer.

Vilhelm’s son, Jacob Bjerknes (1897–1975) was a leading player in the Bergen school. He was responsible for the now-famous frontal model of cyclones, and was one of the first to seriously discuss the role of cyclones in the general circulation of the atmosphere. In collaboration with Halvor Solberg and Tor Berg- eron the frontal model lead to a prescient picture of the life-cycle of extra-tropical cyclones (see chapter 9), in which a wave grows initially on the polar front (akin to baroclinic instability with the meridional temperature gradient compressed to a front, but baroclinic instability theory was not yet developed), develops into a mature cyclone, occludes and decays. In 1939 Bjerknes moved to the U.S. and, largely because of WWII, stayed, joining UCLA and heading its Dept. of Meteorology after its formation in 1945. He developed an interest in air-sea interactions, and notably proposed the essential mechanism governing El Niño, a feedback between sea-surface temperatures and the strength of the trade winds (Bjerknes 1969). [See also Fried- man (1989), Cressman (1996), articles in Shapiro and Grønas (1999), and a memoir by Arnt Eliassen available from http://www.nap.edu/readingroom/books/biomems/jbjerkins.html.]

4. The result (4.52) is sometimes attributed to Poincaré (1893), although it was also known to Bjerknes (1902).

5. The first derivation of the PV conservation law was given for the shallow water equations by Rossby (1936), with a generalization to multiple layers in Rossby (1938). In the 1936 paper Rossby notes [his eq. (75)] that a fluid column satisfies \( f + \zeta = cD \).
where \( c \) is a constant and \( D \) is the thickness of a fluid column; equivalently, \( (f + \zeta) / D \) is a material invariant. In [Rossby (1940)] this was generalized slightly to an isentropic layer, in which \( \zeta \) is computed using horizontal derivatives taken at constant density or potential temperature. In this paper Rossby also introduces the expression 'potential vorticity', as follows: 'This quantity, which may be called the potential vorticity, represents the vorticity the air column would have it were brought, isopycnally or isentropically, to a standard latitude \( (f_0) \) and stretched or shrunk vertically to a standard depth \( D_0 \) or weight \( \Delta_0 \).' (Rossby's italics.) That is,

\[
\text{Potential Vorticity} = \zeta_0 = \left( \frac{\zeta + f}{D} \right) D_0 - f_0, \tag{4.147}
\]

which follows from his eq. (11), and this is the sense he uses it in that paper. However, potential vorticity has come to mean the quantity \( (\zeta + f)/D \), which of course does not have the dimensions of vorticity. We use it in this latter, now conventional, sense throughout this book. Ironically, quasi-geostrophic potential vorticity as usually defined does have the dimensions of vorticity.

The expression for potential vorticity in a continuously stratified fluid was given by Ertel (1942a), and its relationship to circulation was given by Ertel (1942b). It is now commonly known as the Ertel potential vorticity. Interestingly, in Rossby (1940) we find the Fermat-like comment 'It is possible to derive corresponding results for an atmosphere in which the potential temperature varies continuously with elevation. . . . The generalized treatment will be presented in another place.' (!) Opinions differ as to whether Rossby's and Ertel's derivations were independent: Charney (in Lindzen et al. 1990) suggests they were, and Cressman (1996) remarks that the origin of the concept of potential vorticity is a 'delicate one that has aroused some passion in private correspondences'. In fact, Ertel visited MIT in autumn 1937 and presumably talked to Rossby and became aware of his work. The likeliest scenario is that Ertel knew of Rossby's shallow water theorems, and that he subsequently provided an independent and significant generalization. Rossby and Ertel apparently remained on good terms, but further collaboration was stymied by WWII. They later published a pair of short joint papers, one in German and the other in English, describing their conservation theorems [Ertel and Rossby 1949a,b]. (I thank A. Persson and R. Samelson for some historical details on this.)

---

Further Reading


Written in Truesdell’s inimitable style, this book discusses many aspects of vorticity and has many historical references (as well as a generous definition of what constitutes a ‘kinematic’ result).


Contains an extensive discussion of vorticity and vortices.
Contains derivations of the main vorticity theorems as well as more general material on the dynamics of atmospheric flows.

Chapter 4 contains a brief discussion of potential vorticity, and chapter 7 a longer discussion of Hamiltonian fluid dynamics, in which the particle relabeling symmetry that gives rise to potential vorticity conservation is discussed.

**Problems**

4.1 For the $v_r$ vortex, choose a contour of arbitrary shape (e.g., a square) with segments neither parallel nor perpendicular to the radius, and not enclosing the origin. Show explicitly that the circulation around the contour is zero. (This problem is a little perverse.)

4.2 ♦ Vortex stretching and viscosity.
Suppose there is an incompressible swirling flow given in cylindrical coordinates ($r, \phi, z$):

\[
\mathbf{v} = (v_r, v_\phi, v_z) = \left(-\frac{1}{2} \alpha r, v_\phi, \alpha z\right)
\]  

(P4.1)

Show that this satisfies the mass conservation equation. Show too that vorticity is only non-zero in the vertical direction. Show that the vertical component of the vorticity equation contains only the stretching term and that in steady state it is

\[
-\frac{1}{2} \alpha r \frac{\partial \zeta}{\partial r} = \zeta \alpha + \nu \frac{1}{r} \frac{\partial}{\partial r} \left( r \frac{\partial \zeta}{\partial r} \right).
\]  

(P4.2)

Show that this may be integrated to $\zeta = \zeta_0 \exp\left(-\frac{\alpha r^2}{4\nu}\right)$. Thus deduce that there is a rotational core of thickness $r_o = 2 \left(\frac{\nu}{\alpha}\right)^{1/2}$, and that the radial velocity field is given by

\[
v_\phi(r) = -\frac{1}{r} \frac{2\nu}{\alpha} \zeta_0 \exp\left[\frac{-\frac{\alpha r^2}{4\nu}}\right] + \frac{A}{r}
\]  

(P4.3)

where $A = 2\nu \zeta_0 / \alpha$. What is the swirling velocity field? (From Batchelor 1967)

4.3 Beginning with the three-dimensional vorticity equation in a rotating frame of reference [e.g., (4.60) or (4.61)], or otherwise, obtain an expression for the evolution of the vertical and radial coordinate of relative vorticity in Cartesian and spherical coordinates, respectively. Discuss the differences (if any) between the resulting equations and (4.68). Show carefully how the $\beta$-term arises, and in particular that it may be interpreted as arising from tilting term in the vorticity equation.

4.4 ♦ Making use of Kelvin’s circulation theorem obtain an expression for the potential vorticity that is conserved following the flow (for an adiabatic and unforced fluid) and that is appropriate for the hydrostatic primitive equations on a spherical planet. Express this in terms of the components of the spherical coordinate system.

4.5 ♦ In pressure coordinates for hydrostatic flow on the f-plane, the horizontal momentum equation takes the form

\[
\frac{Du}{Dt} + f \times u = -\nabla \phi
\]  

(P4.4)

On taking the curl of this, there appears to be no baroclinic term. Show that Kelvin’s circulation theorem is nevertheless not in general satisfied, even for unforced, adiabatic flow. By appropriately choosing a path on which to evaluate the circulation obtain an expression for potential vorticity, in this coordinate system, that is conserved following the flow. Hint: look at the hydrostatic equation.
4.6 Solenoids and sea-breezes.

A land-sea temperature contrast of 20 K forces a sea breeze in the surface "mixed layer" (potential temperature nearly uniform with height), as illustrated schematically. The layer extends through the lowest 10% of the mass of the atmosphere.

(a) In the absence of dissipation and diffusion, at what rate does the circulation change on a material circuit indicated? You may assume the horizontal flow is isobaric, and express your answer in m/s per hour.

(b) Suppose the sea breeze is equilibrated by a nonlinear surface drag of the form

\[
\frac{dV}{dt} = -\frac{V^2}{L_F}
\]

with \(L_F = (3 \text{ m s}^{-1})(3600 \text{ s})\). What is the steady speed of the horizontal wind in the case \(L = 50 \text{ km}\)?

(c) Suppose that the width of the circulation is determined by a horizontal thermal diffusion of the form

\[
\frac{D\theta}{Dt} = \kappa_H \frac{\partial^2 \theta}{\partial x^2}
\]

Provide an estimate of \(\kappa_H\) that is consistent with \(L = 50 \text{ km}\). Comment on whether you think the extent of real sea breezes is really determined this way.
Large-scale flow in the ocean and atmosphere is characterized by an approximate balance in the vertical between the pressure gradient and gravity (hydrostatic balance), and in the horizontal between the pressure gradient and the Coriolis force (geostrophic balance). In this chapter we exploit these balances to simplify the Navier-Stokes equations and thereby obtain various sets of simplified ‘geostrophic equations’. Depending on the precise nature of the assumptions we make, we are led to the quasi-geostrophic system for horizontal scales similar to that on which most synoptic activity takes place and, for very large-scale motion, to the planetary-geostrophic set of equations. By eliminating unwanted or unimportant modes of motion, in particular sound waves and gravity waves, and by building in the important balances between flow fields, these filtered equation sets allow the investigator to better focus on a particular class of phenomenon and to potentially achieve a deeper understanding than might otherwise be possible.\(^1\)

Simplifying the equations in this way relies first on scaling the equations. The idea is that we choose the scales we wish to describe, typically either on some a priori basis or by using observations as a guide. We then attempt to derive a set of equations that is simpler than the original set but that consistently describes motion of the chosen scale. An asymptotic method is one approach to this, for it systematically tells us which terms we can drop and which we should keep. The combined approach — scaling plus asymptotics — has proven enormously useful, but it is useful to always remember two things: (i) that scaling is a choice; (ii) that the approach does not explain the existence of particular scales of motion, it just describes the motion that might occur on such scales. We have already employed
this general approach in deriving the hydrostatic primitive equations, but now we go further.

5.1 GEOSTROPHIC SCALING

\( I \) have no satisfaction in formulas unless I feel their numerical magnitude. \( \)\( \) William Thomson, Lord Kelvin (1824–1907).

5.1.1 Scaling in the Shallow Water Equations
Postponing the complications that come with stratification, we begin with the shallow water equations. With the odd exception, we will denote the scales of variables by capital letters; thus, if \( L \) is a typical length scale of the motion we wish to describe, and \( U \) a typical velocity scale, and assuming the scales are horizontally isotropic, we write

\[
(x, y) \sim L \quad \text{or} \quad (x, y) = \mathcal{O}(L) \\
(u, v) \sim L \quad \text{or} \quad (u, v) = \mathcal{O}(U) \tag{5.1}
\]

and similarly for other variables. We may then nondimensionalize the variables by writing

\[
(x, y) = L(\hat{x}, \hat{y}), \quad (u, v) = U(\hat{u}, \hat{v}) \tag{5.2}
\]

where the hatted variables are nondimensional and, by supposition, are \( \mathcal{O}(1) \). The various terms in the momentum equation then scale as:

\[
\frac{\partial u}{\partial t} + u \cdot \nabla u + f \times u = -g \nabla \eta \\
\frac{U}{T} \quad \frac{U^2}{L} \quad fU \quad \frac{\mathcal{H}}{L} \tag{5.3b}
\]

where the \( \nabla \) operator acts in the \( x, y \) plane and \( \mathcal{H} \) is the amplitude of the variations in the surface displacement. (We use \( \eta \) to denote the height of the free surface above some arbitrary reference level, as in Fig. [3.1]. Thus, \( \eta = H + \Delta \eta \), where \( \Delta \eta \) denotes the variation of \( \eta \) about its mean position.)

The ratio of the advective term to the rotational term in the momentum equation (5.3) is \( (U^2/L)/(fU) = U/fL \); this is the Rossby number, first encountered in chapter 2.\(^2\) Using values typical of the large-scale circulation (e.g., from table 2.1) we find that \( Ro \approx 0.1 \) for the atmosphere and \( Ro \approx 0.01 \) for the ocean, small in both cases. If we are interested in motion that has the advective timescale \( T = L/U \) then we scale time by \( L/U \) so that

\[
t = \frac{L}{U} \hat{t}, \tag{5.4}
\]

and the local time derivative and the advective term then both scale as \( U^2/L \), and both are order Rossby number smaller than the rotation term. Then, either the Coriolis term is the dominant term in the equation, in which case we have a state of no motion with \( -fv = 0 \), or else the Coriolis force is balanced by the pressure force, and the dominant balance is

\[
-fv = -g \frac{\partial \eta}{\partial x}, \tag{5.5}
\]
5.1 Geostrophic Scaling

namely geostrophic balance, as encountered in chapter 2. If we make this nontrivial choice, then the equation informs us that variations in \( \eta \) scale according to

\[
\Delta \eta \sim \mathcal{H} = \frac{fUL}{g}
\]  

(5.6)

We can also write \( \mathcal{H} \) as

\[
\mathcal{H} = Ro \frac{f^2L^2}{g} = RoH \frac{L^2}{L^2_d},
\]  

(5.7)

where \( L_d = \sqrt{gH/f} \) is the deformation radius, and \( H \) is the mean depth of the fluid. The variations in fluid height thus scale as

\[
\frac{\Delta \eta}{H} \sim Ro \frac{L^2}{L^2_d},
\]  

(5.8)

and the height of the fluid may be written as

\[
\eta = H \left( 1 + Ro \frac{L^2}{L^2_d} \hat{\eta} \right) \quad \text{and} \quad \Delta \eta = Ro \frac{L^2}{L^2_d} H \hat{\eta},
\]  

(5.9)

where \( \hat{\eta} \) is the \( O(1) \) nondimensional value of the surface height deviation.

**Nondimensional momentum equation**

If we use (5.9) to scale height variations, (5.2) to scale lengths and velocities, and (5.4) to scale time, then the momentum equation (5.3) becomes

\[
Ro \left\{ \frac{\partial \hat{\mathbf{u}}}{\partial \hat{t}} + (\hat{\mathbf{u}} \cdot \nabla) \hat{\mathbf{u}} \right\} + \hat{f} \times \hat{\mathbf{u}} = -\nabla \hat{\eta},
\]  

(5.10)

where \( \hat{f} = k \hat{f} = k f / f_0 \), where \( f_0 \) is a representative value of the Coriolis parameter. (If \( f \) is a constant, then \( \hat{f} = 1 \), but it is useful to keep it in the equations to indicate the presence of Coriolis parameter. Also, where the operator \( \nabla \) operates on a nondimensional variable then the differentials are taken with respect to the nondimensional variables \( \hat{x}, \hat{y} \).) All the variables in (5.10) will be supposed to be of order unity, and the Rossby number multiplying the local time derivative and the advective terms indicates the smallness of those terms. By construction, the dominant balance in this equation is the geostrophic balance between the last two terms.

**Nondimensional mass continuity (height) equation**

The (dimensional) mass continuity equation can be written

\[
\frac{1}{H} \frac{D\eta}{Dt} + \left( 1 + \frac{\Delta \eta}{H} \right) \nabla \cdot \mathbf{u} = 0
\]  

(5.11)

Using (5.2), (5.4) and (5.9) this equation may be written

\[
Ro \left( \frac{L}{L_d} \right)^2 \frac{D\hat{\eta}}{Dt} + \left[ 1 + Ro \left( \frac{L}{L_d} \right)^2 \hat{\eta} \right] \nabla \cdot \hat{\mathbf{u}} = 0.
\]  

(5.12)
Equations (5.10) and (5.12) are the nondimensional versions of the full shallow water equations of motion. Evidently, some terms in the equations of motion are small and may be eliminated with little loss of accuracy, and the way this is done will depend on the size of the second nondimensional parameter, \((L/L_d)^2\). We explore this in sections 5.2 and 5.3.

**Froude and Burger numbers**

The Froude number may be generally defined as the ratio of a fluid particle speed to a wave speed. In a shallow-water system this gives

\[
Fr \equiv \frac{U}{\sqrt{gH}} = \frac{U}{f_0 L_d} = \frac{Ro}{L_d}.
\]  

(5.13)

The Burger number\(^3\) is a useful measure of scale of motion of the fluid, relative to the deformation radius, and may be defined by

\[
Bu \equiv \left(\frac{L_d}{L}\right)^2 = \frac{gH}{f_0^2 L^2} = \left(\frac{Ro}{Fr}\right)^2.
\]  

(5.14)

It is also useful to define the parameter \(F \equiv Bu^{-1}\), which is like the square of a Froude number but uses the rotational speed \(fL\) instead of \(U\) in the numerator.

### 5.1.2 Geostrophic scaling in the stratified equations

We now apply the same scaling ideas, *mutatis mutandis*, to the stratified primitive equations. We use the hydrostatic anelastic equations, which we write as:

\[
\frac{D}{Dt} u + f \times u = -\nabla_z \phi,
\]

(5.15a)

\[
\frac{\partial \phi}{\partial z} = b,
\]

(5.15b)

\[
\frac{Db}{Dt} = 0,
\]

(5.15c)

\[
\nabla \cdot (\tilde{\rho} \nu) = 0.
\]

(5.15d)

where \(b\) is the buoyancy and \(\tilde{\rho}\) is a reference density profile. Anticipating that the average stratification may not scale in the same way as the deviation from it, let us separate out the contribution of the advection of a reference stratification in (5.15c) by writing

\[
b = \tilde{b}(z) + b'(x,y,z,t).
\]

(5.16)

Then the thermodynamic equation becomes

\[
\frac{Db'}{Dt} + N^2 w = 0,
\]

(5.17)

where \(N^2 \equiv \partial \tilde{b}/\partial z\) (and the advective derivative is still three-dimensional). We then let \(\phi = \tilde{\phi}(z) + \phi'\) where \(\tilde{\phi}\) is hydrostatically balanced by \(\tilde{b}\), and the hydrostatic equation becomes

\[
\frac{\partial \phi'}{\partial z} = b'.
\]

(5.18)

Equations (5.17) and (5.18) replace (5.15c) and (5.15b), and \(\phi'\) is used in (5.15a).
Non-dimensional equations

We scale the basic variables by supposing that

\[(x, y) \sim L, \quad (u, v) \sim U, \quad t \sim \frac{L}{U}, \quad z \sim H, \quad f \sim f_0,\]  

(5.19)

where the scaling variables (capitalized, except for \(f_0\)) are chosen such that the nondimensional values have values of order unity. We presume that the scales chosen are such that the Rossby number is small; that is \(Ro = U/(f_0 L) \ll 1\). In the momentum equation the pressure term then balances the Coriolis force,

\[|f \times u| \sim |\nabla \phi'|\]

(5.20)

and so the pressure scales as

\[\phi' \sim \Phi = f_0 UL.\]

(5.21)

Using the hydrostatic relation, (5.21) implies that the buoyancy scales as

\[b' \sim B = f_0 UL.\]

(5.22)

and from this we obtain

\[\left(\frac{\partial b'/\partial z}{N^2}\right) \sim \frac{Ro L^2}{L_d^2},\]

(5.23)

where \(L_d = NH/f_0\) is the deformation radius in the continuously stratified fluid, analogous to the quantity \(\sqrt{gH}/f_0\) in the shallow water system, and we use the same symbol, \(L_d\), for both. In the continuously stratified system, if the scale of motion is the same as or smaller than the deformation radius, and the Rossby number is small, then the variations in stratification are small. The choice of scale is the key difference between the planetary geostrophic and quasi-geostrophic equations.

Finally, we will nondimensionalize the vertical velocity by using the mass conservation equation,

\[\frac{1}{\tilde{\rho}} \frac{\partial \tilde{\rho} w}{\partial z} = -\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right),\]

(5.24)

and we suppose that this implies

\[w \sim W = \frac{UH}{L}.\]

(5.25)

This is a naïve scaling for rotating flow: if the Coriolis parameter is nearly constant the geostrophic velocity is nearly horizontally non-divergent and the right-hand side of (5.24) is small, and \(W \ll UH/L\). We might then estimate \(w\) by cross-differentiating geostrophic balance to obtain the linear geostrophic vorticity equation and corresponding scaling:

\[\beta v = f \frac{\partial w}{\partial z}, \quad w \sim W = \frac{\beta U H}{f_0}.\]

(5.26a,b)

However, rather than using (5.26b) from the outset, we will use (5.25) and let the asymptotics guide us to a proper scaling in the fullness of time. Note that if variations in the Coriolis parameter are large and \(\beta \sim f_0/L\), then (5.26b) is the same as (5.25).
Given the scalings above [using (5.25) for $w$] we nondimensionalize by setting

\begin{align}
(\hat{x}, \hat{y}) &= L^{-1}(x, y), \quad \hat{z} = H^{-1}z, \quad (\hat{u}, \hat{v}) = U^{-1}(u, v), \quad \hat{t} = \frac{U}{L} t, \\
\hat{w} &= \frac{L}{UH} w, \quad \hat{f} = f_0^{-1} f, \quad \hat{\phi} = \frac{\phi'}{f_0 UL}, \quad \hat{b} = \frac{H}{f_0 UL} b'.
\end{align}

(5.27)

where the hatted variables are nondimensional. The horizontal momentum and hydrostatic equations then become

\begin{align}
Ro \frac{D \hat{u}}{Dt} + \hat{f} \times \hat{u} &= -\nabla \hat{\phi}, \\
\frac{\partial \hat{\phi}}{\partial \hat{z}} &= \hat{b}.
\end{align}

(5.28) \hspace{1cm} (5.29)

The non-dimensional mass conservation equation is simply

\begin{align}
\frac{1}{\hat{\rho}} \nabla \cdot (\hat{\rho} \hat{\mathbf{v}}) &= \left( \frac{\partial \hat{u}}{\partial \hat{x}} + \frac{\partial \hat{v}}{\partial \hat{y}} + \frac{1}{\hat{\rho}} \frac{\partial \hat{\rho} \hat{w}}{\partial \hat{z}} \right) = 0.
\end{align}

(5.30)

and the nondimensional thermodynamic equation is

\begin{align}
\frac{f_0 UL}{H} \frac{D \hat{b}}{Dt} + N^2 \frac{HU}{L} \hat{w} &= 0,
\end{align}

(5.31)

or

\begin{align}
Ro \frac{D \hat{b}}{Dt} + \left( \frac{L_d}{L} \right)^2 \hat{w} &= 0.
\end{align}

(5.32)

The nondimensional primitive equations are summarized in the box on the facing page.

5.2 THE PLANETARY GEOSTROPHIC EQUATIONS

We now use the low Rossby number scalings above to derive equation sets that are simpler than the original, ‘primitive’, ones. The planetary geostrophic equations are probably the simplest such set of equations, and we derive these equations first for the shallow water equations, and then for the stratified primitive equations.

5.2.1 Using the shallow water equations

Informal derivation

The advection and time derivative terms in the momentum equation (5.10) are order Rossby number smaller than the Coriolis and pressure terms (the term in square brackets is multiplied by $Ro$), and therefore let us neglect them. The momentum equation straightforwardly becomes

\begin{align}
\hat{f} \times \hat{u} &= -\nabla \hat{\eta}.
\end{align}

(5.33)
5.2 The Planetary Geostrophic Equations

Nondimensional Primitive Equations

Horizontal momentum:
\[
\frac{\text{Ro}}{\text{Dt}} \frac{\partial \hat{u}}{\partial t} + \hat{f} \times \hat{u} = -\nabla \hat{\phi} \quad (\text{NDPE.1})
\]

Hydrostatic:
\[
\frac{\partial \hat{\phi}}{\partial \hat{z}} = \hat{b} \quad (\text{NDPE.2})
\]

Mass continuity:
\[
\left(\frac{\partial \hat{u}}{\partial \hat{x}} + \frac{\partial \hat{v}}{\partial \hat{y}} + \frac{1}{\hat{\rho}} \frac{\partial \hat{\rho} \hat{w}}{\partial \hat{z}}\right) = 0 \quad (\text{NDPE.3})
\]

Thermodynamic:
\[
\frac{\text{Ro}}{\text{Dt}} \frac{\partial \hat{b}}{\partial t} + \left(\frac{L_d}{L}\right)^2 \hat{w} = 0 \quad (\text{NDPE.4})
\]

These equations are written for the anelastic equations. The Boussinesq equations result if we take \( \hat{\rho} = 1 \). The equations in pressure coordinates have a very similar form to the Boussinesq equations, but with a slight difference in hydrostatic equation.

The mass conservation equation \( (5.12) \), contains two nondimensional parameters, \( \text{Ro} = U/(f_0 L) \) (the Rossby number), and \( F = L/L_d \) (the ration of the length scale of the motion to the deformation scale) and we must make a choice as to the relationship of these two numbers. We will choose
\[
F \text{Ro} = \mathcal{O}(1), \quad (5.34)
\]
which implies
\[
L^2 \gg L_d^2 \quad \text{or equivalently} \quad F \gg 1, \quad \text{Bu} \ll 1. \quad (5.35)
\]
That is to say, we suppose that the scales of motion are much larger than the deformation scale. Given this choice, all the terms in the mass conservation equation, \( (5.12) \), are of roughly the same size, and we retain them all. Thus, the shallow water planetary geostrophic equations are the full mass continuity equation along with geostrophic balance and a geometric relationship between the height field and fluid thickness, and in dimensional form these are:
\[
\frac{\text{D}h}{\text{Dt}} + h \nabla \cdot \mathbf{u} = 0 \quad (5.36a, b)
\]
\[
f \times \mathbf{u} = -g \nabla \eta, \quad \eta = h + \eta_b
\]
We emphasize that the planetary geostrophic equations are only valid for scales of motion much larger than the deformation radius. The height variations are then as large as the mean height field itself; that is, using \( (5.8) \), \( \Delta \eta/H = \mathcal{O}(1) \).

**Formal derivation**

We assume that:
(i) The Rossby number is small. \( Ro = \frac{U}{f_0 L} \ll 1 \).

(ii) The scale of the motion is significantly larger than the deformation scale. That is, (5.34) holds or

\[
F = Bu^{-1} = \left( \frac{L}{L_d} \right)^2 \gg 1
\]

and in particular

\[
FRo = \mathcal{O}(1).
\]

(iii) Time scales advectively, so that \( T = \frac{L}{U} \).

The idea is now to expand the nondimensional variables velocity and height fields in an asymptotic series with Rossby number as the small parameter, substitute into the equations of motion, and derive a simpler set of equations. It is a nearly trivial exercise in this instance, and so it illustrates well the methodology. The expansions are

\[
\hat{u} = \hat{u}_0 + Ro\hat{u}_1 + Ro^2\hat{u}_2 + \cdots
\]

and

\[
\hat{\eta} = \hat{\eta}_0 + Ro\hat{\eta}_1 + Ro^2\hat{\eta}_2 + \cdots
\]

Then substituting (5.39a) and (5.39b) into the momentum equation gives

\[
Ro \left[ \frac{\partial \hat{u}_0}{\partial t} + \hat{u}_0 \cdot \nabla \hat{u}_0 + \hat{f} \times \hat{u}_1 \right] + \hat{f} \times \hat{u}_0 = -\nabla \hat{\eta}_0 - Ro \left[ \nabla \hat{\eta}_1 \right] + \mathcal{O}(Ro^2)
\]

The Rossby number is now an asymptotic ordering parameter; thus, the sum of all the terms at any particular order in Rossby number must vanish. At lowest order we obtain the simple expression

\[
\hat{f} \times \hat{u}_0 = -\nabla \hat{\eta}_0.
\]

Note that although \( f_0 \) is a representative value of \( f \), we have made no assumptions about the constancy of \( f \). In particular, \( f \) is allowed to vary by an order one amount, provided that it does not become so small that the Rossby number \((U/f_0 L)\) is not small.

The appropriate height (mass conservation) equation is similarly obtained by substituting (5.39a) and (5.39b) into the shallow water mass conservation equation. Because \( FRo = \mathcal{O}(1) \) at lowest order we simply retain all the terms in the equation to give

\[
FRo \left[ \frac{\partial \hat{\eta}_0}{\partial t} + \hat{u}_0 \cdot \nabla \hat{\eta}_0 \right] + [1 + FRo\hat{\eta}] \nabla \cdot \hat{u}_0 = 0.
\]

Equations (5.41) and (5.42) are a closed set, and constitute the nondimensional planetary geostrophic equations. The dimensional forms of these equations are just (5.36).

**Variation of the Coriolis parameter**

Suppose then that \( f \) is a constant \((f_0)\), or nearly so. Then, from the curl of (5.41), \( \nabla \cdot \hat{u}_0 = 0 \). This means that we can define a streamfunction for the flow and,
from geostrophic balance, the height field is just that streamfunction. That is, in
dimensional form,
\[ \psi = \frac{g}{f_0} \eta, \quad u = -k \times \nabla \psi, \tag{5.43} \]
and (5.42) becomes, in dimensional form,
\[ \frac{\partial \eta}{\partial t} + u \cdot \nabla \eta = 0, \quad \text{or} \quad \frac{\partial \eta}{\partial t} + J(\psi, \eta) = 0. \tag{5.44} \]
where \( J(a, b) = a_x b_y - a_y b_x \). But since \( \eta \propto \psi \) the advective term is proportional to
\( J(\psi, \psi) \), which is zero. Thus, the flow does not evolve at this order. The planetary
geostrophic equations are uninteresting if the scale of the motion is such that the
Coriolis parameter is not variable. On earth, the scale of motion on which this
parameter regime exists is rather limited, since the planetary geostrophic equations
require that the scale of motion is also larger than the deformation radius. In the
earth’s atmosphere, any scale that is larger than the deformation radius will be such
that the Coriolis parameter varies significantly over it, and we do not encounter this
parameter regime. On the other hand, in the earth’s ocean the deformation radius
is relatively small and there exists a small parameter regime that has scales larger
than the deformation radius but smaller than that on which the Coriolis parameter
varies.  

Potential vorticity
The shallow water PG equations may be written as an evolution equation for an ap-
proximated potential vorticity. A little manipulation reveals that (5.36) are equiva-
 lent to:

\[ \frac{DQ}{Dt} = 0, \quad Q = \frac{f}{h}, \quad f \times u = -g \nabla \eta, \quad \eta = h + \eta_h \tag{5.45} \]

Thus, potential vorticity is a material invariant in the approximate equation set, just as it is in the full equations. The other variables — the free surface height and the velocity — are diagnosed from it, a process known as potential vorticity
inversion. In the planetary geostrophic approximation, the inversion proceeds using
the approximate form \( f/h \) rather than the full potential vorticity, \( (f + \zeta)/h \). (Strictly
speaking, we do not approximate potential vorticity, because this is the evolving
variable. Rather, we approximate the inversion relations from which we derive the
height and velocity fields.) The simplest way of all to derive the shallow water PG
equations is to begin with the conservation of potential vorticity, and to note that at
small Rossby number the expression \( (\zeta + f)/h \) may be approximated by \( f/h \). Then,
noting in addition that the flow is geostrophic, (5.45) immediately emerges. Every
approximate set of equations that we derive in this chapter may be expressed as the
evolution of potential vorticity, with the other fields being obtained diagnostically
from it.

5.2.2 The Planetary geostrophic equations for stratified flow
To explore the stratified system we will use the (inviscid and adiabatic) Boussinesq equations of motion with the hydrostatic approximation. The derivation carries through easily enough using the anelastic or pressure-coordinate equations, but as the PG equations have more oceanographic importance than atmospheric using the incompressible equations is quite appropriate.

**Simplifying the equations**

The nondimensional equations we begin with are (5.28)–(5.32). As in the shallow water case we expand these in a series in Rossby number, so that:

\[
\hat{u} = \hat{u}_0 + \epsilon \hat{u}_1 + \epsilon^2 \hat{u}_2 + \cdots, \quad \hat{b} = \hat{b}_0 + \epsilon \hat{b}_1 + \epsilon^2 \hat{b}_2 + \cdots \tag{5.46}
\]

and similarly for \(\hat{v}, \hat{w}\) and \(\hat{\phi}\), where \(\epsilon = Ro\), the Rossby number. Substituting into the nondimensional equations of motion (on page 205) and equating powers of \(\epsilon\) gives the lowest order momentum, hydrostatic, and mass conservation equations:

\[
\hat{f} \times \hat{u}_0 = -\nabla \hat{\phi}_0, \tag{5.47a}
\]

\[
\frac{\partial \hat{\phi}_0}{\partial z} = \hat{b}_0, \tag{5.47b}
\]

\[
\nabla \cdot \hat{v}_0 = 0. \tag{5.47c}
\]

If we also assume that \(L_d/L = O(1)\), then the thermodynamic equation (5.32) becomes

\[
\left(\frac{L_d}{L}\right)^2 \hat{\omega}_0 = 0. \tag{5.48}
\]

Of course we have neglected any diabatic terms in this equation, which would in general provide a non-zero right-hand side. Nevertheless, this is not a useful equation, because the set of the equations we have derived, (5.47), can no longer evolve: all the time derivatives have been scaled away! Thus, although instructive, these equations are not very useful. If instead we assume that the scale of motion is much larger than the deformation scale then the other terms in the thermodynamic equation will become equally important. Thus, we suppose that \(L_d \ll L^2\) or, more formally, that \(L^2 = O(Ro^{-1})L_d^2\), and then all the terms in the thermodynamic equation are retained. A closed set of equations is then given by (5.47) and the thermodynamic equation (5.32).

**Dimensional equations**

Restoring the dimensions, dropping the asymptotic subscripts, and allowing for the possibility of a source term, denoted \(S[b']\), in the thermodynamic equation, the planetary-geostrophic equations of motion are:

\[
\begin{align*}
\frac{\partial b'}{\partial t} + w'N^2 &= S[b'] \\
\hat{f} \times \hat{u} &= -\nabla \hat{\phi}' \\
\frac{\partial \phi'}{\partial z} &= b' \\
\nabla \cdot \hat{v} &= 0
\end{align*} \tag{5.49}
\]
The thermodynamic equation may also be written simply as
\[
\frac{Db}{Dt} = \dot{b}
\]  
(5.50)

where \(b\) now represents the total stratification. The relevant pressure, \(\phi\), is then the pressure that is in hydrostatic balance with \(b\), so that geostrophic and hydrostatic balance are most usefully written as
\[
f \times u = -\nabla \phi, \quad \frac{\partial \phi}{\partial z} = b.
\]  
(5.51a,b)

Potential vorticity
Manipulation of (5.49) reveals that we can equivalently write the equations as an evolution equation for potential vorticity. Thus, the evolution equations may be written
\[
\frac{DQ}{Dt} = \dot{Q}
\]
\[
Q = f \frac{\partial b}{\partial z}
\]
(5.52)

where \(\dot{Q} = f \frac{\partial b}{\partial z}\), and the inversion — i.e., the diagnosis of velocity, pressure and buoyancy — is carried out using the hydrostatic, geostrophic and mass conservation equations.

Applicability to the ocean and atmosphere
In the atmosphere a typical deformation radius \(NH/f\) is about 1,000 km. The constraint that the scale of motion be much larger than the deformation radius is thus quite hard to satisfy, since one quickly runs out of room on a planet whose equator-to-pole distance is 10,000 km. Thus, only the largest planetary waves can satisfy the planetary-geostrophic scaling in the atmosphere and we should then also write the equations in spherical coordinates.

In the ocean the deformation radius is about 100 km, so there is lots of room for the planetary-geostrophic equations to hold, and indeed much of the theory of the large-scale structure of the ocean involves the planetary-geostrophic equations.

5.3 THE SHALLOW WATER QUASI-GEOSTROPHIC EQUATIONS
We now derive a set of geostrophic equations that is valid (unlike the PG equations) when the horizontal scale of motion is similar to that of the deformation radius. These equations are called the quasi-geostrophic equations, and are perhaps the most widely used set of equations for theoretical studies of the atmosphere and ocean. The specific assumptions we make are:

(i) The Rossby number is small, so that the flow is in near-geostrophic balance.

(ii) The scale of the motion is not significantly larger than the deformation scale. Specifically, we shall require that
\[
Ro \left( \frac{L}{L_d} \right)^2 = O(Ro).
\]  
(5.53)
For the shallow water equations, this assumption implies, using (5.9), that the variations in fluid depth are small compared to its total depth. For the continuously stratified system it implies, using (5.23), that the variations in stratification are small compared to the background stratification.

(iii) Variations in the Coriolis parameter are small. That is, \(|\beta L| \ll |f_0|\) where \(L\) is the length-scale of the motion.

(iv) Time scales advectively; that is, the scaling for time is given by \(T = L/U\).

The second and third of these differ from the planetary geostrophic counterparts: we make the second assumption because we wish to explore a different parameter regime, and we then find that the third assumption is necessary to avoid a rather trivial state (i.e., a leading order balance of \(\beta \vec{v} = 0\), see the discussion surrounding (5.77)). All of the assumptions are the same whether we consider the shallow water equations or a continuously stratified flow, and in this section we consider the former.

5.3.1 Single-layer shallow water quasi-geostrophic equations

The algorithm is, again, to expand the variables \(\hat{u}, \hat{v}, \hat{\eta}\) in an asymptotic series with Rossby number as the small parameter, substitute into the equations of motion, and derive a simpler set of equations. Thus we let

\[
\begin{align*}
\hat{u} &= \hat{u}_0 + Ro \hat{u}_1 + Ro^2 \hat{u}_2 + \cdots, \\
\hat{v} &= \hat{v}_0 + Ro \hat{v}_1 + Ro^2 \hat{v}_2 + \cdots \\
\hat{\eta} &= \hat{\eta}_0 + Ro \hat{\eta}_1 + Ro^2 \hat{\eta}_2 + \cdots.
\end{align*}
\]  

We will recognize the smallness of \(\beta\) compared to \(f_0/L\) by letting \(\beta = \beta U/L^2\), where \(\beta\) is assumed to be a parameter of order unity. Then the expression \(f = f_0 + \beta y\) becomes

\[
\hat{f} = f/f_0 = \hat{f}_0 + Ro \hat{\beta} \hat{y}.
\]

where \(\hat{f}_0\) is the nondimensional value of \(f_0\); its value is unity, but it is helpful to denote it explicitly. Substitute (5.54) into the nondimensional momentum equation (5.10), and equate powers of \(Ro\). At lowest order we obtain

\[
\hat{f}_0 \hat{u}_0 = -\frac{\partial \hat{\eta}_0}{\partial \hat{y}}, \quad \hat{f}_0 \hat{v}_0 = \frac{\partial \hat{\eta}_0}{\partial \hat{x}}.
\]

Cross-differentiating gives

\[
\nabla \cdot \hat{\mathbf{u}}_0 = 0,
\]

where, when \(\nabla\) operates on a nondimensional variable, the derivatives are taken with respect to the nondimensional variables \(\hat{x}\) and \(\hat{y}\). From (5.57) we see that the velocity field is divergence-free, and that this arises from the momentum equation rather than the mass conservation equation.

The mass conservation equation is also, at lowest order, \(\nabla \cdot \hat{\mathbf{u}}_0 = 0\), and at next order we have

\[
F \frac{\partial \hat{\eta}_0}{\partial t} + F \hat{\mathbf{u}}_0 \cdot \nabla \hat{\eta}_0 + \nabla \cdot \hat{\mathbf{u}}_1 = 0
\]
This equation is not closed, because the evolution of the zeroth order term involves evaluation of a first order quantity. For closure, we go to next order in the momentum equation,

$$\frac{\partial \hat{u}_0}{\partial t} + (\hat{u}_0 \cdot \nabla) \hat{u}_0 + \hat{\beta} \hat{y} \hat{k} \times \hat{u}_0 - \hat{f}_0 \hat{k} \times \hat{u}_1 = -\nabla \hat{\eta}_1,$$

(5.59)

and take its curl to give the vorticity equation:

$$\frac{\partial \hat{\zeta}_0}{\partial t} + (\hat{u}_0 \cdot \nabla)(\hat{\zeta}_0 + \hat{\beta} \hat{y}) = -\hat{f}_0 \nabla \cdot \hat{u}_1.$$

(5.60)

The term on the right-hand side is the vortex stretching term. Only vortex stretching by the background or planetary vorticity is present, because the vortex stretching by the relative vorticity is a factor Rossby number smaller. Eq. (5.60) is also not closed; however, we may use (5.58) to eliminate the divergence term to give

$$\frac{\partial \hat{\zeta}_0}{\partial t} + (\hat{u}_0 \cdot \nabla)(\hat{\zeta}_0 + \hat{\beta} \hat{y}) = \hat{f}_0 \left( F \frac{\partial \hat{\eta}_0}{\partial t} + F \hat{u} \cdot \nabla \hat{\eta}_0 \right),$$

(5.61)

or

$$\frac{\partial}{\partial t}(\hat{\zeta}_0 - \hat{f}_0 F \hat{\eta}_0) + (\hat{u}_0 \cdot \nabla)(\hat{\zeta}_0 + \hat{\beta} \hat{y} - \hat{f}_0 \hat{F} \hat{\psi}_0) = 0.$$  

(5.62)

The final step is to note that the lowest order vorticity and height fields are related through geostrophic balance, so that using (5.56) we can write

$$\hat{u}_0 = -\frac{\partial \hat{\psi}_0}{\partial \hat{y}}, \quad \hat{v}_0 = \frac{\partial \hat{\psi}_0}{\partial \hat{x}}, \quad \hat{\zeta}_0 = \nabla^2 \hat{\psi}_0,$$

(5.63)

where $\hat{\psi}_0 = \hat{\eta}_0 / \hat{f}_0$ is the streamfunction. Eq. (5.62) can thus be written,

$$\frac{\partial}{\partial t} (\nabla^2 \hat{\psi}_0 - \hat{f}_0 \hat{F} \hat{\psi}_0) + (\hat{u}_0 \cdot \nabla)(\hat{\zeta}_0 + \hat{\beta} \hat{y} - \hat{f}_0 \hat{F} \hat{\psi}_0) = 0,$$

(5.64)

or

$$\frac{D}{Dt}(\nabla^2 \hat{\psi}_0 + \hat{\beta} \hat{y} - \hat{f}_0 \hat{F} \hat{\psi}_0) = 0,$$

(5.65)

where the subscript ‘0’ on the material derivative indicates that the lowest order velocity, the geostrophic velocity, is the advecting velocity. Restoring the dimensions, (5.65) becomes

$$\frac{D}{Dt}(\nabla^2 \psi + \beta y - \frac{1}{L_d^2} \psi) = 0,$$

(5.66)

where $\psi = (g / f_0) \eta$, $L_d^2 = gH / f_0^2$, and the advective derivative is

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u_g \frac{\partial}{\partial x} + v_g \frac{\partial}{\partial y} = \frac{\partial}{\partial t} - \frac{\partial \psi}{\partial y} \frac{\partial}{\partial x} + \frac{\partial \psi}{\partial x} \frac{\partial}{\partial y} = \frac{\partial}{\partial t} + f(\psi, \cdot).$$

(5.67)

Another form of (5.66) is

$$\frac{D}{Dt}(\zeta + \beta y - \frac{f_0}{H} \eta) = 0,$$

(5.68)
with \( \zeta = (g/f_0) \nabla^2 \eta \). Equations (5.66) and (5.68) are forms of the shallow-water quasi-geostrophic potential vorticity equation. The quantity

\[
q \equiv \zeta + \beta y - f_0 \frac{1}{H} \nabla^2 \eta = \nabla^2 \psi + \beta y - \frac{1}{H^2} \psi
\]

is the shallow water quasi-geostrophic potential vorticity.

**Connection to shallow water potential vorticity**

The quantity \( q \) given by (5.69) is an approximation (except for dynamically unimportant constant additive and multiplicative factors) to the shallow water potential vorticity. To see the truth of this statement, begin with the expression for the shallow water potential vorticity,

\[
Q = f + \zeta h.
\]

Now let \( h = H(1 + \eta'/H) \), where \( \eta' \) is the perturbation of the free-surface height, and assume that \( \eta'/H \) is small to obtain

\[
Q = \frac{f + \zeta H(1 + \eta'/H)}{H} \approx \frac{1}{H} \left( f + \zeta \right) \left( 1 - \frac{\eta'}{H} \right) \approx \frac{1}{H} \left( f_0 + \beta y + \zeta - f_0 \frac{\eta'}{H} \right).
\]

Because \( f_0/H \) is a constant it has no effect in the evolution equation, and the quantity given by

\[
q = \beta y + \zeta - f_0 \frac{\eta'}{H}
\]

is materially conserved. Using geostrophic balance we have \( \zeta = \nabla^2 \psi \) and \( \eta' = f_0 \psi/g \) so that (5.72) is identical to (5.69). [Note that only the variation in \( \eta \) are important in (5.68) or (5.69).]

The approximations needed to go from (5.70) to (5.69) are the same as those used in our earlier, more long-winded, derivation of the quasi-geostrophic equations. That is, we assumed that \( f \) itself is nearly constant, and that \( f_0 \) is much larger than \( \zeta \), equivalent to a low Rossby number assumption. It was also necessary to assume that \( H \gg \eta' \) to enable the expansion of the height field which, using assumption (4) on page 209, is equivalent to requiring that scale of motion be not significantly larger than the deformation scale. The derivation is completed by noting that the advection of the potential vorticity should be by the geostrophic velocity alone, and we recover (5.66) or (5.68).

**Two interesting limits**

There are two interesting limits to the quasi-geostrophic potential vorticity equation:

(i) *Motion on scales much smaller than the deformation radius.*

That is, \( L \ll L_d \) and thus \( Bu \gg 1 \) or \( F \ll 1 \). Then (5.65) becomes

\[
\frac{\partial \zeta}{\partial t} + u_\psi \cdot \nabla \zeta = 0 \quad \text{or} \quad \frac{\partial \zeta}{\partial t} + J(\psi, \zeta) = 0,
\]

where \( \zeta = \nabla^2 \psi \) and \( J(\psi, \zeta) = \psi_x \zeta_y - \psi_y \zeta_x \). Thus, the motion obeys the two-dimensional vorticity equation. Physically, on small length scales the deviations in the height field are very small and may be neglected.
Motion on scales much larger than the deformation radius.

Although scales are not allowed to become so large that \( Ro(L/L_d)^2 \) is of order unity, we may, \textit{a posteriori}, still have \( L \gg L_d \), whence the potential vorticity equation becomes

\[
\frac{\partial \eta}{\partial t} + u \psi \cdot \nabla \eta = 0 \quad \text{or} \quad \frac{\partial \eta}{\partial t} + J(\psi, \eta) = 0. \tag{5.74}
\]

However, because \( \psi = g \eta / f_0 \), the Jacobian term vanishes. Thus, one is left with a trivial equation that implies there is no adveotive evolution of the height field. There is nothing wrong with our reasoning; the mathematics have indeed pointed out a limit interesting in its uninterestingness. From a physical point of view, however, such (lack of) motion is likely to be rare, because on such large scales the Coriolis parameter varies considerably, and we are led to the planetary geostrophic equations.

In practice, often the most severe restriction of quasi-geostrophy is that variations in layer thickness are small: what does this have to do with geostrophy? If we scale \( \eta \) assuming geostrophic balance then \( \eta \sim fUL/g \) and \( \eta/H \sim Ro(L/L_d)^2 \). Thus, if \( Ro \) is to remain small, \( \eta/H \) can only be order one if \( (L/L_d)^2 \gg 1 \). That is, the height variations must occur on a large scale, or we are led to a scaling inconsistency. Put another way, if there are order-one height variations over a length-scale less than or order of the deformation scale, the Rossby number will not be small. Large height variations are allowed if the scale of motion is large, but this contingency is described by the planetary geostrophic equations.

Another flow regime

Although perhaps of little terrestrial interest, we can imagine a regime in which the Coriolis parameter varies fully, but the scale of motion remains no larger than the deformation radius. This parameter regime is not quasigeostrophic, but it gives an interesting result. Because \( \eta'/H \sim Ro(L/L_d)^2 \) deviations of the height field are at least order Rossby number smaller than the reference height and \( |\eta'| \ll H \). The dominant balance in height equation is then

\[
H \nabla \cdot u = 0, \tag{5.75}
\]

presuming that time still scales advectively. This zero horizontal divergence must remain consistent with geostrophic balance

\[
f \times u = -g \nabla \eta, \tag{5.76}
\]

where now \( f \) is a fully variable Coriolis parameter. Taking the curl of (i.e., cross-differentiating) (5.76) gives

\[
\beta v + f \nabla \cdot u = 0, \tag{5.77}
\]

whence, using (5.75), \( v = 0 \), and the flow is purely zonal. Although not at all useful as an evolution equation, this illustrates the constraining effect that differential rotation has on meridional velocity. This effect may be the cause of the banded, highly zonal flow on some of the the giant planets, and we will revisit this issue in our discussion of geostrophic turbulence.
5.3.2 Two-layer and multi-layer quasi-geostrophic systems

Just as for the one-layer case, the multi-layer shallow water equations simplify to a corresponding quasi-geostrophic system in appropriate circumstances. The assumptions are virtually same as before, although we assume that the variation in the thickness of each layer is small compared to its mean thickness. The basic fluid system for a two-layer case is sketched in Fig. 5.1 (and see also Fig. 3.5), and for the multi-layer case in Fig. 5.2.

Let us proceed directly from the potential vorticity equation for each layer. We will also stay in dimensional variables, foregoing a strict asymptotic approach for the sake of informality and insight, and use the Boussinesq approximation. For each layer the potential vorticity equation is just

$$\frac{DQ_i}{Dt} = 0, \quad Q_i = \frac{\zeta_i + f}{h_i}. \quad (5.78)$$

Let $h_i = H_i + h_i'$ where $|h_i'| \ll H_i$. The potential vorticity then becomes

$$Q_i \approx \frac{1}{H_i}(\zeta_i + f) \left(1 - \frac{h_i'}{H_i}\right) \quad \text{— variations in layer thickness are small} \quad (5.79a)$$

$$\approx \frac{1}{H_i} \left(f + \zeta_i - f \frac{h_i'}{H_i}\right) \quad \text{— the Rossby number is small} \quad (5.79b)$$

$$\approx \frac{1}{H_i} \left(f + \zeta_i - f_0 \frac{h_i'}{H_i}\right) \quad \text{— variations in Coriolis parameter are small} \quad (5.79c)$$
Now, because \( Q \) appears in the equations only as an advected quantity, it is only the variations in Coriolis parameter that are important in the first term on the right-hand side of (5.79c), and given this all three terms are of the same approximate magnitude. Then, because mean layer thicknesses are constant, we can define the quasi-geostrophic potential vorticity in each layer by

\[
q_i = \left( \beta y + \zeta_i - f_0 \frac{h'_i}{H_i} \right),
\]

and this will evolve according to \( Dq_i /Dt = 0 \), where the advective derivative is by the geostrophic wind. As in the one-layer case, quasi-geostrophic potential vorticity has different dimensions from the full shallow water potential vorticity.

**Two-layer model**

To obtain a closed set of equations we must obtain an advecting field from the potential vorticity. We use geostrophic balance to do this, and neglecting the advective derivative in (3.51) gives

\[
f_0 \times u_1 = -g \nabla \eta_0 = -g \nabla (h'_1 + h'_2 + \eta_b),
\]

\[
f_0 \times u_2 = -g \nabla \eta_0 - g' \nabla \eta_1 = -g \nabla (h'_1 + h'_2 + \eta_b) - g' \nabla (h_1 + \eta_b),
\]

where \( g' = (\rho_2 - \rho_1) / \rho_1 \) and \( \eta_b \) is the height of any bottom topography, and, because variations in the Coriolis parameter are presumptively small, we use a constant value of \( f \) (i.e., \( f_0 \)) on the left-hand side. For each layer there is therefore a streamfunction, given by

\[
\psi_1 = \frac{g}{f_0} (h'_1 + h'_2 + \eta_b), \quad \psi_2 = \frac{g}{f_0} (h'_1 + h'_2 + \eta_b) + \frac{g'}{f_0} (h'_2 + \eta_b),
\]

and these two equations may be manipulated to give

\[
h'_1 = \frac{f_0}{g'} (\psi_1 - \psi_2) + \frac{f_0}{g} \psi_1, \quad h'_2 = \frac{f_0}{g'} (\psi_2 - \psi_1) - \eta_b.
\]

We note as an aside that the interface displacements are given by

\[
\eta'_0 = \frac{f_0}{g} \psi_1, \quad \eta'_1 = \frac{f_0}{g'} (\psi_2 - \psi_1).
\]

Using (5.80) and (5.83) the quasi-geostrophic potential vorticity for each layer becomes

\[
q_1 = \beta y + \nabla^2 \psi_1 + \frac{f_0^2}{g' H_1} (\psi_2 - \psi_1) + \frac{f_0^2}{g H_1} \psi_1
\]

\[
q_2 = \beta y + \nabla^2 \psi_2 + \frac{f_0^2}{g' H_2} (\psi_1 - \psi_2) + f_0 \frac{\eta_b}{H_2}
\]

In the rigid-lid approximation the last term in (5.85b) is neglected. The potential vorticity in each layer is just advected by the geostrophic velocity, so that the evolution equation for each layer is just

\[
\frac{\partial q_i}{\partial t} + f(\psi_i, q_i) = 0, \quad i = 1, 2.
\]
Fig. 5.2 A multi-layer quasi-geostrophic fluid system. Layers are numbered from the top down, $i$ denotes a general interior layer and $N$ denotes the bottom layer.

*Multi-layer model*

A multi-layer quasi-geostrophic model may be constructed by a straightforward extension of the above two-layer procedure (see Fig. 5.2). The quasi-geostrophic potential vorticity for each layer is still given by (5.80). The pressure field in each layer can be expressed in terms of the thickness of each layer using (3.46) and (3.47) on page 132 and by geostrophic balance the pressure is proportional to the streamfunction, $\psi$, for each layer. Carrying out these steps we obtain, after a little algebra, the following expression for the quasi-geostrophic potential vorticity of an interior layer, in the Boussinesq approximation:

$$q_i = \beta y + \nabla^2 \psi_i + f_0^2 H_i \left( \frac{\psi_{i-1} - \psi_i}{g_{i-1}} - \frac{\psi_i - \psi_{i+1}}{g_i} \right),$$

(5.87)

and for the top and bottom layers,

$$q_1 = \beta y + \nabla^2 \psi_1 + \frac{f_0^2}{H_1} \left( \frac{\psi_2 - \psi_1}{g_1} \right) + \frac{f_0^2}{g H_2} \psi_1,$$

(5.88a)

$$q_N = \beta y + \nabla^2 \psi_N + \frac{f_0^2}{H_N} \left( \frac{\psi_{N-1} - \psi_N}{g_{N-1}} \right) + \frac{f_0}{H_N} \eta_b.$$

(5.88b)

In these equations $H_i$ is the basic-state thickness of the $i$'th layer, and $g_i' = g(\rho_{i+1} - \rho_i)/\rho_i$. In each layer the evolution equation is (5.86), now for $i = 1 \cdots N$. The displacements of each interface are given, similarly to (5.84), by

$$\eta_0' = \frac{f_0}{g} \psi_1, \quad \eta_i' = \frac{f_0}{g_i} (\psi_{i+1} - \psi_i).$$

(5.89a,b)
5.3.3 †Non-asymptotic and intermediate models

The form of the derivation of the previous section suggests that we might be able to improve on the accuracy and range of applicability of the quasi-geostrophic equations, whilst still filtering gravity waves. For example, a seemingly improved set of geostrophic evolution equations might be

\[ \frac{\partial q_i}{\partial t} + u_i \cdot \nabla q_i = 0, \]  

(5.90)

with

\[ q_i = f + \zeta_i h_i, \quad \zeta_i = \frac{\partial v_i}{\partial x} - \frac{\partial u_i}{\partial y}, \]  

(5.91a,b)

and with the velocities given by geostrophic balance, and therefore a function of the layer depths. Thus, the vorticity, height, and velocity fields may all be inverted from potential vorticity. Note that the inversion does not involve the linearization of potential vorticity about a resting state [compare (5.91a) with (5.80)], and we might also choose to keep the full variation of the Coriolis parameter in (5.81). Thus, the model consisting of (5.90) and (5.91) contains both the planetary geostrophic and quasi-geostrophic equations. However, the informality of the derivation hides the fact that this is not an asymptotically consistent set of equations: it mixes asymptotic orders in the same equation, and good conservation properties are not assured. The set above does not, in fact, exactly conserve energy. Models that are either more accurate or more general than the quasi-geostrophic or planetary geostrophic equations yet that still filter gravity waves are called ‘intermediate models’.5

A model that is derived asymptotically will, in general, maintain the conservation properties of the original set. To see this, albeit in a rather abstract way, suppose that the original equations (e.g., the primitive equations) may be written in non-dimensional form, as

\[ \frac{\partial \phi}{\partial t} = F(\phi, \epsilon) \]  

(5.92)

where \( \phi \) is a set of variables, \( F \) is some operator, and \( \epsilon \) is a small parameter, like the Rossby number. Suppose also that this set of equations has various invariants (such as energy and potential vorticity) that hold for any value of \( \epsilon \). The asymptotically-derived lowest order model (such as quasi-geostrophy) is simply a version of this equation set valid in the limit \( \epsilon = 0 \), and therefore it will preserve the invariants of the original set. These invariants may seem to have a different form in the simplified set: for example, in deriving the hydrostatic primitive equations from the Navier-Stokes equations the small parameter is the aspect ratio, and this multiplies the vertical velocity. Thus, in the limit of zero aspect ratio, and therefore in the primitive equations, the conserved kinetic energy has contributions only from the horizontal velocity. In other cases, some invariants may be reduced to trivialities in the simplified set. On the other hand, there is nothing to preclude new invariants emerging that hold only in the limit \( \epsilon = 0 \), and enstrophy (considered later in this chapter) is one example.6

5.4 THE CONTINUOUSLY STRATIFIED QUASI-GEOSTROPHIC SYSTEM

We now consider the quasi-geostrophic equations for the continuously stratified hydrostatic system. The primitive equations of motion are given by (5.15), and
we extract the mean stratification so that the thermodynamic equation is given by (5.17). We also stay on the $\beta$-plane for simplicity. Readers who wish for a briefer, more informal derivation may peruse the box on page 224; however, it is important to realize that there is a systematic asymptotic derivation of the quasi-geostrophic equations, for it is this that ensures that the resulting equations have good conservation properties, as explained in section 5.3.3.

### 5.4.1 Scaling and assumptions

The scaling assumptions we make are just those we made for the shallow water system on page 209, with a deformation radius now given by $L_d = NH/f_0$. The nondimensionalization and scaling is initially precisely that of section 5.1.2, and so we obtain the following non-dimensional equations:

**Horizontal momentum:**
\[
\frac{D\hat{u}}{Dt} + \hat{f} \times \hat{u} = -\nabla_z \hat{\phi},
\]
\[(5.93)\]

**Hydrostatic:**
\[
\frac{\partial \hat{\phi}}{\partial \hat{z}} = \hat{b},
\]
\[(5.94)\]

**Mass continuity:**
\[
\frac{\partial \hat{u}}{\partial \hat{x}} + \frac{\partial \hat{v}}{\partial \hat{y}} + \frac{1}{\tilde{\rho}} \frac{\partial \tilde{\rho} \hat{w}}{\partial \hat{z}} = 0,
\]
\[(5.95)\]

**Thermodynamic:**
\[
\frac{D\hat{b}}{Dt} + \left(\frac{L_d}{L}\right)^2 \tilde{\omega} = 0.
\]
\[(5.96)\]

In Cartesian coordinates we may express the Coriolis parameter as
\[
f = f_0 + \beta y \mathbf{k}
\]
where $f_0 = f_0 \mathbf{k}$. The variation of the Coriolis parameter is assumed small (this is a key difference between the quasi-geostrophic system and the planetary geostrophic system), and in particular we shall assume that $\beta y$ is approximately the size of the relative vorticity, and so much smaller than $f_0$ itself. Thus,
\[
\beta y \sim \frac{U}{L}, \quad \beta \sim \frac{U}{L^2},
\]
\[(5.97)\]

and so we define an $O(1)$ non-dimensional beta parameter by
\[
\hat{\beta} = \frac{\beta L^2}{U} = \frac{\beta L}{Ro f_0}.
\]
\[(5.98)\]

From this it follows that if $f = f_0 + \beta y$, the corresponding nondimensional version is
\[
\hat{f} = \hat{f}_0 + Ro \hat{\beta} \hat{y}.
\]
\[(5.99)\]

where $\hat{f} = f / f_0$ and $\hat{f}_0 = f_0 / f_0 = 1$.

### 5.4.2 Asymptotics

We now expand the nondimensional dependent variables in an asymptotic series in Rossby number, and write
\[
\hat{u} = \hat{u}_0 + Ro \hat{u}_1 + \cdots, \quad \hat{\phi} = \hat{\phi}_0 + Ro \hat{\phi}_1 + \cdots, \quad \hat{b} = \hat{b}_0 + Ro \hat{\beta}_1 + \cdots.
\]
\[(5.100)\]
Substituting these into the equations of motion, the lowest order momentum equation is simply geostrophic balance,
\[ \hat{f}_0 \times \hat{u}_0 = -\nabla \hat{\phi}_0 \] (5.102)
with a constant value of the Coriolis parameter. (For the rest of this chapter we drop the subscript \( z \) from the \( \nabla \) operator.) From (5.102) it is evident that
\[ \nabla \cdot \hat{u}_0 = 0. \] (5.103)
Thus, the horizontal flow is, to leading order, non-divergent; this is a consequence of geostrophic balance, and is not a mass conservation equation. Using (5.103) in the mass conservation equation, (5.95), gives
\[ \frac{\partial}{\partial \hat{z}} (\tilde{\rho} \hat{w}_0) = 0, \] (5.104)
which implies that if \( w_0 \) is zero somewhere (e.g., at a solid surface) then \( w_0 \) is zero everywhere (essentially the Taylor-Proudman effect). A physical way of saying this is that the scaling estimate \( W = UH/L \) is an overestimate of the size of the vertical velocity, because even though \( \partial w/\partial z \approx -\nabla \cdot \mathbf{u} \), the horizontal divergence of the geostrophic flow is small if \( f \) is nearly constant and \( |\nabla \cdot \mathbf{u}| \ll U/L \). We might have anticipated this from the outset, and scaled \( w \) differently, perhaps using the geostrophic vorticity balance estimate, \( w \sim \beta U H/f_0 = Ro U H/L \) as the scaling factor for \( w \), but there is no a priori guarantee that this would be correct.

At next order the momentum equation is
\[ \frac{D_0 \hat{u}_0}{Dt} + \hat{\beta} \hat{y} \times \hat{u}_0 + \hat{f} \times \hat{u}_1 = -\nabla \hat{\phi}_1, \] (5.105)
where \( D_0/Dt = \partial/\partial \hat{t} + (\hat{u}_0 \cdot \nabla) \), and the next order mass conservation equation is
\[ \nabla \cdot (\tilde{\rho} \hat{u}_1) + \frac{\partial}{\partial \hat{z}} (\tilde{\rho} \hat{w}_1) = 0. \] (5.106)

From (5.96), the lowest order thermodynamic equation is just
\[ \left( \frac{L_d}{L} \right)^2 \tilde{w}_0 = 0 \] (5.107)
provided that, as we have assumed, the scales of motion are not sufficiently large that \( Ro(L/L_d)^2 = O(1) \). (This is a key difference between quasi-geostrophy and planetary geostrophy.) At next order we obtain an evolution equation for the buoyancy, and this is
\[ \frac{D_0 \hat{b}_0}{Dt} + \hat{w}_1 \left( \frac{L_d}{L} \right)^2 = 0. \] (5.108)

**The potential vorticity equation**
To obtain a single evolution equation for lowest order quantities we eliminate \( w_1 \) between the thermodynamic and momentum equations. Cross differentiating the first order momentum equation (5.105) gives the vorticity equation,
\[ \frac{\partial \hat{\zeta}_0}{\partial \hat{t}} + (\hat{u}_0 \cdot \nabla) \hat{\zeta}_0 + \hat{v}_0 \hat{\beta} = -\hat{f}_0 \nabla \cdot \hat{u}_1. \] (5.109)
(In dimensional terms, the divergence on the right-hand side is small, but is multi-
plied by the large term $f_0$, and their product is the same order as the terms on the
left-hand side.) Using the mass conservation equation (5.106), (5.109) becomes
\[
\frac{D_0}{Dt} (\zeta_0 + \hat{f}) = \frac{\hat{f}_0}{\rho_0} \frac{\partial}{\partial z} (w_1 \rho_1)
\]  
(5.110)

Combining (5.110) and (5.108) gives
\[
\frac{D_0}{Dt} (\zeta_0 + \hat{f}) = -\frac{\hat{f}_0}{\rho_0} \frac{\partial}{\partial z} \left( F \rho \hat{b}_0 \right)
\]  
(5.111)

where $F \equiv (L/L_d)^2$. The right-hand side of this equation is
\[
\frac{\partial}{\partial z} \left( \frac{D_0}{Dt} \hat{b}_0 \right)
\]  
(5.112)

The second term on the right-hand side vanishes identically using the thermal wind
equation
\[
k \times \frac{\partial \hat{u}_0}{\partial z} = -\frac{1}{f_0} \nabla \hat{b}_0,
\]  
(5.113)

and so (5.111) becomes
\[
\frac{D_0}{Dt} \left[ \zeta_0 + \hat{f} + \frac{\hat{f}_0}{\rho_0} \frac{\partial}{\partial z} \left( \rho F \hat{b}_0 \right) \right] = 0
\]  
(5.114)

or, after using the hydrostatic equation,
\[
\frac{D_0}{Dt} \left[ \zeta_0 + \hat{f} + \frac{\hat{f}_0}{\rho_0} \frac{\partial}{\partial z} \left( \rho F \frac{\partial \tilde{\phi}_0}{\partial z} \right) \right] = 0
\]  
(5.115)

Since the lowest-order horizontal velocity is divergence-free, we can define a
streamfunction $\tilde{\psi}$ such that
\[
\hat{u}_0 = -\frac{\partial \tilde{\psi}}{\partial \hat{y}}, \quad \hat{v}_0 = \frac{\partial \tilde{\psi}}{\partial \hat{x}}
\]  
(5.116)

where also, using (5.102), $\phi_0 = \hat{f}_0 \tilde{\psi}$. The vorticity is then given by $\zeta_0 = \nabla^2 \tilde{\psi}$ and
(5.115) becomes a single equation in a single unknown, to wit
\[
\frac{D_0}{Dt} \left[ \nabla^2 \tilde{\psi} + \hat{f} \hat{y} + \frac{\hat{f}_0}{\rho_0} \frac{\partial}{\partial z} \left( \rho F \frac{\partial \tilde{\psi}}{\partial z} \right) \right] = 0
\]  
(5.117)

where the material derivative is evaluated using $\hat{u}_0 = k \times \nabla \tilde{\psi}$. This is the nondi-
imensional form of the quasi-geostrophic potential vorticity equation, one of the
most important equations in dynamical meteorology and oceanography. In deriving
it we have reduced the Navier Stokes equations, which are six coupled nonlinear
partial differential equations in six unknowns $(u, v, w, T, p, \rho)$ to a single (albeit nonlinear) first-order partial differential equation in a single unknown.\(^8\)
The dimensional version of the quasi-geostrophic potential vorticity equation may be written,

$$\frac{Dq}{Dt} = 0, \quad q = \nabla^2 \psi + f + \frac{f_0^2}{\tilde{\rho}} \frac{\partial}{\partial z} \left( \frac{\tilde{\rho}}{N^2} \frac{\partial \psi}{\partial z} \right).$$

(5.118a,b)

where only the variable part of \( f \) (e.g., \( \beta y \)) is relevant in the second term on the right-hand side of the expression for \( q \). The quantity \( q \) is known as the quasi-geostrophic potential vorticity. It is analogous to the exact (Ertel) potential vorticity (see section 5.5 for more about this), and it is conserved when advected by the horizontal geostrophic flow. All the other dynamical variables may be obtained from potential vorticity as follows:

(i) Streamfunction, using (5.118b).

(ii) Velocity: \( u = k \times \nabla \psi \equiv \nabla \times (k \psi) \).

(iii) Relative vorticity: \( \zeta = \nabla^2 \psi \).

(iv) Perturbation pressure: \( \phi = f_0 \psi \).

(v) Perturbation buoyancy: \( b' = f_0 \frac{\partial \psi}{\partial z} \).

The length-scale \( L_d = NH/f_0 \), emerges naturally from the QG dynamics. It is the scale at which buoyancy and relative vorticity effects contribute equally to the potential vorticity, and is called the deformation radius; it is analogous to the quantity \( \sqrt{gH/f_0} \) arising in shallow water theory. In the upper ocean, with \( N \approx 10^{-2} \text{s}^{-1} \), \( H \approx 10^3 \text{m} \), and \( f_0 \approx 10^{-4} \text{s}^{-1} \), then \( L_d \approx 100 \text{km} \). At high latitudes the ocean is much less stratified and \( f \) is somewhat larger, and the deformation radius may be as little as 30 km (see Fig. 9.11 on 407, where the deformation radius is defined slightly differently). In the atmosphere, with \( N \approx 10^{-2} \text{s}^{-1} \), \( H \approx 10^4 \text{m} \), then \( L_d \approx 1000 \text{km} \). It is this order of magnitude difference in the deformation scales that accounts for a great deal of the quantitative difference in the dynamics of the ocean and atmosphere. If we take the limit \( L_d \to \infty \) then the stratified quasi-geostrophic equations reduce to

$$\frac{Dq}{Dt} = 0, \quad q = \nabla^2 \psi + f$$

(5.119)

This is the two-dimensional vorticity equation, identical to (4.69). The high stratification of this limit has suppressed all vertical motion, and variations in the flow become confined to the horizontal plane. Finally, we note that it is typical in quasi-geostrophic applications to omit the prime on the buoyancy perturbations, and write \( b = f_0 \frac{\partial \psi}{\partial z} \); however, we will keep the prime in this chapter.

### 5.4.3 Buoyancy advection at the surface

The solution of the elliptic equation in (5.118) requires vertical boundary conditions on \( \psi \) at the ground and at the top of the atmosphere, and these are given by use of the thermodynamic equation. For a flat, slippery, rigid surface the vertical velocity
is zero so that the thermodynamic equation may be written

\[
\frac{D b'}{D t} = 0, \quad b' = f_0 \frac{\partial \psi}{\partial z}. \tag{5.120}
\]

We apply this at the ground and at the tropopause, treating the latter as a lid on the lower atmosphere. In the presence of friction and topography the vertical velocity is not zero, but is given by

\[
w = r \nabla^2 \psi + u \cdot \nabla \eta_b \tag{5.121}
\]

where the first term represents Ekman friction (with the constant \(r\) proportional to the thickness of the Ekman layer) and the second term represents topographic forcing. The boundary condition becomes

\[
\frac{\partial}{\partial t} \left( \frac{\partial \psi}{\partial z} \right) + u \cdot \nabla \left( \frac{\partial \psi}{\partial z} + N^2 \eta_b \right) + N^2 r \nabla^2 \psi = 0, \tag{5.122}
\]

where all the fields are evaluated at \(z = 0\) and \(z = H\), the height of the lid. Thus, the quasi-geostrophic system is characterized by the horizontal advection of potential vorticity in the interior and the advection of buoyancy at the boundary. Instead of a lid at the top, then in a compressible fluid like the atmosphere we may suppose that all disturbances tend to zero as \(z \to \infty\).

*A potential vorticity sheet at the boundary*

Rather than regarding buoyancy advection as providing the boundary condition, it is sometimes useful to think of there being a very thin sheet of potential vorticity just above the ground and another just below the lid, specifically with a vertical distribution proportional to \(\delta(z - \epsilon)\) or \(\delta(z - H + \epsilon)\). The boundary condition \(\frac{\partial \psi}{\partial z} = 0\) at \(z = 0\) and \(z = H\) can be replaced by this, along with the condition that there are no variations of buoyancy at the boundary and \(\frac{\partial \psi}{\partial z} = 0\) at \(z = 0\) and \(z = H\).

To see this, we first note that the differential of a step function is a delta function. Thus, a discontinuity in \(\frac{\partial \psi}{\partial z}\) at a level \(z = z_1\) is equivalent to a delta function in potential vorticity there:

\[
q(z_1) = \left[ \frac{f_0^2}{N^2} \frac{\partial \psi}{\partial z} \right]_{z_1^-}^{z_1^+} \delta(z - z_1). \tag{5.123}
\]

Now, suppose that the lower boundary condition, given by \(\frac{\partial \psi}{\partial z} = 0\), has some arbitrary distribution of buoyancy on it. We can replace this condition by the simpler condition \(\frac{\partial \psi}{\partial z} = 0\) at \(z = 0\), provided we also add to our definition of potential vorticity a term given by \(\frac{f_0^2}{N^2} \frac{\partial \psi}{\partial z} \delta(z - z_1)\) with \(z_1 = \epsilon\). This term is then advected by the horizontal flow, as are the other contributions. A buoyancy source at the boundary must similarly be treated as a sheet of potential vorticity source in the interior. Any flow with buoyancy variations over a horizontal boundary is thus equivalent to a flow with uniform buoyancy at the boundary, but with a spike in potential vorticity adjacent to the boundary. This approach brings notational and conceptual advantages, in that now everything is expressed in terms of potential vorticity and its advection. However, in practice there may be less to be gained, because the boundary terms must still be included in any particular calculation that is to be performed.
5.4.4 Quasi-geostrophy in pressure coordinates

The derivation of the quasi-geostrophic system in pressure coordinates is very similar to that in height coordinates, with the main difference coming at the boundaries, and we give only the results. The starting point is the primitive equations in pressure coordinates, (2.151). In pressure coordinates quasi-geostrophic potential vorticity is found to be

\[ q = f + \nabla^2 \psi + \frac{\partial}{\partial p} \left( \frac{f_0^2 \partial \psi}{S_2^2 \partial p} \right), \tag{5.124} \]

where \( \psi = \Phi / f_0 \) is the streamfunction and \( \Phi \) the geopotential, and

\[ S_2^2 \equiv -\frac{R}{p} \left( \frac{p}{p_R} \right)^{\kappa} \frac{d \tilde{\theta}}{d p} = -\frac{1}{\rho} \frac{d \tilde{\theta}}{d p} \tag{5.125} \]

where \( \tilde{\theta} \) is a reference profile, a function of pressure only. In log-pressure coordinates, with \( Z = -H \ln p \), the potential vorticity may be written

\[ q = f + \nabla^2 \psi + \frac{1}{\rho_*} \frac{\partial}{\partial Z} \left( \rho_* f_0^2 \frac{\partial \psi}{N_Z^2 \partial Z} \right), \tag{5.126} \]

where

\[ N_Z^2 = S_2^2 \left( \frac{p}{H} \right)^2 = -\left( \frac{R}{H} \right) \left( \frac{p}{p_R} \right)^{\kappa} \frac{d \tilde{\theta}}{d Z} \tag{5.127} \]

is the buoyancy frequency and \( \rho_* = \exp(-z/H) \). Temperature and potential temperature are related to the streamfunction by

\[ T = -\frac{f_0}{g} \frac{\partial \psi}{\partial p} = \frac{H f_0}{R} \frac{\partial \psi}{\partial Z}, \tag{5.128a} \]

\[ \theta = -\left( \frac{p_R}{p} \right)^{\kappa} \left( \frac{f_0}{R} \right) \frac{\partial \psi}{\partial p} = \left( \frac{p_R}{p} \right)^{\kappa} \left( \frac{H f_0}{R} \right) \frac{\partial \psi}{\partial Z}. \tag{5.128b} \]

In pressure or log-pressure co-ordinates, potential vorticity is advected along isobaric surfaces, analogous to the horizontal advection in height co-ordinates.

The surface boundary condition again is derived from the thermodynamic equation. This is, in log-pressure coordinates,

\[ \frac{D}{Dt} \left( \frac{\partial \psi}{\partial Z} \right) + \frac{N_Z^2}{f_0} W = 0. \tag{5.129} \]

where \( W = DZ/Dt \). This is not the real vertical velocity, \( w \), but it is related to it by

\[ w = \frac{f_0}{g} \frac{\partial \psi}{\partial t} + \frac{RT}{gH} W. \tag{5.130} \]

Thus, choosing \( H = RT(0)/g \), we have, at \( Z = 0 \),

\[ \frac{\partial}{\partial t} \left( \frac{\partial \psi}{\partial Z} \right) - \frac{N_Z^2}{g} \psi + \mathbf{u} \cdot \nabla \psi = -\frac{N_Z^2}{f_0} w, \tag{5.131} \]

where

\[ w = \mathbf{u} \cdot \nabla \eta_b + r \nabla^2 \psi. \tag{5.132} \]

This differs from the expression in height coordinates only by the second term in the local time derivative. In applications where accuracy is not the main issue the simpler boundary condition \( D(\tilde{\psi}_Z)/Dt = 0 \) is sometimes used.
Informal Derivation of Stratified QG Equations

We will use the Boussinesq equations, but similar derivations could be given for the anelastic equations or pressure coordinates. The first ingredient is the vertical component of the vorticity equation, (4.68); in the Boussinesq version (or the pressure coordinate or anelastic versions) there is no baroclinic term and we have:

$$\frac{D}{Dt}(\zeta + f) = -\left(\zeta + f\right) \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + \left(\frac{\partial u}{\partial z} \frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \frac{\partial w}{\partial x}\right).$$  \hspace{1cm} (QG.1)

We now apply the assumptions on page 209. The advection and the vorticity on the left-hand side are geostrophic, but we keep the horizontal divergence (which is small) on the right-hand side where it is multiplied by the big term $f$. Furthermore, because $f$ is nearly constant we replace it with $f_0$ except where it is differentiated. The second term (tilting) on the right-hand side is smaller than the advection terms on the left-hand side by the ratio $[UW/(HL)]/[U^2/L^2] = [W/H]/[U/L] \ll 1$, because $w$ is small ($\partial w/\partial z$ equals the divergence of the ageostrophic velocity). We therefore neglect it, and given all this (QG.1) becomes

$$\frac{D}{Dt}(\zeta_g + f) = -f_0 \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = f_0 \frac{\partial w}{\partial z},$$ \hspace{1cm} (QG.2)

where the second equality uses mass continuity and $D/Dt = \partial/\partial t + \mathbf{u}_g \cdot \nabla$.

The second ingredient is the three-dimensional thermodynamic equation,

$$D_3b/Dt = 0.$$ \hspace{1cm} (QG.3)

The stratification is assumed nearly constant, so we write $b = \tilde{b}(z) + b'(x,y,z,t)$, where $\tilde{b}$ is the basic state buoyancy. Furthermore, because $w$ is small it only advects the basic state, and with $N^2 = \partial \tilde{b}/\partial z$ (QG.3) becomes

$$D_3b'/Dt + wN^2 = 0.$$ \hspace{1cm} (QG.4)

Hydrostatic and geostrophic wind balance enable us to write the geostrophic velocity, vorticity, and buoyancy in terms of streamfunction $\psi$ [$= p/(f_0 \rho_0)$]:

$$\mathbf{u}_g = \mathbf{k} \times \nabla \psi, \quad \zeta_g = \nabla^2 \psi, \quad b' = f_0 \partial \psi/\partial z.$$ \hspace{1cm} (QG.5)

The quasi-geostrophic potential vorticity equation is obtained by eliminating $w$ between (QG.2) and (QG.4), and this gives

$$\frac{D_3q}{Dt} = 0, \quad q = \zeta_g + f + \frac{\partial}{\partial z} \left(\frac{f_0 b'}{N^2}\right).$$ \hspace{1cm} (QG.6)

This equation is the Boussinesq version of (5.118), and using (QG.5) it may be expressed entirely in terms of the streamfunction, with $D_3/ Dt = \partial/\partial t + J(\psi, \cdot)$. The vertical boundary conditions, at $z = 0$ and $z = H$ say, are given by (QG.4) with $w = 0$, with straightforward generalizations if topography or friction are present.
5.4.5 The two-level quasi-geostrophic system

The quasi-geostrophic system has, in general, continuous variation in the vertical (and horizontal, of course). By finite-differencing the continuous equations we can obtain a multi-level model, and a crude but important special case of this is the two-level model, which allows only two-degrees of freedom in the vertical. To obtain the equations of motion one way to proceed is to take a crude finite difference of the continuous relation between potential vorticity and streamfunction given in (5.118b). In the Boussinesq case (or in pressure coordinates, with a slight reinterpretation of the meaning of the symbols) the continuous expression for potential vorticity is

\[ q = \zeta + f + \frac{\partial}{\partial z} \left( \frac{f_0 b'}{N^2} \right), \quad (5.133) \]

where \( b' = f_0 \delta \psi / \delta z \). In the case with a flat bottom and rigid lid at the top (and incorporating topography is an easy extension) the boundary condition of \( w = 0 \) is satisfied by \( D \delta \psi / Dt = 0 \) at the top and bottom. An obvious finite differencing of (5.133) in the vertical (see Fig. [5.3]) then gives

\[ q_1 = \zeta_1 + f + \frac{2 f_0^2}{N^2 H_1 H} (\psi_2 - \psi_1), \quad q_2 = \zeta_2 + f + \frac{2 f_0^2}{N^2 H_2 H} (\psi_1 - \psi_2). \quad (5.134) \]

In atmospheric problems it is common to choose \( H_1 = H_2 \), whereas in oceanic problems we might choose to have a thinner upper layer, representing the flow above the main thermocline. Note that the boundary conditions of \( w = 0 \) at the top and bottom are already taken care of (5.134): they are incorporated into the definition of the potential vorticity — a finite-difference analog of the delta-function construction of section 5.4.3. At each level the potential vorticity is advected by the streamfunction so that the evolution equation for each level is:

\[ \frac{Dq_i}{Dt} = \frac{\partial q_i}{\partial t} + u_i \cdot \frac{\partial q_i}{\partial t} + J(\psi_i, q_i) = 0, \quad i = 1, 2. \quad (5.135) \]
Models with more than two levels can be easily constructed by extending the finite-differencing procedure.

Connection to the layered system

The two-level expressions, (5.134), have an obvious similarity with the two-layer expressions, (5.85). Noting that \( N^2 = \partial \hat{b}/\partial z \) and that \( b = -g\delta \rho/\rho_0 \) it is natural to let

\[
N^2 = -\frac{g}{\rho_0} \frac{\rho_1 - \rho_2}{H/2} = \frac{g'}{H/2}.
\]

(5.136)

With this identification we find that (5.134) becomes

\[
q_1 = \zeta_1 + f + \frac{f_0^2}{\beta H_1}(\psi_2 - \psi_1), \quad q_2 = \zeta_2 + f + \frac{f_0^2}{\beta H_2}(\psi_1 - \psi_2).
\]

(5.137)

These expressions are identical with (5.85) in the flat-bottomed, rigid lid case. Similarly, a multi-layered system with \( n \) layers is equivalent to a finite-difference representation with \( n \) levels. It should be said, though, that in the pantheon of quasi-geostrophic models the two-level and two-layer models hold distinguished places.

5.5 * QUASI-GEOSTROPHY AND ERTEL POTENTIAL VORTICITY

When using the shallow water equations, quasi-geostrophic theory could be naturally developed beginning with the expression for potential vorticity. Is such an approach possible for the stratified primitive equations? The answer is yes, although the algebra is more complicated, as we see.

5.5.1 * Using height coordinates

Noting the general expression, (4.121), for potential vorticity in a hydrostatic fluid, potential vorticity in the Boussinesq hydrostatic equations is given by

\[
Q = \left[ (v_x - u_y)b_z - v_x b_x + u_x b_y + f b_z \right],
\]

(5.138)

where the \( x, y, z \) subscripts denote derivatives. Without approximation, we write the stratification as \( b = \tilde{b}(z) + b'(x, y, z, t) \), and (5.138) becomes

\[
Q = [f_0 N^2] + [(\beta y + \zeta)N^2 + f_0 b'_z] + [(\beta y + \zeta)b'_z - (v_x b_x' - u_x b_y')],
\]

(5.139)

where, under quasi-geostrophic scaling, the terms in square brackets are in decreasing order of size. Neglecting the third term, and taking the velocity and buoyancy fields to be in geostrophic and thermal wind balance, we can write the potential vorticity as \( Q \approx \tilde{Q} + Q' \), where \( \tilde{Q} = f_0 N^2 \) and

\[
Q' = (\beta y + \zeta)N^2 + f_0 b'_z = (\beta y + \nabla^2 \psi)N^2 + f_0^2 \frac{\partial \psi}{\partial z}.
\]

(5.140)

The potential vorticity evolution equation is then

\[
\frac{DQ'}{Dt} + u \frac{\partial \tilde{Q}}{\partial z} = 0.
\]

(5.141)
The vertical advection is important only in advecting the basic state potential vorticity $\tilde{Q}$. Thus, after dividing by $N^2$, (5.141) becomes
\[
\frac{\partial q_*}{\partial t} + u_\beta \cdot \nabla q_* + \frac{w}{N^2} \frac{\partial \tilde{q}}{\partial z} = 0,
\]
where
\[
q_* = (\beta y + \zeta) + \frac{f_0}{N^2} b_z'.
\]
This is the approximation to the (perturbation) Ertel potential vorticity in the quasi-geostrophic limit. However, it is not the same as the expression for the quasi-geostrophic potential vorticity, (5.118) and, furthermore, (5.142) involves a vertical advection. (Thus, we might refer to the expression in (5.118) as the ‘quasi-geostrophic pseudo-potential vorticity’, but the prefix ‘quasi-geostrophic’ alone normally suffices.) We can derive (5.118) by eliminating $w$ between (5.142) and the quasi-geostrophic thermodynamic equation $\partial b'/\partial t + u_\beta \cdot \nabla b' + w \partial \tilde{b}/\partial z = 0$.

5.5.2 Using isentropic coordinates
An illuminating and somewhat simpler path from Ertel potential vorticity to the quasi-geostrophic equations goes by way of isentropic coordinates. We begin with the isentropic expression for Ertel potential vorticity for an ideal gas,
\[
Q = \frac{f + \zeta}{\sigma},
\]
where $\sigma = -\partial p/\partial \theta$ is the thickness density (which we will just call the thickness), and in adiabatic flow potential vorticity is advected along isopycnals. We now employ quasi-geostrophic scaling to derive an approximate equation set from this. First assume that variations in thickness are small compared to the reference state, so that
\[
\sigma = \tilde{\sigma}(\theta) + \sigma', \quad |\sigma'| \ll |\sigma|.
\]
and similarly for pressure and density. Assuming also that the variations in Coriolis parameter are small, (5.144) becomes
\[
Q \approx \left[ \frac{f_0}{\tilde{\sigma}} \right] + \left[ \frac{1}{\tilde{\sigma}} (\zeta + \beta y) - \frac{f_0}{\tilde{\sigma}} \frac{\sigma'}{\sigma} \right].
\]
We now use geostrophic and hydrostatic balance to express the terms on the right-hand side in terms of a single variable, noting that the first term does not vary along isentropic surfaces. Hydrostatic balance is
\[
\frac{\partial M}{\partial \theta} = \Pi
\]
where $M = c_p T + g z$ and $\Pi = c_p (p/p_R)^\kappa$. Writing $M = \tilde{M}(\theta) + M'$ and $\Pi = \tilde{\Pi}(\theta) + \Pi'$, where $\tilde{M}$ and $\tilde{\Pi}$ are hydrostatically balanced reference profiles, we obtain
\[
\frac{\partial M'}{\partial \theta} = \Pi' = \frac{d\tilde{\Pi}}{dp'} p' = \frac{1}{\partial \tilde{\rho}} p'
\]
where the last equality follows using the equation of state for an ideal gas and \( \tilde{\rho} \) is a reference profile. The perturbation thickness field may then be written as

\[
\sigma' = -\frac{\partial}{\partial \theta} \left( \tilde{\rho} \theta \frac{\partial M'}{\partial \sigma} \right).
\]  

(5.149)

Geostrophic balance is \( f_0 \times u = -\nabla \phi M' \) where the velocity, and the horizontal derivatives, are along isentropic surfaces. This enables us to define a flow streamfunction by

\[
\psi \equiv \frac{M'}{f_0}.
\]  

(5.150)

and we can then write all the variables in terms of \( \psi \):

\[
\begin{align*}
u &= \frac{\partial \psi}{\partial y}, \\
v &= \frac{\partial \psi}{\partial x}, \\
\zeta &= \nabla^2 \psi, \\
\sigma' &= f_0 \frac{\partial}{\partial \theta} \left( \tilde{\rho} \theta \frac{\partial M'}{\partial \sigma} \right).
\end{align*}
\]  

(5.151)

Using (5.146), (5.150) and (5.151), the quasi-geostrophic system in isentropic coordinates may be written

\[
\begin{align*}
\frac{Dq}{Dt} &= 0 \\
q &= f + \nabla^2 \psi + f_0 \frac{\partial}{\partial \theta} \left( \tilde{\rho} \theta \frac{\partial \psi}{\partial \theta} \right).
\end{align*}
\]  

(5.152a,b)

where the advection of potential vorticity is by the geostrophically balanced flow, along isentropes. The variable \( q \) is an approximation to the second term in square brackets in (5.146), multiplied by \( \tilde{\sigma} \).

**Projection back to physical-space coordinates**

We can recover the height or pressure coordinate quasi-geostrophic systems by projecting (5.152) onto the appropriate coordinate. This is straightforward because, by assumption, the isentropes in a quasi-geostrophic system are nearly flat. Recall that [c.f., (2.143)] a transformation between vertical coordinates may be effected by

\[
\frac{\partial}{\partial x} \bigg|_\theta = \frac{\partial}{\partial x} \bigg|_p + \frac{\partial p}{\partial x} \bigg|_\theta \frac{\partial}{\partial p},
\]  

(5.153)

but the second term is \( \mathcal{O}(Ro) \) smaller than the first because, under quasi-geostrophic scaling, isentropic slopes are small. Thus \( \nabla^2 \psi \) in (5.152b) may be replaced by \( \nabla^2_p \psi \) or \( \nabla^2_z \psi \). The vortex stretching term in (5.152) becomes, in pressure coordinates,

\[
\begin{align*}
\frac{f_0^2}{\tilde{\sigma}} \frac{\partial}{\partial \theta} \left( \tilde{\rho} \theta \frac{\partial \psi}{\partial \theta} \right) &\approx \frac{f_0^2}{\tilde{\sigma}} \frac{d\tilde{\rho}}{d\tilde{\sigma}} \frac{\partial}{\partial \tilde{\rho}} \left( \tilde{\rho} \theta \frac{d\tilde{\psi}}{d\tilde{\sigma}} \right) = \frac{\partial}{\partial \tilde{\rho}} \left( \frac{f_0^2}{S^2} \frac{\partial \tilde{\psi}}{\partial \tilde{\rho}} \right)
\end{align*}
\]  

(5.154)

where \( S^2 \) is given by (5.125). The expression for the quasi-geostrophic potential vorticity in isentropic coordinates is thus approximately equal to the quasi-geostrophic potential vorticity in pressure coordinates. This near-equality holds
because the isentropic expression, (5.152b), does not contain a component proportional to the mean stratification: the second square-bracketed term on the right-hand side (5.146) is the only dynamically relevant one, and its evolution along isentropes is mirrored by the evolution along isobaric surfaces of quasi-geostrophic potential vorticity in pressure coordinates.

5.6 * ENERGETICS OF QUASI-GEOSTROPHY

If the quasi-geostrophic set of equations is to represent a real fluid system in a physically meaningful way, then it should have a consistent set of energetics. In particular, total energy should be conserved, and there should be analogs of kinetic and potential energy and conversion between the two. We now show that such energetic properties do hold, using the Boussinesq set as an example.

Let us write the governing equations as a potential vorticity equation in the interior,

\[
\frac{D}{Dt} \left[ \nabla^2 \psi + \frac{\partial}{\partial z} \left( \frac{f_0^2}{N^2} \frac{\partial \psi}{\partial z} \right) \right] + \beta \frac{\partial \psi}{\partial x} = 0, \quad 0 < z < 1, (5.155)
\]

and buoyancy advection at the boundary,

\[
\frac{D}{Dt} \left( \frac{\partial \psi}{\partial z} \right) = 0, \quad z = 0, 1. (5.156)
\]

For lateral boundary conditions we may assume that \( \psi = \) constant, or impose periodic conditions. If we multiply (5.155) by \( -\psi \) and integrate over the domain, using the boundary conditions, we easily find

\[
\frac{d\hat{E}}{dt} = 0, \quad \hat{E} = \frac{1}{2} \int_V \left[ (\nabla \psi)^2 + \frac{f_0^2}{N^2} \left( \frac{\partial \psi}{\partial z} \right)^2 \right] dV. (5.157a,b)
\]

The term involving \( \beta \) makes no direct contribution to the energy budget. Eq. (5.157) is the fundamental energy equation for quasi-geostrophic motion, and it states that in the absence of viscous or diabatic terms the total energy is conserved. The two terms in (5.157b) can be identified as the kinetic and available potential energy of the flow, where

\[
KE = \frac{1}{2} \int_V (\nabla \psi)^2 dV, \quad APE = \frac{1}{2} \int_V \frac{f_0^2}{N^2} \left( \frac{\partial \psi}{\partial z} \right)^2 dV. (5.158a,b)
\]

The available potential energy may also be written as

\[
APE = \frac{1}{2} \int_V \frac{H}{L_d^2} \left( \frac{\partial \psi}{\partial z} \right)^2 dV, \quad (5.159)
\]

where \( L_d \) is the deformation radius \( NH/f_0 \) and we may choose \( H \) such that \( z \sim H \). At some scale \( L \) the ratio of the kinetic energy to the potential energy is thus, roughly,

\[
\frac{KE}{APE} \sim \frac{L^2}{L_d^2}. (5.160)
\]

For scales much larger than \( L_d \) the potential energy dominates the kinetic energy, and contrariwise.
5.6.1 Conversion between APE and KE

Let us return to the vorticity and thermodynamic equations,

$$\frac{D\zeta}{Dt} = f \frac{\partial w}{\partial z}$$  \hspace{1cm} (5.161)

where \(\zeta = \nabla^2 \psi\), and

$$\frac{Db'}{Dt} + N^2 w = 0$$  \hspace{1cm} (5.162)

where \(b' = f_0 \frac{\partial \psi}{\partial z}\). From (5.161) we form a kinetic energy equation namely

$$\frac{1}{2} \frac{d}{dt} \int_V (\nabla \psi)^2 dV = -\int_V f_0 \frac{\partial w}{\partial z} \psi dV = \int_V f_0 w \frac{\partial \psi}{\partial z} dV.$$  \hspace{1cm} (5.163)

From (5.162) we form a potential energy equation, namely

$$\frac{d}{dt} \frac{1}{2} \int_V \frac{f_0^2}{N^2} \left( \frac{\partial \psi_1}{\partial z} \right)^2 dV = -\int_V f_0 w \frac{\partial \psi}{\partial z} dV.$$  \hspace{1cm} (5.164)

Thus, the conversion from APE to KE is represented by

$$\frac{d}{dt} \text{KE} = -\frac{d}{dt} \text{APE} = \int_V f_0 w \frac{\partial \psi}{\partial z} dV.$$  \hspace{1cm} (5.165)

Because the buoyancy is proportional to \(\frac{\partial \psi}{\partial z}\), when warm fluid rises there is a correlation between \(w\) and \(\frac{\partial \psi}{\partial z}\) and available potential energy is converted to kinetic energy. Whether such a phenomenon occurs depends of course on the dynamics of the flow; however, such a conversion is in fact a common feature of geophysical flows, and in particular of baroclinic instability, as we see in chapter 6.

5.6.2 Energetics of two-layer flows

Two-layer or two-level flows are an important special case. For layers of equal thickness let us write the evolution equations as

$$\frac{D}{Dt} \left( \nabla^2 \psi_1 - \frac{1}{2} k_0^2 (\psi_1 - \psi_2) \right) + \beta \frac{\partial \psi_1}{\partial x} = 0$$  \hspace{1cm} (5.166a)

$$\frac{D}{Dt} \left( \nabla^2 \psi_2 - \frac{1}{2} k_0^2 (\psi_1 - \psi_2) \right) + \beta \frac{\partial \psi_2}{\partial x} = 0$$  \hspace{1cm} (5.166b)

where \(k_0^2/2 = (2f_0/NH)^2\). On multiplying these two equations by \(-\psi_1\) and \(-\psi_2\) respectively and integrating over the horizontal domain, the advective term in the material derivatives and the beta term all vanish, and we obtain

$$\frac{d}{dt} \int_A \left[ \frac{1}{2} (\nabla \psi_1)^2 + \frac{1}{2} k_0^2 \psi_1 (\psi_1 - \psi_2) \right] dA = 0,$$  \hspace{1cm} (5.167a)

$$\frac{d}{dt} \int_A \left[ \frac{1}{2} (\nabla \psi_2)^2 - \frac{1}{2} k_0^2 \psi_2 (\psi_1 - \psi_2) \right] dA = 0.$$  \hspace{1cm} (5.167b)

Adding these gives

$$\frac{d}{dt} \int_A \left[ \frac{1}{2} (\nabla \psi_1)^2 + \frac{1}{2} (\nabla \psi_2)^2 + k_0^2 \psi_2^2 \right] dA = 0.$$  \hspace{1cm} (5.168)

This is the energy conservation statement for the two layer model. The first two terms represent the kinetic energy and the last term the available potential energy.
5.6 * Energetics of Quasi-Geostrophy

Energy in the baroclinic and barotropic modes

A useful partitioning of the energy is between the energy in the barotropic and baroclinic modes. The barotropic streamfunction, $\psi$, is the vertically averaged streamfunction and the baroclinic mode is the difference between the streamfunctions in the two layers. That is, for equal layer thicknesses,

$$\psi = \frac{1}{2}(\psi_1 + \psi_2), \quad \tau = \frac{1}{2}(\psi_1 - \psi_2) \quad \text{(5.169)}$$

Substituting (5.169) into (5.168) reveals that

$$\frac{d}{dt} \int_A \left[ (\nabla \psi)^2 + (\nabla \tau)^2 + k^2 \tau^2 \right] d\mathbf{x} = 0 \quad \text{(5.170)}$$

The energy density in the barotropic mode is thus just $(\nabla \psi)^2$, and that in the baroclinic mode is $(\nabla \tau)^2 + k^2 \tau^2$. This partitioning will prove particularly useful when we consider baroclinic turbulence in chapter 9.

5.6.3 Enstrophy conservation

Potential vorticity is advected only by the horizontal flow, and thus it is materially conserved on the horizontal surface at every height and

$$\frac{Dq}{Dt} = \frac{\partial q}{\partial t} + \mathbf{u} \cdot \nabla q = 0. \quad \text{(5.171)}$$

Furthermore, the advecting flow is divergence-free so that $\mathbf{u} \cdot \nabla q = \nabla \cdot (\mathbf{u} q)$. Thus, on multiplying (5.171) by $q$ and integrating over a horizontal domain, $A$, using either no-normal flow or periodic boundary conditions, we straightforwardly obtain

$$\frac{d}{dt} \hat{Z} = 0, \quad \hat{Z} = \frac{1}{2} \int_A q^2 dA. \quad \text{(5.172)}$$

The quantity $\hat{Z}$ is known as the enstrophy, and this is conserved at each height as well as, naturally, over the entire volume.

The enstrophy is just one of an infinity of invariants in quasi-geostrophic flow. Because the potential vorticity of a fluid element is conserved, any function of the potential vorticity must be a material invariant and we can immediately write

$$\frac{D}{Dt} F(q) = 0. \quad \text{(5.173)}$$

To verify that this is true, simply note that (5.173) implies that $(dF/dq)Dq/Dt = 0$, which is true by virtue of (5.171). (However, by virtue of the material advection, the function $F(q)$ need not be differentiable in order for (5.173) to hold.) Each of the material invariants corresponding to different choices of $F(q)$ has a corresponding integral invariant; that is

$$\frac{d}{dt} \int_A F(q) dA = 0. \quad \text{(5.174)}$$

The enstrophy invariant corresponds to choosing $F(q) = q^2$; it plays a particularly important role because, like energy, it is a quadratic invariant, and its presence profoundly alters the behaviour of two-dimensional and quasi-geostrophic flow compared to three-dimensional flow (see section 8.3).
5.7 ROSSBY WAVES

In the final topic of this chapter we consider wave motion in a quasi-geostrophic system. (A brief introduction to wave kinematics is given in the appendix to this chapter.) Although we consider the closely related topics of hydrodynamic instability and wave–mean flow interaction in Part II, Rossby waves are such a fundamental part of geophysical fluid dynamics, and intimately tied to quasi-geostrophic dynamics, that they find a natural place in this chapter.

5.7.1 Waves in a single layer

Consider flow of a single homogeneous layer on a flat-bottomed $\beta$-plane. The unforced, inviscid equation of motion is

$$\frac{D}{Dt}(\zeta + f - \psi/L^2) = 0,$$

where $\zeta = \nabla^2 \psi$ is the vorticity and $\psi$ the streamfunction.

**Infinite deformation radius**

If the scale of motion is much less than the deformation scale then the $\beta$-plane the equation of motion is governed by

$$\frac{D}{Dt}(\zeta + \beta y) = 0.$$

Expanding the material derivative gives

$$\frac{\partial \zeta}{\partial t} + u \cdot \nabla \zeta + \beta v = 0 \quad \text{or} \quad \frac{\partial \zeta}{\partial t} + J(\psi, \zeta) + \beta \frac{\partial \psi}{\partial x} = 0.$$  \label{eq:5.177}

We now linearize this equation — that is, we suppose that the flow consists of a time-independent component (the ‘basic state’) plus a perturbation, with the perturbation being small compared to the mean flow. Such a mean flow must satisfy the time independent equation of motion, and purely zonal flow will do this. For simplicity we choose a flow with no meridional dependence and let

$$\psi = \Psi + \psi'(x, y, t)$$  \label{eq:5.178}

where $\Psi = -Uy$ and $|\psi'| \ll |\Psi|$. (The symbol $U$ represents the zonal flow of the basic state, not a magnitude for scaling purposes.) Substitute \eqref{eq:5.178} into \eqref{eq:5.177} and neglect the nonlinear terms involving products of $\psi'$ to give

$$\frac{\partial \zeta'}{\partial t} + J(\Psi, \zeta') + \beta \frac{\partial \psi'}{\partial x} = 0 \quad \text{or} \quad \frac{\partial}{\partial t}\nabla^2 \psi' + U \frac{\partial \nabla^2 \psi'}{\partial x} + \beta \frac{\partial \psi'}{\partial x} = 0.$$  \label{eq:5.179a,b}

Solutions to this equation may be found in the form of a plane wave,

$$\psi' = \Re \tilde{\psi} e^{(kx + iy - \omega t)},$$  \label{eq:5.180}

where Re indicates the real part of the function (and this will sometimes be omitted if no ambiguity is so-caused). Solutions of the form \eqref{eq:5.180} are valid in a domain with doubly-periodic boundary conditions; solutions in a channel can be obtained...
using a meridional variation of $\sin ly$, with no essential changes to the dynamics. The amplitude of the oscillation is given by $\tilde{\psi}$ and the phase by $kx + ly - \omega t$, where $k$ and $l$ are the $x$- and $y$-wavenumbers and $\omega$ is the frequency of the oscillation.

Substituting (5.180) into (5.179) yields

$$[(\omega - U k)(K^2) + \beta k] \tilde{\psi} = 0,$$

(5.181)

where $K^2 = k^2 + l^2$. For nontrivial solutions this implies

$$\omega = U k - \frac{\beta k}{K^2}.$$  

(5.182)

This is the dispersion relation for Rossby waves. The phase speed, $c_p$, and group velocity, $c_g$, in the $x$-direction are

$$c_p^x = \frac{\omega}{k} = U - \frac{\beta}{K^2}, \quad c_g^x = \frac{\partial \omega}{\partial k} = U + \frac{\beta(k^2 - l^2)}{(k^2 + l^2)^2}.$$  

(5.183a,b)

The velocity $U$ provides a uniform translation, and doppler shifts the frequency. The phase speed in the absence of a mean flow westwards, with waves of longer wavelengths travelling more quickly, and the eastward current speed required to hold the waves of a particular wavenumber stationary (i.e., $c_p^x = 0$) is $U = \beta/K^2$. We discuss the meaning of the group velocity in section 5.7.3.

**Finite deformation radius**

For finite deformation radius the basic state $\Psi = -Uy$ is still a solution of the original equations of motion, but the potential vorticity corresponding to this state is $Q = U y/L_d^2 + \beta y$ and its gradient is $\nabla Q = (\beta + U/L_d^2)j$. The linearized equation of motion is thus,

$$\left(\frac{\partial}{\partial t} + U \frac{\partial}{\partial x}\right)(\nabla^2 \psi' - \psi'/L_d^2) + (\beta + U/L_d^2) \frac{\partial \psi'}{\partial x} = 0.$$  

(5.184)

Substituting $\psi' = \tilde{\psi} e^{i(kx + ly - \omega t)}$ we obtain the dispersion relation,

$$\omega = \frac{k(UK^2 - \beta)}{K^2 + 1/L_d^2} = U k - k \frac{\beta + U/L_d^2}{K^2 + 1/L_d^2}.$$  

(5.185)

The corresponding $x$-components of phase speed and group velocity are

$$c_p^x = U - \frac{\beta + U k_d^2}{K^2 + k_d^2} = \frac{UK^2 - \beta}{K^2 + k_d^2}, \quad c_g^x = U + \frac{(\beta + U k_d^2)(k^2 - l^2 - k_d^2)}{(k^2 + l^2 + k_d^2)^2},$$  

(5.186a,b)

where $k_d = 1/L_d$. The uniform velocity field now no longer provides just a simple Doppler shift of the frequency, nor a uniform addition to the phase speed. From (5.186a) the waves are stationary when $K^2 = K_s^2 = \beta/U$; that is, the current speed required to hold waves of a particular wavenumber stationary is $U = \beta/K^2$. However, this is not simply the magnitude of the phase speed of waves of that wavenumber in the absence of a current — this is given by

$$c_p^x = -\frac{\beta}{K_s^2 + k_d^2} = -\frac{U}{1 + k_d^2/K_s^2}.$$  

(5.187)
Why is there a difference? It is because the current does not just provide a uniform translation, but, if $L_d$ is non-zero, also modifies the basic potential vorticity gradient. The basic state height field $\eta_0$ is sloping, that is $\eta_0 = -(f_0/g)\Psi y$, and the ambient potential vorticity field increases with $y$, that is $Q = (\beta + U/L_d^2)y$. Thus, the basic state defines a preferred frame of reference, and the problem is not Galilean invariant.\footnote{11}

The mechanism of Rossby waves

The fundamental mechanism underlying Rossby waves is easily understood. Consider a material line of stationary fluid parcels along a line of constant latitude, and suppose that some disturbance causes their displacement to the line marked $\eta(t = 0)$ in Fig. 5.4. In the displacement, the potential vorticity of the fluid parcels is conserved, and in the simplest case of barotropic flow on the $\beta$-plane the potential vorticity is the absolute vorticity, $\beta y + \zeta$. Thus, in either hemisphere, a northwards displacement leads to the production of negative relative vorticity and a southwards displacement leads to production of positive relative vorticity. The relative vorticity gives rise to a velocity field which in turn advects the parcels in material line in the manner shown, and the material line propagates eastward.

In more complicated situations, such as flow in two layers, considered below, or in a continuously stratified fluid, the mechanism is essentially the same: a displaced fluid parcel conserves its potential vorticity, and in the presence of a potential vorticity gradient in the basic state, the displacement leads to the production of relative vorticity and an associated velocity field. The velocity field then further displaces the fluid parcels, leading to the formation of a Rossby wave. The vital ingredient is a basic state potential vorticity gradient, such as that provided by the change of the Coriolis parameter with latitude.
5.7 Rossby Waves

5.7.2 Rossby waves in two layers

Now consider the dynamics of the two-layer model, linearized about a state of rest. The two (coupled) linear equations describing the motion in each layer are

\[
\frac{\partial}{\partial t} \left[ \nabla^2 \psi' + F_1 (\psi'_2 - \psi'_1) \right] + \beta \frac{\partial \psi'_1}{\partial x} = 0, \tag{5.188a}
\]

\[
\frac{\partial}{\partial t} \left[ \nabla^2 \psi'_2 + F_2 (\psi'_1 - \psi'_2) \right] + \beta \frac{\partial \psi'_2}{\partial x} = 0, \tag{5.188b}
\]

where \( F_1 = \frac{f_0^2}{\gamma' H_1} \) and \( F_2 = \frac{f_0^2}{\gamma' H_2} \). By inspection these may be transformed into two uncoupled equations: one equation is obtained by multiplying the first by \( F_2 \) and the second by \( F_1 \) and adding, and the other is obtained as the difference of the two equations. Then, defining

\[
\psi = F_1 \psi'_2 + F_2 \psi'_1, \quad \tau = \frac{1}{2} (\psi'_1 - \psi'_2), \tag{5.189a,b}
\]

(think ‘\( \tau \) for temperature’), (5.188) become

\[
\frac{\partial}{\partial t} \nabla^2 \overline{\psi} + \beta \frac{\partial \overline{\psi}}{\partial x} = 0, \tag{5.190a}
\]

\[
\frac{\partial}{\partial t} \left[ (\nabla^2 - k_d^2) \tau \right] + \beta \frac{\partial \tau}{\partial x} = 0. \tag{5.190b}
\]

where now \( k_d = (F_1 + F_2)^{1/2} \). The internal radius of deformation for this problem is the inverse of this, namely

\[
L_d = k_d^{-1} = \frac{1}{f_0} \left( \frac{\gamma' H_1 H_2}{H_1 + H_2} \right)^{1/2}. \tag{5.191}
\]

The variables \( \overline{\psi} \) and \( \tau \) are the normal coordinates for the two layer model; they oscillate independently of each other, and the solution in physical space is just their superposition. [For the continuous equations the analogous eigenfunctions are given by solutions of \( \partial_z [(f_0^2/N^2) \partial_z \phi] = \lambda^2 \phi \), where eigenvalue, \( \lambda \), is inversely proportional to the deformation radius.] The equation for \( \overline{\psi} \) is identical to that of the single-layer, rigid-lid model, namely (5.179) with \( U = 0 \), and its dispersion relation is just

\[
\omega = -\frac{\beta k}{K^2}. \tag{5.192}
\]

The barotropic mode corresponds to synchronous, depth-independent, motion in the two layers with a flat interface — the displacement of the interface is given by \( 2f_0 \tau / \gamma' \) and so proportional to the amplitude of the baroclinic mode. The dispersion relation for the baroclinic mode is

\[
\omega = -\frac{\beta k}{K^2 + k_d^2}. \tag{5.193}
\]

The mass transport associated with this mode is identically zero, since from (5.189) we have

\[
\psi_1 = \overline{\psi} + \frac{2F_1 \tau}{F_1 + F_2}, \quad \psi_2 = \overline{\psi} - \frac{2F_2 \tau}{F_1 + F_2}. \tag{5.194a,b}
\]
and this implies
\begin{equation}
H_1 \psi_1 + H_2 \psi_2 = (H_1 + H_2) \overline{\psi}.
\end{equation}

The left-hand side is proportional to the total mass transport, which is evidently associated with the barotropic mode.

The dispersion relation and associated group and phase velocities are plotted in Fig. 5.5. The \(x\)-component of phase speed, \(\omega/k\), is negative (westward) for both baroclinic and barotropic Rossby waves. The group velocity of the barotropic waves is always positive (eastward), but the group velocity of long baroclinic waves may be negative (westward). For very short waves, \(k^2 \gg k_d^2\), the baroclinic and barotropic velocities coincide and their phase and group velocities are equal and opposite. With a deformation radius of 50 km, typical for the mid-latitude ocean, then a nondimensional frequency of unity in the figure corresponds to a dimensional frequency of \(5 \times 10^{-7} \text{s}^{-1}\) or a period of about 100 days. In an atmosphere with a deformation radius of 1000 km a non-dimensional frequency of unity corresponds to \(1 \times 10^{-5} \text{s}^{-1}\) or a period of about 7 days. Nondimensional velocities of unity correspond to respective dimensional velocities of about \(0.25 \text{m s}^{-1}\) (ocean) and \(10 \text{m s}^{-1}\) (atmosphere).

The deformation radius only affects the baroclinic mode. For scales much smaller than the deformation radius, \(K^2 \gg k_d^2\), we see from (5.190b) that the baroclinic mode obeys the same equation as the barotropic mode so that
\begin{equation}
\frac{\partial}{\partial t} \nabla^2 \tau + \beta \frac{\partial \tau}{\partial x} = 0.
\end{equation}

Using this and (5.190a) implies that
\begin{equation}
\frac{\partial}{\partial t} \nabla^2 \psi_i + \beta \frac{\partial \psi_i}{\partial x} = 0, \quad i = 1, 2.
\end{equation}
That is to say, the two layers themselves are uncoupled from each other. At the other extreme, for very long baroclinic waves the relative vorticity is unimportant.

5.8 * Rossby Waves in Stratified Quasi-Geostrophic Flow

5.8.1 Setting up the problem

Let us now consider the dynamics of linear waves in stratified quasi-geostrophic flow on a $\beta$-plane, with a resting basic state. (In chapter 13 we explore the role of Rossby waves in a more realistic setting.) The interior flow is governed by the potential vorticity equation, (5.118), and linearizing this about a state of rest gives

$$
\frac{\partial}{\partial t} \left[ \nabla^2 \psi' + \frac{1}{\tilde{\rho}(z)} \frac{\partial}{\partial z} \left( \tilde{\rho}(z) F(z) \frac{\partial \psi'}{\partial z} \right) \right] + \beta \frac{\partial \psi'}{\partial x} = 0,
$$

(5.198)

where $\tilde{\rho}$ is the density profile of the basic state, and $F(z) = \frac{f_0^2}{N^2}$. ($F$ is the square of the inverse Prandtl ratio, $N/f_0$.) In the Boussinesq approximation $\tilde{\rho} = \rho_0$, a constant. The vertical boundary conditions are determined by the thermodynamic equation, (5.120). If the boundaries are flat, rigid, slippery surfaces then $w = 0$ at the boundaries and if there is no surface buoyancy gradient the linearized thermodynamic equation is

$$
\frac{\partial}{\partial t} \left( \frac{\partial \psi'}{\partial z} \right) = 0.
$$

(5.199)

We apply this at the ground and, with somewhat less justification, at the tropopause — the higher static stability of the stratosphere inhibits vertical motion. If the ground is not flat or if friction provides a vertical velocity by way of an Ekman layer the boundary condition must be correspondingly modified, but we will stay with the simplest case here and apply (5.199) at $z = 0$ and $z = H$.

5.8.2 Wave motion

As in the single-layer case, we seek solutions of the form

$$
\psi' = \text{Re} \tilde{\psi}(z) e^{i(kx + ly - \omega t)}
$$

(5.200)

where $\tilde{\psi}(z)$ will determine the vertical structure of the waves. The case of a sphere is more complicated but introduces no truly new physical phenomena.

Substituting (5.200) into (5.198) gives

$$
\omega \left[ -K^2 \tilde{\psi}(z) + \frac{1}{\tilde{\rho}} \frac{\partial}{\partial z} \left( \tilde{\rho} F(z) \frac{\partial \tilde{\psi}}{\partial z} \right) \right] - \beta k \tilde{\psi}(z) = 0.
$$

(5.201)

Now, if $\tilde{\psi}$ satisfies

$$
\frac{1}{\tilde{\rho}} \frac{\partial}{\partial z} \left( \tilde{\rho} F(z) \frac{\partial \tilde{\psi}}{\partial z} \right) = -\Gamma \tilde{\psi},
$$

(5.202)

where $\Gamma$ is a constant, then the equation of motion becomes

$$
- \omega \left[ K^2 + \Gamma \right] \tilde{\psi} - \beta k \tilde{\psi} = 0,
$$

(5.203)
and the dispersion relation follows, namely

\[ \omega = -\frac{\beta k}{K^2 + \Gamma}. \]  

(5.204)

Equation (5.202) constitutes an eigenvalue problem for the vertical structure; the boundary conditions, derived from (5.199), are \( \partial \tilde{\psi}/\partial z = 0 \) at \( z = 0 \) and \( z = H \). The resulting eigenvalues, \( \Gamma \) are proportional to the inverse of the squares of the deformation radii for the problem and the eigenfunctions are the vertical structure functions.

**A simple example**

Consider the case in which \( F(z) \) and \( \tilde{\rho} \) are constant, and in which the domain is confined between two rigid surfaces at \( z = 0 \) and \( z = H \). Then the eigenvalue problem for the vertical structure is

\[ F \frac{\partial^2 \tilde{\psi}}{\partial z^2} = -\Gamma \tilde{\psi} \]  

(5.205a)

with boundary conditions of

\[ \frac{\partial \tilde{\psi}}{\partial z} = 0, \quad \text{at} \quad z = 0, H. \]  

(5.205b)

There is a sequence of solutions to this, namely

\[ \tilde{\psi}_n(z) = \cos(n\pi z/H), \quad n = 1, 2 \ldots \]  

(5.206)

with corresponding eigenvalues

\[ \Gamma_n = n^2 \frac{F \pi^2}{H^2} = (n\pi)^2 \left( \frac{f_0}{NH} \right)^2, \quad n = 1, 2, \ldots \]  

(5.207)

Eq. (5.207) may be used to define the deformation radii for this problem, namely

\[ L_n \equiv \frac{1}{\sqrt{\Gamma_n}} = \frac{NH}{n\pi f_0}. \]  

(5.208)

The first deformation radius is the same as the expression obtained by dimensional analysis, namely \( NH/f \), except for a factor of \( \pi \). (Definitions of the deformation radii both with and without the factor of \( \pi \) are common in the literature, and neither is obviously more correct. In the latter case, the first deformation radius in a problem with uniform stratification is given by \( NH/f \), equal to \( \pi / \sqrt{\Gamma_1} \).) In addition to these baroclinic modes, the case with \( n = 0 \), that is with \( \tilde{\psi} = 1 \), is also a solution of (5.205) for any \( F(z) \).

Using (5.204) and (5.207) the dispersion relation becomes

\[ \omega = -\frac{\beta k}{K^2 + (n\pi)^2(f_0/NH)^2}, \quad n = 0, 1, 2 \ldots \]  

(5.209)

and, of course, the horizontal wavenumbers \( k \) and \( l \) are also quantized in a finite
domain. The dynamics of the barotropic mode are independent of height and independent of the stratification of the basic state, and so these Rossby waves are identical with the Rossby waves in a homogeneous fluid contained between two flat rigid surfaces. The structure of the baroclinic modes, which in general depends on the structure of the stratification, becomes increasingly complex as the vertical wavenumber $n$ increases. This increasing complexity naturally leads to a certain delicacy, making it rare that they can be unambiguously identified in nature. The eigenproblem for a realistic atmospheric profile is further complicated because of the lack of a rigid lid at the top of the atmosphere.\textsuperscript{12}

**APPENDIX: WAVE KINEMATICS, GROUP VELOCITY AND PHASE SPEED**

This appendix provides a brief and informal look at wave kinematics and the meaning of phase speed and group velocity.\textsuperscript{13}

5.A.1 Kinematics and definitions

A wave may be defined as a disturbance that satisfies a dispersion relation. To see what this means, suppose a disturbance, $\psi(x,t)$ (where $\psi$ might be velocity, streamfunction, pressure, etc), satisfies some equation

$$L(\psi) = 0, \quad (5.210)$$

where $L$ is a linear operator, typically a polynomial in time and space derivatives; an example is $\partial^2 \nabla \cdot / \partial t + \beta \partial \cdot / \partial x$. If $L(\psi)$ has constant coefficients (if $\beta$ is constant in our example) then solutions may often be found as a superposition of plane waves, each of which satisfy

$$\psi = \text{Re} \tilde{\psi} e^{i\theta(x,t)} = \text{Re} \tilde{\psi} e^{i(k \cdot x - \omega t)}, \quad (5.211)$$

where $\tilde{\psi}$ is a constant, $\theta$ is the phase, $k$ is the vector wavenumber $(k_x, k_y, k_z)$, and $\omega$ is the wave frequency. [We also often write the wave vector as $k = (k, l, m)$.] The frequency and wavevector are related by the dispersion relation:

$$\omega = \Omega(k), \quad (5.212)$$

where $\Omega(k)$ [meaning $\Omega(k, l, m)$] is some function determined by the form of $L$. Unless it is necessary to explicitly distinguish the function $\Omega$ from the frequency $\omega$ we will often write $\omega = \omega(k)$. Two examples of dispersion relations are (2.245) and (5.182).

If the medium in which the waves are propagating is inhomogeneous, then (5.210) will probably not have constant coefficients (for example, $\beta$ may vary meridionally). Nevertheless, if the medium is slowly varying, wave solutions may often still be found — although we do not prove it here — with the general form

$$\psi = \text{Re} \tilde{\psi}(x,t) e^{i\theta(x,t)}, \quad (5.213)$$

where $\tilde{\psi}(x,t)$ [meaning $\tilde{\psi}(x,y,z,t)$] varies slowly compared to the variation of the phase, $\theta$. The frequency and wavenumber are then defined by

$$k \equiv \nabla \theta, \quad \omega \equiv -\frac{\partial \theta}{\partial t}, \quad (5.214)$$
which in turn implies the formal relation between \( k \) and \( \omega \):

\[
\frac{\partial k}{\partial t} + \nabla \omega = 0. \tag{5.215}
\]

Even if the medium is inhomogeneous, we may still have a *local* dispersion relation between frequency and wavevector,

\[
\omega = \Omega(k, x, t). \tag{5.216}
\]

### 5.A.2 Wave propagation

#### Phase speed

First consider the propagation of monochromatic plane waves. Given (5.211) a wave will propagate in the direction of \( k \) (Fig. 5.6). At a given instant and location we can align our coordinate axis along this direction, and we may write \( k \cdot x = Kx^* \) where \( x^* \) increases in the direction of \( k \) and \( K^2 = |k|^2 \) is the magnitude of the wavenumber. With this, we can write (5.211) as

\[
\psi = \overline{\psi} e^{i(Kx^* - \omega t)} = \overline{\psi} e^{iK(x^* - ct)}. \tag{5.217}
\]

where \( c = \omega / K \). From this equation it is evident that the phase of the wave propagates at the speed \( c \) in the direction of \( k \), and we define the *phase speed*:

\[
c_p \equiv \frac{\omega}{K}. \tag{5.218}
\]

The wavelength of the wave, \( \lambda \), is the distance between two wavecrests — that is, the distance between two locations along the line of travel whose phase differs by \( 2\pi \) — and evidently this is given by

\[
\lambda = \frac{2\pi}{K}. \tag{5.219}
\]

In (for simplicity) a two-dimensional wave, and referring to Fig. 5.6a, the wavelength and wave vectors in the \( x \)- and \( y \)-directions are given by,

\[
\lambda^x = \frac{\lambda}{\cos \phi}, \quad \lambda^y = \frac{\lambda}{\sin \phi}, \quad k^x = K \cos \phi, \quad k^y = K \sin \phi. \tag{5.220}
\]

In general, lines of constant phase intersect both the coordinate axes and propagate along them. The speed of propagation along these axes is given by

\[
c_p^x = c_p \frac{l^x}{l} = \frac{c_p}{\cos \phi} = c_p \frac{K}{k^x} = \frac{\omega}{k^x}, \quad c_p^y = c_p \frac{l^y}{l} = \frac{c_p}{\sin \phi} = c_p \frac{K}{k^y} = \frac{\omega}{k^y}, \tag{5.221}
\]

using (5.220). The speed of phase propagation along any one of the axis is in general larger than the phase speed in the primary direction of the wave. The phase speeds are clearly not components of a vector: for example, \( c_p^x \neq c_p \cos \phi \).

To summarize, the phase speed and its components are given by

\[
c_p = \frac{\omega}{K}, \quad c_p^x = \frac{\omega}{k^x}, \quad c_p^y = \frac{\omega}{k^y}. \tag{5.222}
\]
Appendix: Wave Kinematics, Group Velocity and Phase Speed

Fig. 5.6 The propagation of a two-dimensional wave. (a) Two lines of constant phase (e.g., two wavecrests) at a time \(t_1\). The wave is propagating in the direction \(k\) with wavelength \(\lambda\). (b) The same line of constant phase at two successive times. The phase speed is the speed of advancement of the wavecrest in the direction of travel, and so \(c_p = \lambda/(t_2 - t_1)\). The phase speed in the \(x\)-direction is the speed of propagation of the wavecrest along the \(x\)-axis, and so \(c_{p,x} = \lambda/(t_2 - t_1) = c_p / \cos \phi\).

**Group velocity**

Let us consider how the wavevector and frequency might change with position and time. Using (5.216) in (5.215) gives

\[
\frac{D c_g}{D t} \equiv \frac{\partial}{\partial t} + c_g \cdot \nabla k = -\nabla \Omega
\]

(5.223)

where

\[
c_g = \frac{\partial \Omega}{\partial k} \equiv \left( \frac{\partial \Omega}{\partial \alpha}, \frac{\partial \Omega}{\partial \beta}, \frac{\partial \Omega}{\partial m} \right),
\]

(5.224)

is the group velocity, sometimes written as \(c_g = \nabla k \omega\) or, in subscript notation, as \(c_{g,i} = \partial \Omega / \partial k_i\). The group velocity is a vector. If the frequency is spatially constant the wavevector is evidently propagated at the group velocity.

The frequency is, in general, a function of space, wavenumber and time and from the dispersion relation, (5.216), its variation is governed by

\[
\frac{\partial \omega}{\partial t} = \frac{\partial \Omega}{\partial t} + \frac{\partial \Omega}{\partial k} \frac{\partial k}{\partial t} = \frac{\partial \Omega}{\partial t} + \frac{\partial \Omega}{\partial k} \nabla \omega
\]

(5.225)

using (5.215). Thus, using the definition of group velocity, we have

\[
\frac{D c_g \omega}{D t} = \frac{\partial \omega}{\partial t} + c_g \cdot \nabla \omega = \frac{\partial \Omega}{\partial t}.
\]

(5.226)

If the dispersion relation is not an function of time, the frequency propagates at the group velocity. We may define a ray as the trajectory traced by the group velocity, and if the frequency is not an explicit function of space or time, then both wavevector and frequency are constant along a ray.
5.A.3 Meaning of group velocity

In the discussion above we introduced the group velocity in a kinematic and rather formal way. What is the group velocity, physically?

Information and energy travel clearly cannot travel at the phase speed, for as the direction of propagation of the phase line tends to a direction parallel to the y-axis, the phase speed in the x-direction tends to infinity! Rather, they travel at the group velocity, and to see this we consider the superposition of plane waves, noting that a monochromatic plane wave already fills space uniformly. We will restrict attention to waves propagating in one direction, but the argument may be extended to two or three dimensions.

Superposition of two waves

Consider the linear superposition of two waves. Limiting attention to the one-dimensional case for simplicity, consider a disturbance represented by

$$\psi = \text{Re} \, \tilde{\psi} (e^{i(k_1 x - \omega_1 t)} + e^{i(k_2 x - \omega_2 t)}).$$

Let us further suppose that the two waves have similar wavenumbers and frequency, and in particular that $k_1 = k + \Delta k$ and $k_2 = k - \Delta k$, and $\omega_1 = \omega + \Delta \omega$ and $\omega_2 = \omega - \Delta \omega$. With this, (5.227) becomes

$$\psi = \text{Re} \, \tilde{\psi} e^{i(kx - \omega t)} \left[ e^{i(\Delta k x - \Delta \omega t)} + e^{-i(\Delta k x - \Delta \omega t)} \right] = 2 \text{Re} \, \tilde{\psi} e^{i(kx - \omega t)} \cos(\Delta k x - \Delta \omega t).$$

The resulting disturbance, illustrated in Fig. [5.7] has two aspects: a rapidly varying component, with wavenumber $k$ and frequency $\omega$, and a more slowly varying envelope, with wavenumber $\Delta k$ and frequency $\Delta \omega$. The envelope modulates the fast oscillation, and moves with the velocity $\Delta \omega / \Delta k$; in the limit $\Delta k \to 0$ and $\Delta \omega \to 0$ this is the group velocity, $c_g = \partial \omega / \partial k$. It evidently differs from the phase speed, $\omega / k$, when the latter depends on the the wavenumber. The energy in the disturbance must move at the group velocity — note that node of the envelope moves at the speed of the envelope and no energy can cross the node. These concepts generalize to more than one dimension, and if the wavenumber is the three-dimensional vector $\mathbf{k} = (k, l, m)$ then the three-dimensional envelope propagates at the group velocity given by (5.224).

* Superposition of many waves

Now consider a generalization of the above arguments. We will suppose that the disturbance is a wave packet of the form

$$\psi = A(x, t) \, e^{i(kx - \omega t)}$$

where $A(x, t)$, like the envelope in Fig. [5.7] modulates the amplitude of the wave on a scale much longer than that of the wavelength $2\pi / k$. We assume that packet is confined to a finite region of space, and that it contains a superposition of Fourier modes with wavenumbers near to $k$. That is, at some fixed time, $t = 0$ say,

$$A(x) = \int_{-\Delta k}^{+\Delta k} \tilde{A}(k') \, e^{ik'x} \, dk'$$

(5.230)
Fig. 5.7 Superposition of two sinusoidal waves with wavenumbers $k$ and $k + \delta k$, producing a wave (solid line) that is modulated by slowly varying wave envelope or wave packet (dashed line). The envelope moves at the group velocity, $c_g = \partial \omega / \partial k$ and the phase of the wave moves at the group speed $c_p = \omega / k$.

where $\Delta k$ is small. Each of the wavenumber components oscillates at a frequency given by the dispersion relation at hand, so that the evolution of the packet is given by

$$A(x, t) = \int_{-\Delta k}^{+\Delta k} \tilde{A}(k') e^{i(k'x - \omega't)} dk' \approx \int_{-\Delta k}^{+\Delta k} \tilde{A}(k') e^{ik'(x - t\partial \omega / \partial k)} dk'$$  \hspace{1cm} (5.231)

where we have used the fact that the wavenumber range is small, so that $\omega' \approx k'(\partial \omega / \partial k)$, where the derivative is evaluated at the central wavenumber $k$. Thus, comparing (5.230) and (5.231), we see that

$$A(x, t) = A(x - c_g t)$$  \hspace{1cm} (5.232)

where $c_g = \partial \omega / \partial k$. Thus, the packet moves at the group velocity.

Notes

1. The phrase ‘quasi-geostrophic’ seems to have been introduced by Durst and Sutcliffe (1938) and the concept used in Sutcliffe’s development theory of baroclinic systems (Sutcliffe 1939, 1947). The first systematic derivation of the quasi-geostrophic equations based on scaling theory was given by Charney (1948). The planetary geostrophic equations were used by Robinson and Stommel (1959) and Welander (1959) in studies of the thermocline (and were first known as the ‘thermocline equations’), and were put in the context of other approximate equation sets by Phillips (1963).

2. Carl-Gustav Rossby (1898-1957) played a dominant role in the development of dynamical meteorology in the early and middle parts of the 20th century, and his work permeates all aspects of dynamical meteorology today. Perhaps the most fundamental non-dimensional number in rotating fluid dynamics, the Rossby number, is named for him, as is the perhaps the most fundamental wave, the Rossby wave. He also discovered the conservation of potential vorticity (later generalised by Ertel) and contributed important ideas to atmospheric turbulence and the theory of air
masses. Swedish born, he studied first with V. Bjerknes before taking a position in Stockholm in 1922 with the Swedish Meteorological Hydrologic Service and receiving a 'Licentiat' from the University of Stockholm in 1925. Shortly thereafter he moved to the United States, joining the Government Weather Bureau, a precursor of NOAA’s National Weather Service. In 1928 he moved to MIT, playing an important role in developing the meteorology department there, while still maintaining connections with the Weather Bureau. In 1940 he moved to the University of Chicago, where he similarly helped develop meteorology there. In 1947 he became director of the newly-formed Institute of Meteorology in Stockholm, and subsequently divided his time between there and the United States. Thus, as well as his scientific contributions, he played a very influential role in the institutional development of the field.

3 Burger (1958)
4 This is the so-called ‘frontal geostrophic’ regime (Cushman-Roisin 1994).
5 Numerical integrations of the potential vorticity equation using (5.91), and performing the inversion without linearizing potential vorticity, do in fact indicate improved accuracy over either the quasi-geostrophic or planetary geostrophic equations (Mundt et al. 1997). In a similar vein, McIntyre and Norton (2000) show how useful potential vorticity inversion can be, and Allen et al. (1990a,b) demonstrate the high accuracy of certain intermediate models. Certainly, asymptotic correctness should not be the only criterion used in constructing a filtered model, because the parameter range in which the model is useful may be too limited. Note that there is a difference between extending the parameter range in which a filtered model is useful, as in the inversion of (5.91), and going to higher asymptotic order accuracy in a given parameter regime, as in Allen (1993) and Warn et al. (1995). Using Hamiltonian mechanics it is possible to derive equations that span different asymptotic regimes, and that also have good conservation properties (Salmon 1983, Allen et al. 2002).

6 I thank T. Warn for a conversation on this matter.
7 There is a difference between the dynamical demands of the quasi-geostrophic system in requiring $\beta$ to be small, and the geometric demands of the Cartesian geometry. On earth the two demands are similar in practice. But without dynamical inconsistency we may imagine a Cartesian system in which $\beta y \sim f$, and indeed this is common in idealized, planetary geostrophic, models of the large-scale ocean circulation.

8 Atmospheric and oceanic sciences are sometimes thought of as not being ‘beautiful’ in the same way as are some branches of theoretical physics. Yet surely quasi-geostrophic theory, and the quasi-geostrophic potential vorticity equation, are quite beautiful, both for their austerity of description and richness of behaviour.

9 Bretherton (1966). Schneider et al. (2003) look at the non QG extension. The equivalence between boundary conditions and delta-function sources is a common feature of elliptic problems, and is analogous to the generation of electromagnetic fields by point charges. It is sometimes exploited in the numerical solution of elliptic equations, both as a simple way to include non-homogeneous boundary conditions and, using the so-called capacitance matrix method, to solve problems in irregular domains (e.g., Hockney 1970, Pares-Sierra and Vallis 1989).

10 Charney and Stern (1962). See also Vallis (1996).
11 This non-Doppler effect also arises quite generally, even in models in height coordinates. See White (1977) and problem 5.9.
13 For a review of waves and group velocity, see Lighthill (1965).

Further Reading
Majda, A., 2003. *Introduction to PDEs and waves for the atmosphere and ocean.*
Provides a compact, somewhat mathematical introduction to various equation sets and their properties, including quasi-geostrophy.

Problems
5.1 In the derivation of the quasi-geostrophic equations, geostrophic balance leads to the lowest order velocity being divergence-free — that is, $\nabla \cdot u_0 = 0$. It seems that this can also be obtained from the mass conservation equation at lowest order. Is this a coincidence? Suppose that the Coriolis parameter varied, and that the momentum equation yielded $\nabla \cdot u_0 \neq 0$. Would there be an inconsistency?
5.2 ♦ In the planetary geostrophic approximation, obtain an evolution equation and corresponding inversion conditions that conserves potential vorticity and that is accurate to one higher order in Rossby number than the usual shallow water planetary geostrophic equations.
5.3 Consider the flat-bottomed shallow water potential vorticity equation in the form
\[
\frac{D}{Dt} \frac{\zeta + f}{h} = 0
\]  
(P5.1)
(a) Suppose that deviations of the height field are small compared to the mean height field, and that the Rossby number is small (so $|\zeta| \ll f$). Further consider flow on a β-plane such that $f = f_0 + \beta y$ where $|\beta y| \ll f_0$. Show that the evolution equation becomes
\[
\frac{D}{Dt} \left( \zeta + \beta y - \frac{f_0 \eta}{H} \right) = 0
\]  
(P5.2)
where $h = H + \eta$ and $|\eta| \ll H$. Using geostrophic balance in the form $f_0 u = -g \partial \eta / \partial y$, $f_0 v = g \partial \eta / \partial x$, obtain an expression for $\zeta$ in terms of $\eta$.
(b) Linearize (P5.2) about a state of rest, and show that the resulting system supports two-dimensional Rossby waves that are similar to those of the usual two-dimensional barotropic system. Discuss the limits in which the wavelength is much shorter or much longer than the deformation radius.
(c) Linearize (P5.2) about a *geostrophically balanced state* that is translating uniformly eastwards. Note that this means that:
\[
u = U + \nu' \quad \eta = \eta(y) + \eta'
\]  
(P5.3)
where $\eta(y)$ is in geostrophic balance with $U$. Obtain an expression for the form of $\eta(y)$.
(d) Obtain the dispersion relation for Rossby waves in this system. Show that their speed is different from that obtained by adding a constant $U$ to the speed of Rossby waves in part (c), and discuss why this should be so. (That is, why is the problem not *Galilean invariant*?)
5.4 Obtain solutions to the two-layer Rossby wave problem by seeking solutions of the form
\[
\psi_1 = \text{Re} \tilde{\psi}_1 e^{i(k_x x + k_y y - \omega t)}, \\
\psi_2 = \text{Re} \tilde{\psi}_2 e^{i(k_x x + k_y y - \omega t)}
\]  
(P5.4)
Substitute (P5.4) directly into (5.188) to obtain the dispersion relation, and show that the ensuing two roots correspond to the baroclinic and barotropic modes. Show that the baroclinic mode has no net (vertically integrated) transport associated with it, and that the motion of the barotropic is independent of depth.

5.5 ♦ (Not difficult, but messy.) Obtain the vertical normal modes and the dispersion relationship of the two-layer quasi-geostrophic problem with a free surface, for which the equations of motion linearized about a state of rest are

\[
\frac{\partial}{\partial t} \left[ \nabla^2 \psi_1 + F_1 (\psi_2 - \psi_1) \right] + \beta \frac{\partial \psi_1}{\partial x} = 0 \quad \text{(P5.5a)}
\]

\[
\frac{\partial}{\partial t} \left[ \nabla^2 \psi_2 + F_2 (\psi_1 - \psi_2) - F_\text{ext} \psi_1 \right] + \beta \frac{\partial \psi_2}{\partial x} = 0. \quad \text{(P5.5b)}
\]

where \( F_\text{ext} = \frac{f_0}{gH_2^2} \).

5.6 Given the baroclinic dispersion relation

\[
\omega = -\frac{\beta k_x}{k_x^2 + k_d^2}, \quad \text{(P5.6)}
\]

for what value of \( k_x \) is the \( x \)-component of group velocity the largest (i.e., the most positive), and what is the corresponding value of the group velocity?

5.7 Beginning with the vorticity and thermodynamic equations for a two layer model, obtain an expression for the conversion between available potential energy and kinetic energy in the two-layer model. Show that these expressions are consistent with the conservation of total energy as expressed by (5.168). Show also that the expression might be considered to be a simple finite-difference approximation to (5.165).

5.8 ♦ The vertical normal modes are the eigenfunctions of

\[
\frac{1}{\tilde{\rho}} \frac{\partial}{\partial z} \left( \frac{\tilde{\rho}}{N^2} \frac{\partial \Psi}{\partial z} \right) = -\Gamma \Psi \quad \text{(P5.7)}
\]

along with boundary conditions on \( \Psi(z) \). Numerically obtain the vertical normal modes for some or all of the following profiles, or others of your choice.

(a) \( \tilde{\rho} = 1, N^2(z) = 1, \Psi_z = 0 \) at \( z = 0, 1 \). (This is profile ‘b’ of Fig. 9.10. An analytic solution is possible.)

(b) \( \tilde{\rho} = 1 \) and an \( N^2(z) \) profile corresponding to a density profile similar to profile ‘a’ of Fig. 9.10 (e.g., an exponential), and \( \Psi_z = 0 \) at \( z = 0, 1 \).

(c) An isothermal atmosphere, with \( \Psi_z = 0 \) at \( z = 0, 1 \). (Similar to (i), except that \( \tilde{\rho} \) varies with height.)

(d) An isothermal atmosphere, now assuming that \( \psi \to 0 \) as \( z \to \infty \).

(e) An fluid with \( N^2 = 1 \) for \( 0 < z < 0.5 \) and \( N^2 = 4 \) for \( 0.5 < z < 1 \), with continuous \( b \), and with \( \Psi_z = 0 \) at \( z = 0, 1 \).

In the atmospheric cases it is easiest to do the problem first with \( \tilde{\rho} = 1 \) (the Boussinesq case) and then extend the problem (and the code) to the compressible case. Then remove the upper boundary to larger and larger values of \( z \).

5.9 ♦ Show that the non-Doppler effect arises using geometric height as the vertical coordinates, using the modified quasi-geostrophic set of [White, 1977]. In particular, obtain the dispersion relation for stratified quasi-geostrophic flow with a resting basic state. Then obtain the dispersion relation for the equations linearized about a uniformly translating state, paying attention to the lower boundary condition, and note the conditions under which the waves are stationary. Discuss.
5.10 Derive the quasi-geostrophic potential vorticity equation in isopycnal coordinates for a Boussinesq fluid. Show that the isopycnal expression for potential vorticity is approximately equal to the corresponding expression in height coordinates, carefully stating any assumptions that may be necessary to show this.

5.11 (a) Obtain the dispersion relationship for free Rossby waves for the single-layer quasi-geostrophic potential vorticity equation with linear drag.

(b) Obtain the dispersion relation for free Rossby waves in the linearized two-layer potential vorticity equation with linear drag in the lowest layer.

(c) ♦ Obtain the dispersion relation for free waves in the continuously stratified quasi-geostrophic equations, with the effects of linear drag appearing in the thermodynamic equation for the lower boundary condition. That is, the boundary condition at $z = 0$ is $\partial_t (\partial_z \psi) + N^2 w = 0$ where $w = \alpha \zeta$ where $\alpha$ is a constant. You may make the Boussinesq approximation and assume $N^2$ is constant if you like.
Part II

INSTABILITY, WAVE–MEAN FLOW INTERACTION AND TURBULENCE
WHAT HYDRODYNAMIC STATES ACTUALLY OCCUR IN NATURE? Any flow must clearly be a solution of the equations of motion, and there are, in fact, many steady solutions to the equations of motion — a purely zonal flow, for example. However, steady solutions do not abound in nature because, in order to persist, they must be stable to those small perturbations that inevitably arise. Indeed, all the steady solutions that are known for the large-scale flow in the earth’s atmosphere and ocean have been found to be unstable.

There are a myriad forms of hydrodynamic instability, but our focus in this chapter is on barotropic and baroclinic instability. The latter is at the heart of the large- and mesoscale motion in the atmosphere and ocean — it gives rise to atmospheric weather systems, for example. Barotropic instability is important to us for two reasons. First, it is important in its own right as an instability mechanism for jets and vortices, and is a driving mechanism in both two- and three-dimensional turbulence. Second, many problems in barotropic and baroclinic instability are formally and dynamically similar, so that the solutions and insight we obtain in the often simpler problems in barotropic instability are often useful in the baroclinic problem.

6.1 KELVIN-HELMHOLTZ INSTABILITY
To introduce the issue, we will first consider, rather informally, perhaps the simplest physically interesting instance of a fluid-dynamical instability — that of a constant-density flow with a shear perpendicular to the fluid’s mean velocity, this being an example of a Kelvin-Helmholtz instability.¹ Let us consider two fluid masses of equal density, with a common interface at \( y = 0 \), moving with velocities \(-U\) and \(+U\) in the \( x\)-direction respectively (Fig. 6.1). There is no variation in the basic flow in the \( z\)-direction (normal to the page), and we will assume this is also true for
Chapter 6. Barotropic and Baroclinic Instability

Figure 6.1 A simple basic state giving rise to shear-flow instability. The velocity profile is discontinuous, the density uniform.

The instability (these restrictions are not essential). This flow is clearly a solution of the Euler equations. What happens if the flow is perturbed slightly? If the perturbation is initially small then even if it grows we can, for small times after the onset of instability, neglect the nonlinear interactions in the governing equations because these are the squares of small quantities. The equations determining the evolution of the initial perturbation are then the Euler equations linearized about the steady solution. Thus, denoting perturbation quantities with a prime and basic state variables with capital letters, for \( y > 0 \) the perturbation satisfies

\[
\frac{\partial u'}{\partial t} + U \frac{\partial u'}{\partial x} = -\nabla p', \quad \nabla \cdot u' = 0 \tag{6.1a,b}
\]

and a similar equation hold for \( y < 0 \), but with \( U \) replaced by \(-U\). Given periodic boundary conditions in the \( x \)-direction, we may seek solutions of the form

\[
\phi'(x, y, t) = \text{Re} \sum_k \tilde{\phi}_k(y) \exp[i k (x - ct)], \tag{6.2}
\]

where \( \phi \) is any field variable (e.g., pressure or velocity), and \( \text{Re} \) denotes that only the real part should be taken. (Typically we use tildes over variables to denote Fourier-like modes, and we will often omit the marker ‘Re’.) Because (6.1a) is linear, the Fourier modes do not interact and we may confine attention to just one. Taking the divergence of (6.1a), the left-hand side vanishes and the pressure satisfies Laplace’s equation

\[
\nabla^2 p' = 0 \tag{6.3}
\]

This has solutions in the form

\[
p' = \begin{cases} 
\text{Re} \tilde{\rho}_1 e^{ikx-ky} e^{\sigma t} & y > 0, \\
\text{Re} \tilde{\rho}_2 e^{ikx+ky} e^{\sigma t} & y < 0, 
\end{cases} \tag{6.4}
\]

where, anticipating the possibility of growing solutions, we write \( \sigma = -i k c \). In general the growth-rate \( \sigma \) is complex: if it has a positive real component, the amplitude
of the perturbation will grow and there is an instability; if $\sigma$ has a non-zero imaginary component, then there will be oscillatory motion, and there may of course be both oscillatory motion and an instability. To obtain the dispersion relationship, we consider the $y$-component of (6.1a), namely (for $y > 0$)

$$\frac{\partial v'_1}{\partial t} + U \frac{\partial v'_1}{\partial x} = - \frac{\partial p'_1}{\partial y} \quad (6.5)$$

Substituting a solution of the form $v'_1 = \tilde{v}_1 \exp(ikx + \sigma t)$ yields, with (6.4),

$$(\sigma + ikU) \tilde{v}_1 = k \tilde{p}_1. \quad (6.6)$$

But the velocity normal to the interface is, at the interface, nothing but the rate of change of the position of interface itself. That is, at $y = +0$

$$v_1 = \frac{\partial \eta'}{\partial t} + U \frac{\partial \eta'}{\partial x}, \quad (6.7)$$

or

$$\tilde{v}_1 = (\sigma + ikU) \tilde{\eta}. \quad (6.8)$$

where $\eta'$ is the displacement of the interface from its equilibrium position. Using this in (6.6) gives

$$(\sigma + ikU)^2 \tilde{\eta} = k \tilde{p}_1. \quad (6.9)$$

The above few equations pertain to motion on the $y > 0$ side of the interface. Similar reasoning on the other side gives (at $y = -0$)

$$(\sigma - ikU)^2 \tilde{\eta} = -k \tilde{p}_2. \quad (6.10)$$

But at the interface $p_1 = p_2$ (because pressure must be continuous). The dispersion relationship then emerges from (6.9) and (6.10), giving

$$\sigma^2 = k^2 U^2. \quad (6.11)$$

This equation has two roots, one of which is positive. Thus, the amplitude of the perturbation grows exponentially, like $e^{\sigma t}$, and the flow is unstable. The instability itself can be seen in the natural world when billow clouds appear wrapped up into spirals: the clouds are acting as tracers of fluid flow, and the billows are a manifestation of the instability at finite amplitude, as in Fig. 6.6.

### 6.2 Instability of Parallel Shear Flow

We now consider a little more systematically the instability of parallel shear flows, such as are illustrated in Fig. 6.2. This is a classic problem in hydrodynamic stability theory, and there are two particular reasons for our own interest:

(i) The instability is an example of barotropic instability, which abounds in the ocean and atmosphere. Loosely, barotropic instability arises when a flow is unstable by virtue of its shear, with gravitational and buoyancy effects being secondary.
(ii) The instability is in many ways analagous to baroclinic instability, which is the main instability giving rise to weather systems in the atmosphere and similar phenomena in the ocean.

Let us will restrict attention to two dimensional, incompressible flow; this illustrates the physical mechanisms in the most transparent way, in part because it allows for the introduction of a streamfunction and the automatic satisfaction of the mass continuity equation. In fact, for parallel two-dimensional shear flows the most unstable disturbances are two-dimensional ones.

The vorticity equation for incompressible two-dimensional flow is just

$$\frac{D\zeta}{Dt} = 0.$$  \hspace{1cm} (6.12)

We suppose the basic state to be a parallel flow in the \(x\)-direction that may vary in \(y\)-direction. That is

$$\mathbf{u} = U(y) \mathbf{i}.$$  \hspace{1cm} (6.13)

The linearized vorticity equation is then

$$\frac{\partial \zeta'}{\partial t} + U \frac{\partial \zeta'}{\partial x} + v' \frac{\partial Z}{\partial y} = 0$$ \hspace{1cm} (6.14)

where \(Z = -U_y\). Because the mass continuity equation has the simple form

$$\frac{\partial u'}{\partial x} + \frac{\partial v'}{\partial y} = 0,$$  \hspace{1cm} (6.15)

we may introduce a streamfunction \(\psi\) such that \(u' = -\partial \psi'/\partial y, v' = \partial \psi'/\partial x\) and \(\zeta' = \nabla^2 \psi'\). The linear vorticity equation is then

$$\frac{\partial \nabla^2 \psi'}{\partial t} + U \frac{\partial \nabla^2 \psi'}{\partial x} + \frac{\partial Z}{\partial y} \frac{\partial \psi'}{\partial x} = 0.$$ \hspace{1cm} (6.16)

The coefficients of the \(x\)-derivatives are not themselves functions of \(x\). Thus we may seek solutions that are harmonic functions (sines and cosines) in the \(x\)-direction, but the \(y\)-dependence must remain arbitrary at this point and we write

$$\psi' = \text{Re} \tilde{\psi}(y) e^{i(k(x-ct))}.$$ \hspace{1cm} (6.17)

The solution is a superposition of all wavenumbers, but since the problem is linear the waves do not interact and it suffices to consider them separately. If \(c\) is purely real then \(c\) is the phase speed of the wave; if \(c\) has a positive imaginary component then the wave will grow exponentially is thus unstable.

From (6.17) we have

$$u' = \tilde{u}(y) e^{i(k(x-ct))} = -\tilde{\psi}_y e^{i(k(x-ct))},$$ \hspace{1cm} (6.18a)

$$v' = \tilde{v}(y) e^{i(k(x-ct))} = ik \tilde{\psi} e^{i(k(x-ct))},$$ \hspace{1cm} (6.18b)

$$\zeta' = \tilde{\zeta}(y) e^{i(k(x-ct))} = (-k^2 \tilde{\psi} + \tilde{\psi}_{yy}) e^{i(k(x-ct))},$$ \hspace{1cm} (6.18c)

where the \(y\) subscript denotes a derivative. Using (6.18) in (6.14) gives

$$(U - c) (\tilde{\psi}_{yy} - k^2 \tilde{\psi}) - U_{yy} \tilde{\psi} = 0,$$ \hspace{1cm} (6.19)
sometimes known as Rayleigh’s equation. It is the linear vorticity equation for disturbances to parallel shear flow, and in the presence of a \( \beta \)-effect it generalizes slightly to
\[
(U - c)(\tilde{\psi}_{yy} - k^2 \tilde{\psi}) + (\beta - U_{yy}) \tilde{\psi} = 0.
\]
(6.20)

6.2.1 Piecewise linear flows

Although Rayleigh’s equation is linear and has a simple form, it is nevertheless quite difficult to analytically solve for an arbitrary smoothly varying profile. It is simpler to consider piecewise linear flows, in which \( U_y \) is a constant over some interval, with \( U \) or \( U_y \) changing abruptly to another value at a line of discontinuity, as illustrated in Fig. 6.2. The curvature, \( U_{yy} \), is accounted for through the satisfaction of matching conditions, analogous to boundary conditions, at the lines of discontinuity (as in section 6.1), and solutions in each interval are then exponential functions.

Jump or Matching conditions

The idea, then, is to solve the linearized vorticity equation separately in the continuous intervals in which vorticity is constant, matching the solution with that in the adjacent regions. The matching conditions arise from two physical conditions:

(i) That normal stress should be continuous across the interface. For an inviscid fluid this implies that pressure be continuous.

(ii) That the normal velocity of the fluid on either side of the interface should be consistent with the motion of the interface itself.

Let us consider the implications of these two conditions.

(i) Continuity of pressure:

The linearized momentum equation in the direction along the interface is:
\[
\frac{\partial u'}{\partial t} + U \frac{\partial u'}{\partial x} + v' \frac{\partial U}{\partial y} = - \frac{\partial p'}{\partial x}.
\]
(6.21)
For normal modes, \( u' = -\tilde{\psi}_y e^{ik(x-ct)} \) and \( v' = ik\tilde{\psi} e^{ik(x-ct)} \) and (6.21) becomes
\[
  ik(U - c)\tilde{\psi}_y - ik\tilde{\psi}U_y = -ik\tilde{p}. \tag{6.22}
\]

Because pressure is continuous across the interface we have the first matching or jump condition,
\[
  \Delta[(U - c)\tilde{\psi}_y - \tilde{\psi}U_y] = 0 \tag{6.23}
\]
where the operator \( \Delta \) denotes the difference in the values of the argument (in square brackets) across the interface. That is, the quantity \((U - c)\tilde{\psi}_y - \tilde{\psi}U_y\) is continuous.

We can obtain this condition directly from Rayleigh’s equation, (6.20), written in the form
\[
  [(U - c)\tilde{\psi}_y - U_y\tilde{\psi}]_y + [\beta - k^2(U - c)]\tilde{\psi} = 0. \tag{6.24}
\]
Integrating across the interface gives (6.23).

(ii) **Material interface condition:**
At the interface, the normal velocity \( v \) is given by the kinematic condition
\[
  v = \frac{D\eta}{Dt} \tag{6.25}
\]
where \( \eta \) is the interface displacement. The linear version of (6.25) is
\[
  \frac{\partial \eta'}{\partial t} + U \frac{\partial \eta'}{\partial x} = \frac{\partial \psi'}{\partial x}. \tag{6.26}
\]
If the fluid itself is continuous then this equation must hold at either side of the interface, giving two equations and their normal mode counterparts, namely,
\[
  \frac{\partial \eta'}{\partial t} + U_1 \frac{\partial \eta'}{\partial x} = \frac{\partial \psi'_1}{\partial x} \quad \rightarrow \quad (U_1 - c)\tilde{\eta} = \tilde{\psi}_1, \tag{6.27}
\]
\[
  \frac{\partial \eta'}{\partial t} + U_2 \frac{\partial \eta'}{\partial x} = \frac{\partial \psi'_2}{\partial x} \quad \rightarrow \quad (U_2 - c)\tilde{\eta} = \tilde{\psi}_2. \tag{6.28}
\]
Material continuity at the interface thus gives the second jump condition:
\[
  \Delta \left[ \frac{\tilde{\psi}}{U - c} \right] = 0. \tag{6.29}
\]
That is, \( \tilde{\psi}/(U - c) \) is continuous at the interface. Note that if \( U \) is continuous across the interface the condition becomes one of continuity of the normal velocity.
6.2 Instability of Parallel Shear Flow

6.2.2 Kevin-Helmholtz instability, revisited

We now use Rayleigh’s equation and the jump conditions to consider the situation illustrated in Fig. 6.1, that is, vorticity is everywhere zero except in a thin sheet at \( y = 0 \). On either side of the interface, Rayleigh’s equation is simply

\[
(U - c)(\partial_{yy} \tilde{\psi}_i - k^2 \tilde{\psi}_i) = 0 \quad i = 1, 2
\]

(6.30)

or, assuming that \( U \neq c \),

\[
\tilde{\psi}_{yy} - k^2 \tilde{\psi} = 0.
\]

This is just Laplace’s equation, coming from \( \nabla^2 \psi' = \zeta' \), with \( \zeta' = 0 \) everywhere except at the interface. Solutions of this that decay away on either side of the interface are

\[
\begin{align*}
  y > 0 : & \quad \tilde{\psi}_1 = \Psi_1 e^{-ky}, \\
  y < 0 : & \quad \tilde{\psi}_2 = \Psi_2 e^{ky},
\end{align*}
\]

(6.31a)

(6.31b)

where \( \Psi_1 \) and \( \Psi_2 \) are constants. The boundary condition (6.23) gives

\[
(U_1 - c)(-k)\Psi_1 = (U_2 - c)(k)\Psi_2,
\]

(6.32)

and (6.29) gives

\[
\frac{\Psi_1}{(U_1 - c)} = \frac{\Psi_2}{(U_2 - c)}.
\]

(6.33)

The last two equations combine to give \( (U_1 - c)^2 = -(U_2 - c)^2 \), which, supposing that \( U = U_1 = -U_2 \) gives \( c^2 = -U^2 \). Thus, since \( U \) is purely real, \( c = \pm iU \), and the disturbance grows exponentially as \( \exp(kU_1 t) \), just as we obtained in section 6.1. All wavelengths are unstable, and indeed the shorter the wavelength the greater the instability. In reality, viscosity will damp the smallest waves, but at the same time the presence of viscosity would also mean that the initial profile is not an exact, steady solution of the equations of motion.

6.2.3 Edge Waves

We now consider a case sketched in Fig. 6.3 in which the velocity is continuous, but the vorticity is discontinuous. Since on either side of the interface \( U_{yy} = 0 \),...
Rayleigh’s equation is just
\[(U - c)(\tilde{\psi}_{yy} - k^2 \tilde{\psi}) = 0.\]
\[(6.34)\]
Provided \(c \neq U\) this has solutions,
\[\tilde{\psi} = \begin{cases} 
\Phi_1 e^{-ky} & y > 0 \\
\Phi_2 e^{ky} & y < 0.
\end{cases}\]
\[(6.35)\]
The value of \(c\) is found by applying the jump conditions \((6.23)\) and \((6.29)\) at \(y = 0\). Using \((6.35)\) these give
\[-k(U_0 - c)\Phi_1 - \Phi_1 U_{1y} = k(U_0 - c)\Phi_2 - \Phi_2 U_{2y}\]
\[(6.36a)\]
\[\Phi_1 = \Phi_2\]
\[(6.36b)\]
where \(U_1\) and \(U_2\) are the values of \(U\) at either side of the interface, and both are equal to \(U_0\) at the interface. After a line of algebra these equations give
\[c = U_0 + \frac{\partial_y U_1 - \partial_y U_2}{2k}.\]
\[(6.37)\]
This is the dispersion relationship for edge waves that propagate along the interface a speed equal to the sum of the fluid speed and a factor proportional to the difference in the vorticity between the two layers. No matter what the shear is on either side of the interface, the phase speed is purely real and there is no instability. Eq. \((6.37)\) is imperfectly analogous to the Rossby wave dispersion relation \(c = U_0 - \beta/K^2\), and reflects a similarity in the physics — \(\beta\) is a planetary vorticity gradient, which in \((6.37)\) is collapsed to a front and represented by the difference \(U_{1y} - U_{2y} = -(Z_1 - Z_2)\), where \(Z_1\) and \(Z_2\) are the basic-state vorticities on either side of the interface.

### 6.2.4 Interacting edge waves producing instability

Now we consider a slightly more complicated case in which edge waves may interact giving rise, as we shall see, to an instability. The physical situation is illustrated in Fig. \ref{fig:6.4}. We consider the simplest case, that of a shear layer (which we denote region 2) sandwiched between two semi-infinite layers, regions 1 and 3, as in the left panel of the figure. Thus, the basic state is:
\[y > a : \quad U = U_1 = U_0 \quad \text{(a constant)},\]
\[(6.38a)\]
\[-a < y < a : \quad U = U_2 = \frac{U_0}{a} y,\]
\[(6.38b)\]
\[y < -a : \quad U = U_3 = -U_0.\]
\[(6.38c)\]
We assume a solution of Rayleigh’s equation of the form:
\[y > a : \quad \tilde{\psi}_1 = A e^{-k(y-a)},\]
\[(6.39a)\]
\[-a < y < a : \quad \tilde{\psi}_2 = B e^{-k(y-a)} + C e^{k(y+a)},\]
\[(6.39b)\]
\[y < -a : \quad \tilde{\psi}_3 = D e^{k(y+a)}.\]
\[(6.39c)\]
6.2 Instability of Parallel Shear Flow

Fig. 6.4 Barotropically unstable velocity profiles. In the simplest case, on the left, a region of shear is sandwiched between two infinite regions of constant velocity. The edge waves at \( y = \pm a \) interact to produce an instability. If \( a = 0 \), then the situation corresponds to that of Fig. 6.1, giving Kelvin-Helmholtz instability. In the case on the right, the flow is bounded at \( y = \pm b \). It may be shown that the flow is still unstable, provided that \( b \) is sufficiently larger than \( a \). If \( b = a \) (plane Couette flow) the flow is stable to infinitesimal disturbances.

Applying the jump conditions (6.23) and (6.29) at the interfaces at \( y = a \) and \( y = -a \) gives the following relations between the coefficients:

\[
A[(U_0 - c)k] = B \left[(U_0 - c)k + \frac{U_0}{a}\right] + C e^{2ka} \left[\frac{U_0}{a} - (U_0 - c)k\right], \tag{6.40a}
\]

\[A = B + C e^{2ka}, \tag{6.40b}\]

\[D[(U_0 + c)k] = B e^{2ka} \left[-(U_0 + c)k + \frac{U_0}{a}\right] + C \left[\frac{U_0}{a} + (U_0 + c)k\right], \tag{6.40c}\]

\[D = B e^{2ka} + C. \tag{6.40d}\]

These are a set of four homogeneous equations, with the unknown parameters \( A, B, C \) and \( D \), which may be written in the form of a matrix equation,

\[
\begin{pmatrix}
  k(U_0 - c) & -k(U_0 - c) - U_0/a & e^{2ka} & 0 \\
  1 & -1 & -e^{2ka} & 0 \\
  0 & e^{2ka}[k(U_0 + c) - (U_0/a)] & -k(U_0 + c) - (U_0/a) & 0 \\
  0 & e^{2ka} & 1 & -1
\end{pmatrix}
\begin{pmatrix}
  A \\
  B \\
  C \\
  D
\end{pmatrix} = 0. \tag{6.41}

For non-trivial solutions the determinant of the matrix must be zero, and solving the ensuing equation gives the dispersion relationship

\[c^2 = \left(\frac{U_0}{2ka}\right)^2 \left[(1 - 2ka)^2 - e^{-4ka}\right], \tag{6.42}\]

and this is plotted in Fig. 6.5. The flow is unstable for sufficiently long wavelengths,
for then the right-hand side of (6.42) is negative. The critical wavenumber below which instability occurs is found by solving $(1 - 2ka)^2 = e^{-4ka}$, which gives instability for $ka < 0.63293$. A numerical solution of the initial value problem is illustrated in Fig. 6.6 and Fig. 6.7. Here, the initial perturbation is small and random, containing components at all wavenumbers. All the modes in the unstable range grow exponentially, and the pattern is soon dominated by the mode that grows fastest — a horizontal wavenumber three in this problem. Eventually, the perturbation grows sufficiently that the linear equations are no longer valid and, as is seen in the second column of Fig. 6.6, vortices form and pinch off. Eventually, the vortices interact and the flow develops into two-dimensional turbulence, considered in chapter 8.

The mechanism of the instability — an informal view

[A similar mechanism is discussed in section 6.7, and the reader may wish to read the two descriptions in tandem.] We have seen that an edge wave in isolation is stable, the instability arising when two edge waves have sufficient cross-stream extent that they can interact with each other. This occurs for sufficiently long wavelengths because the cross-stream decay scale is proportional to the along-stream wavelength — hence the high-wavenumber cut-off. To transparently see the mechanism of the instability, let us first suppose that the interfaces are, in fact, sufficiently far away that the edge waves at each interface do not interact. Using (6.37) the edge waves at $y = -a$ and $y = +a$ have dispersion relationships

$$c_{+a} = U_0 - \frac{U_0/a}{2k}, \quad c_{-a} = -U_0 + \frac{U_0/a}{2k}$$

(6.43a,b)

If the two waves are to interact these phase speeds must be equal, giving the condition

$$c = 0, \quad k = 1/(2a).$$

(6.44a,b)

That is, the waves are stationary, and their wavelength is proportional to the separation of the two edges. In fact, (6.44) approximately characterizes the conditions.
Fig. 6.6 A sequence of plots of the vorticity, at equal time intervals, from a numerical solution of the nonlinear vorticity equation (6.12), with initial conditions as in Fig. 6.4 with $a = 0.1$, plus a very small random perturbation. Time increases first down the left column, then down the right column. The solution is obtained in a rectangular $(4 \times 1)$ domain, with periodic conditions in the $x$-direction and slippery walls at $y = (0, 1)$. The maximum linear instability occurs for a wavelength of 1.57, which for a domain of length 4 corresponds to a wavenumber of 2.55. Since the periodic domain quantizes the allowable wavenumbers, the maximum instability is at wavenumber 3, and this is what emerges. Only in the first two or three frames is the linear approximation valid.

Fig. 6.7 Total streamfunction (top panel) and perturbation streamfunction from the same numerical calculation as in Fig. 6.6 at a time corresponding to the second frame. Positive values (a clockwise circulation) are solid lines, and negative values are dashed. The perturbation pattern leans into the shear, and grows exponentially in place.
at the critical wavenumber $k = 0.63/a$ (see Fig. 5.3). In the region of the shear the two waves have the form

$$
\psi_{+a} = \text{Re} \tilde{\psi}_{+a}(t) e^{i(y-a)} e^{i\phi} e^{i k x}, \quad \psi_{-a} = \text{Re} \tilde{\psi}_{-a}(t) e^{-i(y+a)} e^{i k x} \quad (6.45a, b)
$$

where $\phi$ is the phase shift between the waves; in the case of pure edge waves we have $\tilde{\psi}_{\pm a} = A_{\pm a} e^{-i k c t}$ where we may take $A_{\pm a}$ to be real.

Now consider how the wave generated at $y = -a$ might affect the wave at $y = +a$ and vice versa. The contribution of $\psi_{-a}$ to acceleration of $\psi_{+a}$ is given by applying the $x$-momentum equation, (6.21), at either side of the interface at $y = \pm a$, and similarly for the acceleration at $y = -a$. Thus we take the kinematic solutions, (6.45), and use them in a dynamical equation, the momentum equation, to calculate the ensuing acceleration. We obtain

$$
\frac{\partial \tilde{u}_{+a}}{\partial t} \approx -v_{-a}(+a) \frac{\partial U}{\partial y}, \quad \frac{\partial \tilde{u}_{-a}}{\partial t} \approx -v_{-a}(-a) \frac{\partial U}{\partial y} \quad (6.46a, b)
$$

at $y = +a$ and $y = -a$ respectively, omitting the terms that give the neutral edge waves. Here $v_{-a}(+a)$ denotes the value of $v$ at $y = +a$ due to the edge wave generated at $-a$. If the spatial dependence of the waves is given by (6.45) this gives, at $y = \pm a$,

$$
-k e^{i \phi} \frac{\partial \tilde{\psi}_{+a}}{\partial t} \approx -i k \tilde{\psi}_{-a} \frac{\partial U}{\partial y}, \quad k \frac{\partial \tilde{\psi}_{-a}}{\partial t} \approx -i k e^{i \phi} \tilde{\psi}_{+a} \frac{\partial U}{\partial y} \quad (6.47a, b)
$$

If $\psi_{+a}$ and $\psi_{-a}$ have the appropriate phase with respect to each other, then the two edge waves can feed back on each other. In particular, from (6.47) we see that the system is unstable when $\phi = \pi/2$, for then (6.47) gives

$$
\frac{\partial^2 \tilde{\psi}_{+a}}{\partial t^2} \approx \tilde{\psi}_{+a} \left( \frac{\partial U}{\partial y} \right)^2 \quad (6.48)
$$

In this case, the wave at $y = +a$ lags the wave at $y = -a$. That is, the perturbation is unstable when it tilts into the shear, and this is seen in the full solution, Fig. 6.7.

### 6.3 NECESSARY CONDITIONS FOR INSTABILITY

#### 6.3.1 Rayleigh’s criterion

For simple profiles it may be possible to calculate, or even intuit, the instability properties, but for continuous profiles of $U(y)$ this is often impossible and it would be nice to have some general guidelines as to when a profile might be unstable. To this end, we will derive a couple of necessary conditions for instability, or sufficient conditions for stability, that will at least tell us if a flow might be unstable.

We first write Rayleigh’s equation, (6.20), as

$$
\tilde{\psi}_{yy} - k^2 \tilde{\psi} + \frac{B - U_{yy}}{U - c} \tilde{\psi} = 0 \quad (6.49)
$$

Multiply by $\tilde{\psi}^*$ (the complex conjugate of $\tilde{\psi}$) and integrate over the domain of interest. After integrating the first term by parts, this gives

$$
\int_{y_1}^{y_2} \left[ \left| \frac{\partial \tilde{\psi}}{\partial y} \right|^2 + k^2 |\tilde{\psi}|^2 \right] dy - \int_{y_1}^{y_2} \frac{B - U_{yy}}{U - c} |\tilde{\psi}|^2 dy = 0 \quad (6.50)
$$
assuming that $\tilde{\psi}$ vanishes at the boundaries. (The limits to the integral may be infinite, in which case it is assumed that $\tilde{\psi}$ decays to zero as $|y|$ approaches $\infty$.) The only variable in this expression that is complex is $c$, and thus the first integral is real. The imaginary component of the second integral is

$$c_i \int \frac{\beta - U_{yy}}{|U - c|^2} |\tilde{\psi}|^2 \, dy = 0. \quad (6.51)$$

Thus, either $c_i$ vanishes or the integral does. For there to be an instability, $c_i$ must be nonzero and because the eigenvalues of Rayleigh’s equation come in pairs (because it is a second order ODE), and for each decaying mode (negative $c_i$) there is a corresponding growing mode (positive $c_i$). Therefore:

A necessary condition for instability is that the expression $\beta - U_{yy}$ change sign somewhere in the domain.

Equivalently, a sufficient criterion for stability is that $\beta - U_{yy}$ not vanish in the domain interior. This condition is known as Rayleigh’s inflection-point criterion, or when $\beta \neq 0$, the Rayleigh-Kuo inflection point criterion.\(^7\)

An alternate, more general, derivation

Consider again the vorticity equation, linearized about a parallel shear flow [c.f., (6.14) with a $\beta$ term],

$$\frac{\partial \zeta}{\partial t} + U \frac{\partial \zeta}{\partial x} + v \left( \frac{\partial Z}{\partial y} + \beta \right) = 0, \quad (6.52)$$

(dropping the primes on the perturbation quantities). Multiply by $\zeta$ and divide by $\beta + Z_y$ to obtain

$$\frac{\partial}{\partial t} \left( \frac{\zeta^2}{\beta + Z_y} \right) + \frac{U}{\beta + Z_y} \frac{\partial \zeta^2}{\partial x} + v \zeta = 0, \quad (6.53)$$

and then integrate with respect to $x$ to give

$$\frac{\partial}{\partial t} \int \left( \frac{\zeta^2}{\beta + Z_y} \right) \, dx = - \int v \zeta \, dx. \quad (6.54)$$

Now, using $\nabla \cdot \mathbf{u} = 0$,

$$v \zeta = - \frac{\partial}{\partial y} (uv) + \frac{1}{2} \frac{\partial}{\partial x} (u^2 + v^2). \quad (6.55)$$

That is, the flux of vorticity is the divergence of some quantity. Its integral therefore vanishes provided there are no contributions from the boundary, and integrating (6.54) with respect to $y$ gives

$$\frac{d}{dt} \int \left( \frac{\zeta^2}{\beta + Z_y} \right) \, dx \, dy = 0. \quad (6.56)$$
If there is to be an instability $\zeta$ must grow, but the integral is identically zero. These two conditions can only be simultaneously satisfied if $\beta + Z_y$, or equivalently $\beta - U_{yy}$, is zero somewhere in the domain.

This derivation shows that the inflection-point criterion applies even if disturbances are not of normal-mode form. The quantity $\zeta^2/(\beta + Z_y)$ is an example of a wave-activity density — a wave activity being a conserved quantity, quadratic in the amplitude of the wave. Such quantities play an important role in instabilities, and we consider then further in chapter 7.

### 6.3.2 Fjørtoft's criterion

Another necessary condition for instability was obtained by Fjørtoft. In this section we will derive the condition for normal mode disturbances, and provide a more general derivation in section 7.7. From the real part of (6.50) we find

$$\int_{y_1}^{y_2} (\beta - U_{yy})(U - U_s) |\tilde{\psi}|^2 \, dy = \int_{y_1}^{y_2} \frac{\partial \tilde{\psi}}{\partial y} \, \frac{\partial \tilde{\psi}}{\partial y} \, dy + k^2 |\tilde{\psi}|^2 \, dy > 0. \quad (6.57)$$

Now, from (6.51), we know that for an instability we must have

$$\int_{y_1}^{y_2} \frac{\beta - U_{yy}}{|U - c|^2} |\tilde{\psi}|^2 \, dy = 0. \quad (6.58)$$

Using this and (6.57), it is clear that, for an instability,

$$\int_{y_1}^{y_2} (\beta - U_{yy})(U - U_s) |\tilde{\psi}|^2 \, dy > 0 \quad (6.59)$$

where $U_s$ is any real constant. It is most useful to choose this constant to be the value of $U(y)$ at which $\beta - U_{yy}$ vanishes. This leads directly to the criterion:

A necessary condition for instability is that the expression

$$(\beta - U_{yy})(U - U_s)$$

where $U_s$ is the value of $U(y)$ at which $\beta - U_{yy}$ vanishes, be positive somewhere in the domain.

This is equivalent to saying that the magnitude of the vorticity must have an extremum inside the domain, and not at the boundary or at infinity, as can be seen by perusing Fig. 6.8. Why choose $U_s$ in the manner we did? Suppose we chose $U_s$ to have a very large negative or large positive value, so that $U - U_s$ is of one sign everywhere. Then (6.59) just implies that $\beta - U_{yy}$ must be negative somewhere and must be positive somewhere, which is already known from Rayleigh’s criterion. The most stringent criterion is obtained by choosing $U_s$ to be the value of $U(y)$ at which $\beta - U_{yy}$ vanishes. Both Fjørtoft’s and Rayleigh’s criteria are necessary conditions for instability, and examples may be constructed which do satisfy their criterion, yet which are stable to infinitesimal perturbations. Note that the $\beta$-effect can stabilize the middle two profiles of Fig. 6.8, because if it is large enough $\beta - U_{yy}$ will be one-signed. However, the $\beta$-effect can destabilize a westward point jet, $U(y) = -(1 - |y|)$ (the negative of the jet in Fig. 6.3), because $\beta - U_{yy}$ is negative at $y = 0$ and positive elsewhere. An eastward point jet is stable, with or without $\beta$. 
6.4 Baroclinic Instability

Baroclinic instability is a hydrodynamic instability that occurs in stably stratified, rotating fluids, and it is ubiquitous in the planetary atmospheres and oceans. It gives rise to weather, and thus is perhaps the form of hydrodynamic instability that most affects the human condition.

6.4.1 A physical picture

We will first draw a picture of baroclinic instability as a form of ‘sloping convection’ in which the fluid, although statically stable, is able to release available potential energy when parcels move along a sloping path. To this end, let us first ask: what is the basic state that is baroclinically unstable? In a stably stratified fluid potential density decreases with height; we can also easily imagine a state in which the basic state temperature decreases, and the potential density increases, polewards. (We
will couch most of our discussion in terms of the Boussinesq equations, and henceforth drop the qualifier ‘potential’ from density.) Can we construct a steady solution from these two conditions? The answer is yes, provided the fluid is also rotating; rotation is necessary because the meridional temperature gradient generally implies a meridional pressure gradient; there is nothing to balance this in the absence of rotation, and a fluid parcel would therefore accelerate. In a rotating fluid this pressure gradient can be balanced by the Coriolis force and a steady solution maintained even in the absence of viscosity. Consider a stably-stratified Boussinesq fluid in geostrophic and hydrostatic balance on an \( f \)-plane, with buoyancy decreasing uniformly polewards. Then \( fu = -\partial \phi / \partial y \) and \( \partial \phi / \partial z = b \), where \( b = -g \rho' / \rho_0 \) is the buoyancy. These together give the thermal wind relation, \( \partial u / \partial z = \partial b / \partial y \).

If there is no variation of these fields in the zonal direction, then, for any variation of \( b \) with \( y \), this is a steady solution to the primitive equations of motion, with \( v = w = 0 \).

The density structure corresponding to a uniform increase of density in the meridional direction is illustrated in Fig. 6.9. Is this structure stable to perturbations? The answer is no, although the perturbations must be a little special. Suppose the particle at ‘A’ is displaced upwards; then, since the fluid is (by assumption) stably stratified it will be denser than its surroundings and hence experience a restoring force, and similarly if displaced downwards. Suppose, however, we interchange the two parcels at positions ‘A’ and ‘B’. Parcel ‘A’ finds itself surrounded by parcels of higher density that itself, and it is therefore buoyant; it is also higher than where it started. Parcel ‘B’ is negatively buoyant, and at a lower altitude than where it started. Thus, overall, the centre of gravity of the fluid has been lowered, and so its overall potential energy lowered. This loss in potential energy of the basic state must be accompanied by a gain in kinetic energy of the perturbation. Thus, the perturbation amplifies and converts potential energy to kinetic energy.
6.4 Baroclinic Instability

The loss of potential energy is easily calculated. Since

$$PE = \int \rho g dz$$  \hspace{1cm} (6.60)$$
the change in potential energy due to the interchange is

$$\Delta PE = g(\rho_A z_A + \rho_B z_B - \rho_A z_B - \rho_B z_B) = g(z_A - z_B)(\rho_A - \rho_B) = g\Delta \rho \Delta z$$  \hspace{1cm} (6.61)$$

If both $\rho_B > \rho_A$ and $z_B > z_A$ then the initial potential energy is larger than the final, energy is released and the state is unstable. If the slope of the isopycnals is $\phi$ [so that $\phi = - (\partial z \rho / \partial x \rho)]$ and the slope of the displacements is $\alpha$, then for a displacement of horizontal distance $L$ the change in potential energy is given by

$$\Delta PE = g\Delta \rho \Delta z = g \left( L \frac{\partial \rho}{\partial y} + L \alpha \frac{\partial \rho}{\partial z} \right) \alpha L = gL^2 \alpha \frac{\partial \rho}{\partial y} \left( 1 - \frac{\alpha}{\phi} \right),$$  \hspace{1cm} (6.62)$$

if $\alpha$ and $\phi$ are small. If $0 < \alpha < \phi$ then energy is released by the perturbation, and it is maximized when $\alpha = \phi/2$. For the atmosphere the actual slope of the isotherms is about $10^{-3}$, so that the slope and potential parcel trajectories are indeed shallow.

Although intuitively appealing, the thermodynamic arguments presented in this section pay no attention to satisfying the dynamical constraints of the equations of motion, and we now turn our attention to that,

6.4.2 Linearized quasi-geostrophic equations

To explore the dynamics of baroclinic instability we use the quasi-geostrophic equations, specifically a potential vorticity equation for the fluid interior and a buoyancy or temperature equation at two vertical boundaries, one representing the ground and the other the tropopause — the boundary between the troposphere and stratosphere at about 10 km. (The tropopause is not a true rigid surface, but the higher static stability of the stratosphere does inhibit vertical motion. We return to this in section 6.9.) For a Boussinesq fluid the equations are

$$\frac{\partial q}{\partial t} + \mathbf{u} \cdot \nabla q = 0, \hspace{1cm} 0 < z < H,$$

$$q = \nabla^2 \psi + \beta y + \frac{\partial}{\partial z} \left( \frac{F}{\partial z} \frac{\partial \psi}{\partial z} \right),$$  \hspace{1cm} (6.63)$$

where $F = f_0^2 / N^2$, and the buoyancy equation with $w = 0$,

$$\frac{\partial b}{\partial t} + \mathbf{u} \cdot \nabla b = 0, \hspace{1cm} z = 0, H,$$

$$b = f_0 \frac{\partial \psi}{\partial z}.$$  \hspace{1cm} (6.64)$$

A solution of these equations is a purely zonal flow, $\mathbf{u} = U(y, z)i$ with a corresponding temperature field given by thermal wind balance. The potential vorticity of this basic state is

$$Q = \beta y - U_y + \frac{\partial}{\partial z} \frac{\partial \psi}{\partial z} = \beta y + \Psi_{yy} + \frac{\partial}{\partial z} F \frac{\partial \psi}{\partial z}$$  \hspace{1cm} (6.65)$$
where $\Psi$ is the streamfunction of the basic state, related to $U$ by $U = -\partial \Psi / \partial y$. Linearizing (6.63) about this zonal flow gives the potential vorticity equation for the interior,

$$\frac{\partial q'}{\partial t} + U \frac{\partial q'}{\partial x} + v' \frac{\partial Q}{\partial y} = 0, \quad 0 < z < H \quad (6.66)$$

where $q' = \nabla^2 \psi' + \partial_z (F \partial_z \psi')$ and $v' = \partial \psi' / \partial x$. Similarly, the linearized buoyancy equation is

$$\frac{\partial b'}{\partial t} + U \frac{\partial b'}{\partial x} + v' \frac{\partial B}{\partial y} = 0, \quad z = 0, H \quad (6.67)$$

where $b' = f_0 \partial \psi' / \partial z$ and $\partial B / \partial y = \partial_y (f_0 \partial_z \Psi) = -f_0 \partial U / \partial z$.

Just as for the barotropic problem, a standard way to proceed is to seek normal-mode solutions of these equations. Since the coefficients of the equations are functions of $y$ and $z$, but not of $x$, we seek solutions of the form

$$\psi'(x, y, z, t) = \text{Re} \tilde{\psi}(y, z) e^{i(k(x - ct))}, \quad (6.68)$$

and similarly for the derived quantities $u'$, $v'$ and $q'$. In particular

$$\hat{q} = \tilde{\psi}_{yy} + \frac{\partial}{\partial z} F \frac{\partial \tilde{\psi}}{\partial z} - k^2 \tilde{\psi}. \quad (6.69)$$

Substituting (6.68) and (6.69) into (6.66) into (6.67) gives

$$(U - c) \left( \tilde{\psi}_{yy} + (F \tilde{\psi} z - k^2 \tilde{\psi}) + Q_y \tilde{\psi} \right) = 0, \quad 0 < z < H, \quad (6.70a)$$

$$z = 0, H. \quad (6.70b)$$

These equations are analogous to Rayleigh's equations for parallel shear flow, and emphasize the similarity between baroclinic instability and that of a parallel shear flow.

### 6.4.3 Necessary conditions for baroclinic instability

Necessary conditions for instability may be obtained following a procedure analogous to that used for parallel shear flows. First, integrating by parts, we note that

$$\int_{\gamma_1}^{\gamma_2} \tilde{\psi}^* \tilde{\psi} y' dy = \left[ \tilde{\psi}^* \tilde{\psi} y' \right]_{\gamma_1}^{\gamma_2} - \int_{\gamma_1}^{\gamma_2} |\tilde{\psi} y'|^2 dy. \quad (6.71)$$

If the integral is performed between two quiescent latitudes, or the domain is a channel with $\psi = 0$ at the boundaries, then the first term on the right-hand side vanishes. Similarly,

$$\int_0^H \tilde{\psi}^* (F \tilde{\psi} z) z' dz = \left[ F \tilde{\psi}^* \tilde{\psi} z \right]_0^H - \int_0^H F |\tilde{\psi} z|^2 dz
= \left[ FU \frac{|\tilde{\psi}|^2}{(U - c)} \right]_0^H - \int_0^H F |\tilde{\psi} z|^2 dz, \quad (6.72)$$
using (6.70b). Now, multiply (6.70a) by \( \tilde{\psi}^* \) and integrate over \( y \) and \( z \), and use (6.71) and (6.72) to obtain

\[
\int_{y_1}^{y_2} \int_0^H |\psi_y|^2 + F|\tilde{\psi}_z|^2 + k^2|\tilde{\psi}|^2 \, dy \, dz - \int_{y_1}^{y_2} \left[ \int_0^H \frac{Q_y}{U - c} |\tilde{\psi}|^2 \, dz + \left[ \frac{FU_z|\tilde{\psi}|^2}{U - c} \right]_0^H \right] \, dy = 0. \tag{6.73}
\]

The term on the first line is purely real. The term on the second line is complex, and its imaginary component is given by

\[
-c_1 \int_{y_1}^{y_2} \left[ \int_0^H \frac{Q_y}{U - c} |\tilde{\psi}|^2 \, dz + \left[ \frac{FU_z|\tilde{\psi}|^2}{U - c} \right]_0^H \right] \, dy = 0. \tag{6.74}
\]

If there is to be instability \( c_1 \) must be non-zero, and the integrand must therefore vanish. This gives the Charney-Stern-Pedlosky (CSP) necessary condition for instability, namely that one of the following criteria be satisfied:

(i) \( Q_y \) changes sign in the interior.
(ii) \( Q_y \) is the opposite sign to \( U_z \) at the upper boundary, \( z = H \).
(iii) \( Q_y \) is the same sign as \( U_z \) at the lower boundary, \( z = 0 \).
(iv) \( U_z \) is the same sign at the upper and lower boundaries, a condition which differs from (ii) or (iii) if \( Q_y = 0 \).

In the earth’s atmosphere, \( Q_y \) is often dominated by \( \beta \), and is positive everywhere, as, frequently, is the shear. The instability criterion is then normally satisfied through (iii): that is, both \( Q_y \) and \( U_z(0) \) are positive. A more general derivation that does not rely on normal mode disturbances is given in section 7.7.2.

6.5 THE EADY PROBLEM

We now proceed to explicitly calculate the stability properties of a particular configuration that has become known as the Eady problem. This was one of the first two mathematical descriptions of baroclinic instability, the other being the Charney problem. The two were formulated independently, each being the (largely unsupervised) Ph.D. thesis of its respective author, and although the Charney problem is in some respects more complete (for example in allowing a \( \beta \)-effect) the Eady problem displays the instability in a more transparent form. The Charney problem in its entirety is also quite mathematically opaque, and for these reasons we will first consider the Eady problem. The \( \beta \)-effect can be incorporated relatively simply in the two-layer model (the ‘Phillips problem’) considered in the next section, and in section 6.9.1 we look at some aspects of the Charney problem approximately. These problems were all initially envisioned as models for instabilities in the atmosphere, but the process of baroclinic instability is also ubiquitous in the ocean. To begin, let us make the following simplifying assumptions:

(i) The motion is on the \( f \)-plane (\( \beta = 0 \)). This assumption, although not particularly realistic, greatly simplifies the analysis.
(ii) The fluid is uniformly stratified; that is, \( N^2 \) is a constant. This is a decent approximation for the atmosphere below the tropopause, but less so for the
ocean where the stratification is quite non-uniform, being much larger in the upper ocean.

(iii) The basic state has uniform shear; that is, \( U_0(z) = \Lambda z = U z / H \) where \( \Lambda \) is the (constant) shear and \( U \) is the zonal velocity at \( z = H \) where \( H \) the domain depth. Again, this profile is more appropriate for the atmosphere than the ocean — below the thermocline the ocean is relatively quiescent and the shear small.

(iv) The motion is contained between two rigid, flat horizontal surfaces. In the atmosphere this corresponds to the ground and a ‘lid’ at a constant-height tropopause.

Assumptions (iii)–(iv) are rather inappropriate for the ocean, and will preclude us from drawing any quantitative conclusions about that system from our analysis. The most restrictive assumption vis-a-vis the atmosphere is (i).

### 6.5.1 The linearized problem

With a basic state streamfunction of \( \Psi = -\Lambda z y \), the basic state potential vorticity, \( Q \), is

\[
Q = \nabla^2 \Psi + \frac{H^2}{L_d^2} \frac{\partial}{\partial z} \left( \frac{\partial \Psi}{\partial z} \right) = 0.
\] (6.75)

The fact that \( Q = 0 \) makes the Eady problem a special case, albeit an illuminating one. The linearized potential vorticity equation is

\[
\left( \frac{\partial}{\partial t} + \Lambda \frac{\partial}{\partial x} \right) \left( \nabla^2 \psi' + \frac{H^2}{L_d^2} \frac{\partial^2 \psi'}{\partial z^2} \right) = 0
\] (6.76)

This equation has no \( x \)-dependent coefficients and in a periodic channel we may seek solutions in the form (6.68), namely \( \psi' (x, y, z, t) = \text{Re} \tilde{\psi}(y, z) e^{ik(x-ct)} \). Substituting this into (6.76) yields

\[
(\Lambda - c) \left( \frac{\partial^2 \tilde{\psi}}{\partial y^2} + \frac{H^2}{L_d^2} \frac{\partial^2 \tilde{\psi}}{\partial z^2} - k^2 \tilde{\psi} \right) = 0,
\] (6.77)

which is (6.70a) applied to the Eady problem.

### Boundary Conditions

There are two sets of boundary conditions to satisfy, the vertical boundary conditions at \( z = 0 \) and \( z = 1 \) and the lateral boundary conditions. In the horizontal plane we may either consider the flow to be confined to a channel, periodic in \( x \) and confined between two meridional walls, or, with a slightly greater degree of idealization but with little change to the essential dynamics, suppose that domain is doubly-periodic. Either case is dealt with easily enough by the choice of geometric basis function; we will choose a channel of width \( L \) and impose \( \psi = 0 \) at \( y = \pm L/2 \) and \( y = -L/2 \) and, to satisfy this, seek solutions of the form \( \Psi = \Phi(z) \sin ly \) or, using (6.68)

\[
\psi' (x, y, z, t) = \text{Re} \Phi(z) \sin ly e^{ik(x-ct)}.
\] (6.78)

where \( l = n \pi / L \) where \( n \) is a positive integer.
The vertical boundary conditions are that \( w = 0 \) at \( z = 0 \) and \( z = H \). We follow the procedure of section 6.4.2 and from (6.67) we obtain
\[
\left( \frac{\partial}{\partial t} + \Lambda z \frac{\partial}{\partial x} \right) \frac{\partial \psi'}{\partial z} - \Lambda \frac{\partial \psi'}{\partial x} = 0, \quad \text{at } z = 0, H. \tag{6.79}
\]

**Solutions**

Substituting (6.78) into (6.77) gives the interior potential vorticity equation
\[
(\Lambda z - c) \left[ \frac{H^2}{L_d^2} \frac{\partial^2 \Phi}{\partial z^2} - (k^2 + l^2) \Phi \right] = 0, \tag{6.80}
\]

and substituting (6.78) into (6.79) gives, at \( z = 0 \) and \( z = H \),
\[
c \frac{d\Phi}{dz} + \Lambda \Phi = 0 \quad \text{and} \quad (c - \Lambda H) \frac{d\Phi}{dz} + \Lambda \Phi = 0. \tag{6.81a,b}
\]

These are equivalent to (6.70b) applied to the Eady problem. If \( \Lambda z = c \) then (6.80) becomes
\[
H^2 \frac{d^2 \Phi}{dz^2} - \mu^2 \Phi = 0, \tag{6.82}
\]
where \( \mu^2 = L_d^2 (k^2 + l^2) \). The nondimensional parameter \( \mu \) is a horizontal wavenumber, scaled by the inverse of the Rossby radius of deformation. Solutions of (6.82) are
\[
\Phi(z) = A \cosh \mu \hat{z} + B \sinh \mu \hat{z}, \tag{6.83}
\]
where \( \hat{z} = z/H \); thus, \( \mu \) determines the vertical structure of the solution. The boundary conditions (6.81) are satisfied if
\[
A [\Lambda H] + B [\mu c] = 0, \quad A [(c - \Lambda H) \mu \sinh \mu + \Lambda H \cosh \mu] + B [(c - \Lambda H) \mu \cosh \mu + \Lambda H \sinh \mu] = 0. \tag{6.84}
\]

Equations (6.84) are two coupled homogeneous equations in the two unknowns \( A \) and \( B \). Non-trivial solutions will only exist if the determinant of their coefficients (the terms in square brackets) vanishes, and this leads to
\[
c^2 - UC + U^2 (\mu^{-1} \coth \mu - \mu^{-2}) = 0, \tag{6.85}
\]
where \( U \equiv \Lambda H \) and \( \coth \mu = \cosh \mu / \sinh \mu \). The solution of (6.85) is
\[
c = \frac{U}{2} \pm \frac{U}{\mu} \left[ \left( \frac{\mu}{2} - \coth \frac{\mu}{2} \right) \left( \frac{\mu}{2} - \tanh \frac{\mu}{2} \right) \right]^{1/2}. \tag{6.86}
\]

The waves, being proportional to \( \exp(-ikct) \), will grow exponentially if \( c \) has an imaginary part. Since \( \mu/2 > \tanh(\mu/2) \) for all \( \mu \), for an instability we require that
\[
\frac{\mu}{2} < \coth \frac{\mu}{2}, \tag{6.87}
\]
which is satisfied when \( \mu < \mu_c \) where \( \mu_c = 2.399 \). The growth rates of the instabilities themselves are given by the imaginary part of (6.86), multiplied by the \( x \)-wavenumber. That is
\[
\sigma = kc_i = k \frac{U}{\mu} \left[ \left( \coth \frac{\mu}{2} \right) \left( \frac{\mu}{2} - \tanh \frac{\mu}{2} \right) \right]^{1/2}. \tag{6.88}
\]
These solutions suggest a natural nondimensionalization: scale length by \( L_d \), height by \( H \) and time by \( L_d/U \). The *Eady growth rate* is the inverse of the time scaling, and is defined by

\[
\sigma_E = \frac{\Lambda H}{L_d} = \frac{U}{L_d}.
\]  

(6.89)

Its inverse, the Eady timescale, may also be written as

\[
T_E = \frac{L_d}{U} = \frac{NH}{f_0 U} = \frac{1}{\sqrt{Ri}} f_0,
\]

(6.90)

where \( Fr = U/(NH) \) and \( Ri = N^2/\Lambda^2 \) are the Froude and Richardson numbers for this problem.

From (6.88) we may determine that the maximum growth rate occurs when \( \mu = \mu_m = 1.61 \), with associated (nondimensional) growth rate of \( kc_i/\sigma_E = 0.31 \), and phase speed \( c_r/U = 0.5 \). Note that for any given \( x \)-wavenumber, the most unstable wavenumber has \( l = 0 \), so that \( L_d k = \mu \). The unstable \( x \)-wavenumbers and corresponding wavelengths occur for

\[
k < k_c = \frac{\mu_c}{L_d} = \frac{2.4}{L_d}, \quad \lambda > \lambda_c = \frac{2\pi L_d}{\mu_c} = 2.6L_d.
\]

(6.91a,b)

The wavenumber and wavelength at which the instability is greatest are:

\[
k_m = \frac{1.6}{L_d}, \quad \lambda_m = \frac{2\pi L_d}{\mu_m} = 3.9L_d.
\]

(6.92a,b)

These properties are illustrated in the left panels of Fig. 6.10 and in Fig. 6.11.

Given \( c \), we may use (6.84) to determine the vertical structure of the Eady wave and this is, to within an arbitrary constant factor,

\[
\Phi(\hat{z}) = \cosh \mu \hat{z} - \frac{U}{\mu_c} \sinh \mu \hat{z} = \left[ \cosh \mu \hat{z} - \frac{U c_r \sinh \mu \hat{z}}{\mu |c^2|} + i \frac{U c_i \sinh \mu \hat{z}}{\mu |c^2|} \right].
\]

(6.93)

The wave therefore has a phase, \( \theta(\hat{z}) \), given by

\[
\theta(\hat{z}) = \tan^{-1} \left( \frac{U c_i \sinh \mu \hat{z}}{\mu |c^2| \cosh \mu \hat{z} - U c_r \sinh \mu \hat{z}} \right).
\]

(6.94)

The phase and amplitude of the Eady waves are plotted in the right panels of Fig. 6.10 and their overall structure in Fig. 6.12 where we see the unstable wave tilting into the shear.

### 6.5.2 Atmospheric and oceanic parameters

To get a qualitative sense of the nature of the instability we choose some typical parameters, as follows.

*For the atmosphere*  
Let us choose

\[
H \sim 10 \text{ km}, \quad U \sim 10 \text{ m s}^{-1}, \quad N \sim 10^{-2} \text{ s}^{-1}.
\]

(6.95)
Fig. 6.10 Solution of the Eady problem, in non-dimensional units. (a) Growth rate, $k_c$, of the most unstable Eady modes (i.e., those with the gravest meridional scale) as a function of scaled wavenumber $\mu$, from (6.88) with $\Lambda = H = 1$. (b) The real (solid) and imaginary (dashed) wave speeds of those modes, as a function of horizontal wavenumber. (c) The phase of the single most unstable mode as a function of height. (d) The amplitude of that mode as a function of height. To obtain dimensional values, multiply the growth rate by $\Lambda H/L_d$ and the wavenumber by $1/L_d$.

Figure 6.11 Contours of growth rate, $\sigma$, in the Eady problem, in the $k$-$l$ plane using (6.88), nondimensionalized as in Fig. 6.10. The growth rate peaks near the deformation scale, and for any given zonal wavenumber the most unstable wavenumber is that with the gravest meridional scale.
Fig. 6.12 Left column: Vertical structure of the most unstable Eady mode. Top: contours of streamfunction. Middle: temperature, $\partial \psi / \partial z$. Bottom: meridional velocity, $\partial \psi / \partial x$. Negative contours are dashed, and two complete wavelengths are present in the horizontal. Poleward flowing (positive $v$) air is generally warmer than equatorward flowing air. Right column: Same, but now for a wave just beyond the short-wave cut-off.

We then obtain:

\begin{align*}
\text{Deformation Radius:} & \quad L_d = \frac{NH}{f} \approx \frac{10^{-2} \times 10^4}{10^{-4}} \approx 1000 \text{ km}, \quad (6.96) \\
\text{Scale of maximum instability:} & \quad L_{\text{max}} \approx 3.9L_d \approx 4000 \text{ km}, \quad (6.97) \\
\text{Growth Rate:} & \quad \sigma \approx 0.3 \frac{U}{L_d} \approx 0.3 \times 10^6 \text{ s}^{-1} \approx 0.26 \text{ day}^{-1}. \quad (6.98)
\end{align*}

For the ocean

For the main thermocline in the ocean let us choose

\[ H \sim 1 \text{ km} \quad U \approx 0.1 \text{ m s}^{-1} \quad N \sim 10^{-2} \text{ s}^{-1}. \quad (6.99) \]

We then obtain:

\begin{align*}
\text{Deformation Radius:} & \quad L_d = \frac{NH}{f} \approx \frac{10^{-2} \times 1000}{10^{-4}} = 100 \text{ km}, \quad (6.100) \\
\text{Scale of maximum instability:} & \quad L_{\text{max}} \approx 3.9L_d \approx 400 \text{ km}, \quad (6.101) \\
\text{Growth Rate:} & \quad \sigma \approx 0.3 \frac{U}{L_d} \approx 0.3 \times 10^5 \text{ s}^{-1} \approx 0.026 \text{ day}^{-1}. \quad (6.102)
\end{align*}
6.6 Two-Layer Baroclinic Instability

The length-scales of the instability are characterized by the deformation scale. The most unstable scale has a wavelength about four times the deformation radius \( L_d \), where \( L_d = NH/f_0 \).

The growth rate of the instability is approximately

\[
\sigma_E \sim \frac{U}{L_d} = \frac{f_0}{N}.
\]

That is, it is proportional to the shear, and scaled by the Prandtl ratio \( f_0/N \). The value \( \sigma_E \) is known as the Eady growth rate.

The most unstable waves for a given zonal scale are those with the gravest meridional scale.

There is a short-wave cutoff beyond which (i.e., at higher wavenumber than) there is no instability. This occurs near the deformation radius.

The instability relies on an interaction between waves at the upper and lower boundaries. If either boundary is removed, the instability dies. This point is be considered further in section 6.7.

In the ocean, the Eady problem is not quantitatively applicable because of the non-uniformity of the stratification and non-zonality of the flow. Nevertheless, the above estimates give a qualitative sense of the scale and growth rate of the instability relative to the corresponding values in the atmosphere. A summary of the main points of the Eady problem is given in the shaded box on the next page.

6.6 TWO-LAYER BAROCLINIC INSTABILITY

The eigenfunctions displaying the largest growth rates in the Eady problem have a relatively simple vertical structure. This suggests that an even simpler mathematical model of baroclinic instability might be constructed in which the vertical structure is a priori restricted to a very simple form, namely the two-layer QG model of sections 5.3.2 and 5.4.5. One notable advantage over the Eady model is that it is possible to include the \( \beta \)-effect in a simple way.

6.6.1 Posing the problem

For two layers or two levels of equal thickness, we write the potential vorticity equations in the dimensional form,

\[
\frac{D}{Dt} \left[ \zeta_i + \beta y + \frac{k_d^2}{2} (\psi_j - \psi_i) \right] = 0, \quad i = 1, 2, \quad j = 3 - i, \quad (6.103)
\]
where, using two-level notation for definiteness,

\[
\frac{k_d^2}{2} = \left( \frac{2f_0}{NH} \right)^2 \quad \rightarrow \quad k_d = \frac{\sqrt{8}}{L_d},
\]

(6.104)

where \(H\) is the total depth of the domain, as in the Eady problem. The basic state we choose is:

\[
\Psi_1 = -U_1 y, \quad \Psi_2 = -U_2 y = +U_1 y.
\]

(6.105)

It is possible to choose \(U_2 = -U_1\) without loss of generality because there is no topography and the system is Galilean invariant. The basic basic state potential vorticity gradient is then given by

\[
Q_1 = \beta y + k_d^2 U y, \quad Q_2 = \beta y - k_d^2 U y
\]

(6.106)

where \(U = U_1\). (Note that this differs by a constant factor from the \(U\) in the Eady problem.) Even in the absence of \(\beta\) there is a non-zero potential vorticity gradient. Why should this be different from the Eady problem? — after all, the shear is uniform in both problems. The difference arises from the vertical boundary conditions. In the standard layered formulation the temperature gradient at the boundary is absorbed into the definition of the potential vorticity in the interior. This results in a nonzero interior potential vorticity gradient at the two levels adjacent to the boundary (the only layers in the two-layer problem), but isothermal boundary conditions \(D/Dt(\partial \psi/\partial z) = 0\). In the Eady problem we have a zero interior gradient of potential vorticity but a temperature gradient at the boundary. The two formulations are physically equivalent — a finite-difference example of the Bretherton boundary layer.

The linearized potential vorticity equation is, for each layer,

\[
\frac{\partial q_i'}{\partial t} + U_i \frac{\partial q_i'}{\partial x} + v_i' \frac{\partial Q_i}{\partial y} = 0, \quad i = 1, 2
\]

(6.107)

or, more explicitly,

\[
\left[ \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right] \left[ \nabla^2 \psi_i' + \frac{k_d^2}{2} (\psi_2' - \psi_1') \right] + \frac{\partial \psi_i'}{\partial x} (\beta + k_d^2 U) = 0,
\]

(6.108a)

\[
\left[ \frac{\partial}{\partial t} - U \frac{\partial}{\partial x} \right] \left[ \nabla^2 \psi_2' + \frac{k_d^2}{2} (\psi_1' - \psi_2') \right] + \frac{\partial \psi_2'}{\partial x} (\beta - k_d^2 U) = 0.
\]

(6.108b)

For simplicity we will set the problem in a square, doubly-periodic domain, and so seek solutions in the form,

\[
\psi_i' = \text{Re} \tilde{\psi}_i e^{i(kx + ly - \omega t)} = \text{Re} \tilde{\psi}_i e^{ik(x - ct)} e^{ily}, \quad i = 1, 2.
\]

(6.109)

Here, \(k\) and \(l\) are the \(x\)- and \(y\)-wavenumbers, and \((k, l) = (2\pi/L)(m, n)\) where \(L\) is the size of the domain and \(m\) and \(n\) are integers. The constant \(\tilde{\psi}_i\) is the complex amplitude.
6.6.2 The solution
Substituting (6.109) into (6.108) we obtain
\[\begin{align*}
[ik(U - c)]\left[-K^2 \tilde{\psi}_1 + k_d^2 (\tilde{\psi}_2 - \tilde{\psi}_1) / 2\right] + ik\tilde{\psi}_1 (\beta + k_d^2 U) &= 0, \\
[-ik(U + c)]\left[-K^2 \tilde{\psi}_2 + k_d^2 (\tilde{\psi}_1 - \tilde{\psi}_2) / 2\right] + ik\tilde{\psi}_2 (\beta - k_d^2 U) &= 0,
\end{align*}\]
where \(K^2 = k^2 + l^2\). Re-arranging these two equations gives
\[\begin{align*}
\left[(U - c)(k_d^2/2 + K^2) - (\beta + k_d^2 U)\right] \tilde{\psi}_1 - \left[k_d^2 (U - c)/2\right] \tilde{\psi}_2 &= 0, \\
-\left[k_d^2 (U + c)/2\right] \tilde{\psi}_1 + \left[(U + c)(k_d^2/2 + K^2) + (\beta - k_d^2 U)\right] \tilde{\psi}_2 &= 0.
\end{align*}\]
These equations are of the form
\[\begin{align*}
[A] \tilde{\psi}_1 + [B] \tilde{\psi}_2 &= 0, \\
[C] \tilde{\psi}_1 + [D] \tilde{\psi}_2 &= 0,
\end{align*}\]
and for nontrivial solutions the determinant of coefficients must be zero; that is \(AD - BC = 0\). This gives a quadratic equation in \(c\) and solving this we obtain
\[c = -\frac{\beta}{K^2 + k_d^2} \left\{ 1 + \frac{k_d^2}{2K^2} \left[ 1 + \frac{4K^4(k^4 - k_d^4)}{k_\beta k_d^4} \right]^{1/2} \right\}, \tag{6.113}\]
where \(K^4 = (k^2 + l^2)^2\) and \(k_\beta = \sqrt{\beta U}\) (its inverse is known as the Kuo scale). We may nondimensionalize this equation using the deformation radius \(L_d\) as the length scale and the shear velocity \(U\) as the velocity scale. Then, denoting non-dimensional parameters with hats, we have
\[\begin{align*}
k &= \hat{k}, \\
c &= \hat{c} U, \\
t &= \frac{L_d}{U} \hat{t},
\end{align*}\]
and the nondimensional form of (6.113) is just
\[\hat{c} = -\frac{\hat{k}_\beta^2}{K^2 + k_d^2} \left\{ 1 + \frac{k_d^2}{2K^2} \left[ 1 + \frac{4\hat{K}^4(\hat{k}^4 - \hat{k}_d^4)}{\hat{k}_\beta \hat{k}_d^4} \right]^{1/2} \right\}, \tag{6.115}\]
where \(\hat{k}_\beta = k_\beta L_d\) and \(\hat{k}_d = \sqrt{\hat{d}}\), as in (6.104). The nondimensional parameter
\[\gamma = \frac{1}{4} \hat{k}_\beta^2 = \frac{\beta L_d^2}{4U}, \tag{6.116}\]
is often useful as a measure of the importance of \(\beta\); it is proportional to the square of the ratio of the deformation radius to the Kuo scale \(\sqrt{U/\beta}\). (It is also the two layer version of the ‘Charney-Green number’ considered more in section 6.9.1.) Let us look at two special cases first, before considering the general solution to these equations.

1. Zero shear, non-zero \(\beta\)
If there is no shear (i.e., \( U = 0 \)) then (6.111a) and (6.111b) are identical and two roots of the equation give the purely real phase speeds \( c \)

\[
c = -\frac{\beta}{K^2}, \quad c = -\frac{\beta}{K^2 + k_d^2}
\]

(6.117)

The first of these is the dispersion relationship for Rossby waves in a purely barotropic flow, and corresponds to the eigenfunction \( \tilde{\psi}_1 = \tilde{\psi}_2 \). The second solution corresponds to the baroclinic eigenfunction \( \tilde{\psi}_1 + \tilde{\psi}_2 = 0 \).

II. Zero \( \beta \), non-zero shear

If \( \beta = 0 \), then (6.111) yields, after a little algebra,

\[
c = \pm U \left( \frac{K^2 - k_d^2}{K^2 + k_d^2} \right)^{1/2}
\]

(6.118)

or, defining the growth rate \( \sigma \) by \( \sigma = -i\omega \),

\[
\sigma = U k \left( \frac{k_d^2 - K^2}{K^2 + k_d^2} \right)^{1/2}
\]

(6.119)

These expressions are very similar to those in the Eady problem. Indeed, as we increase the number of layers (using a numerical method to perform the calculation) the growth rate converges to that of the Eady problem (Fig. 6.13).

We note that:

* There is an instability for all values of \( U \).
* There is a high-wavenumber cut-off, at a scale proportional to the radius of deformation. For the two-layer model, if \( K > k_d = 2.82/L_d \) there is no growth. For the Eady problem, the high wavenumber cut-off occurs at \( 2.4/L_d \).
* There is no low wavenumber cut-off.
For any given $k$, the highest growth rate occurs for $l = 0$. In the two-layer model, from Eq. (6.119), for $l = 0$ the maximum growth rate occurs when $k = 0.634k_d = 1.79/L_d$. For the Eady problem, the maximum growth rate occurs at $1.61/L_d$.

Solution in the general case: non-zero shear and non-zero $\beta$

Using Eq. (6.115), the growth rate and wave speeds as function of wavenumber are plotted in Fig. 6.14. We observe that there still appears to be a high wavenumber cut-off and, for $\beta = 0$, there is a low-wavenumber cut-off. A little analysis elucidates these features.

The neutral curve:

For instability, there must be an imaginary component to the phase speed in Eq. (6.115). That is, we require

$$k^4_p k^4_d + 4K^4(K^4 - k^4_d) < 0.$$  \hfill (6.120)

This is a quadratic equation in $K^4$ for the value of $K$, $K_c$ say, at which the growth rate is zero. Solving, we find

$$K_c^4 = \frac{1}{2} k^4_d \left( 1 \pm \sqrt{1 - k^4_p/k^4_d} \right),$$  \hfill (6.121)

and this is plotted in Fig. 6.15. From Eq. (6.120) useful approximate expressions can be obtained for the critical shear as a function of wavenumber in the limits of small $K$ and $K \approx k_d$, and these are left as exercises for the reader.
Fig. 6.15 Contours of growth rate in the two-layer baroclinic instability problem. The dashed line is the neutral stability curve obtained from (6.121), and the other curves are contours of growth rates obtained from (6.115). Outside of the dashed line, the flow is stable. The wavenumber is scaled by $1/L_d$ (i.e., by $k_d/\sqrt{8}$) and growth rates are scaled by the inverse of the Eady timescale (i.e., by $U/L_d$). Thus, for $L_d = 1000$ km and $U = 10$ m s$^{-1}$, a nondimensional growth rate of 0.25 corresponds to a dimensional growth rate of $0.25 \times 10^{-5}$ s$^{-1} = 0.216$ day$^{-1}$.

**Minimum shear for instability:**

From (6.120), instability arises when $\beta^2 \hat{k}_d^4/U^2 < 4K^4(\hat{k}_d^4 - K^4)$. The maximum value of the right-hand side of this expression arises when $K^4 = \hat{k}_d^4/2$; thus, instability arises only when

$$\frac{\beta^2 \hat{k}_d^4}{U^2} < 4 \frac{\hat{k}_d^4 \hat{k}_d^4}{2}$$

(6.122)

or

$$U_s > \frac{2\beta}{k_d^4}$$

(6.123)

where $U_s = U_1 - U_2 = 2U$. In terms of the deformation radius itself the minimum shear for instability is

$$U_s > \frac{1}{4} \beta L_d^2.$$  

(6.124)

Fig. 6.16 sketches how this might vary with latitude in the atmosphere and ocean. (In (6.124), the shear is the difference in the velocity between level 1 and level 2, whereas the deformation radius, $NH/f_0$, is based on the total height of the fluid. If we were to use half the depth of the fluid in the definition of the deformation radius, the factor of 4 would disappear.) If the shear is just this critical value, the instability occurs at
6.6 Two-Layer Baroclinic Instability

The minimum shear \( (U_s = U_1 - U_2, \text{in m/s}) \) required for baroclinic instability in a two-layer model, calculated using (6.124), i.e. \( U_s = \beta L_d^2 / 4 \) where \( \beta = 2\Omega a^{-1} \cos \vartheta \) and \( L_d = NH / f \), where \( f = 2\Omega \sin \vartheta \). The left panel uses \( H = 10\) km and \( N = 10^{-2}\) s\(^{-1}\), the right panel uses parameters representative of the main thermocline, \( H = 1\) km and \( N = 10^{-2}\) s\(^{-1}\). The results are not quantitatively accurate, but the implications that the minimum shear is much less for the ocean, and that in both atmosphere and ocean the shear increases rapidly at low latitudes, are robust.

\[ k = 2^{-1/4}k_d = 0.84k_d = 2.37/L_d. \]
As the shear increases, the wavenumber at which the growth rate is maximum decreases slightly (see Fig. 6.15), and for a sufficiently large shear the \( \beta \)-effect is negligible and the wavenumber of maximum instability is, as we saw earlier, 0.634 \( k_d \) or 1.79/\( L_d \).

Note the relationship of the minimum shear to the basic state potential vorticity gradient in the respective layers. In the upper and lower layers the potential vorticity gradients are given by, respectively,

\[
\frac{\partial Q_1}{\partial y} = \beta + k_d^2U, \quad \frac{\partial Q_2}{\partial y} = \beta - k_d^2U
\]

(6.125a,b)
Thus, the requirement for instability is exactly that which causes the potential vorticity gradient to change sign somewhere in the domain, in this case becoming negative in the lower layer. This is an example of the general rule that potential vorticity (suitably generalized to include the surface boundary conditions) must change sign somewhere in order for there to be an instability.

**High-wavenumber cut-off:**
Instability can only arise when, from (6.120),

\[ 4K^4(k_d^4 - K^4) > k_d^4k_\beta^4 \]

so that a necessary condition for instability is

\[ k_d^2 > K^2. \]
(6.127)
Thus, waves shorter than the deformation radius are always stable, no matter what the value of \( \beta \). We also see from Fig. 6.14 and Fig. 6.15 that
the high wavenumber cut-off in fact varies little with \( \beta \) if \( k_d \gg k_\beta \). Note that the critical shear required for instability approaches infinity as \( K \) approaches \( k_d \).

**Low-wavenumber cut-off:**
Suppose that \( k \ll k_d \). Then (6.120) simplifies to \( k_\beta^4 < 4K^4 \). That is, for instability we require

\[
K^2 > \frac{1}{2} k_\beta^2 = \frac{\beta}{2U}.
\]  

(6.128)

Thus, using (6.127) and (6.128) the unstable waves lie approximately in the interval \( \beta/(\sqrt{2}U) < k < k_d \).

6.7 AN INFORMAL VIEW OF THE MECHANISM OF BAROCLINIC INSTABILITY

In this section we take a more intuitive look at baroclinic instability, trying to understand the mechanism without treating the problem in full generality or exactness. We will do this by way of a semi-kinematic argument that shows how the waves in each layer of a two-layer model, or the waves on the top and bottom boundaries in the Eady model, can constructively interact to produce a growing instability. It is kinematic in the sense that we initially treat the waves independently, and only subsequently allow them to interact — but it is this dynamical interaction that gives the instability. We first revisit the two-layer model and simplify it to its bare essentials.

6.7.1 The two-layer model

**A simple dynamical model**

We first re-derive the instability *ab initio* from the equations of motion written in terms of the baroclinic streamfunction \( \tau \) and the barotropic streamfunction \( \psi \) where

\[
\tau = \frac{1}{2}(\psi_1 - \psi_2), \quad \psi = \frac{1}{2}(\psi_1 + \psi_2).
\]  

(6.129)

We linearize about a sheared basic state of with zero barotropic velocity and with \( \beta = 0 \). Thus, with \( \psi = 0 + \psi' \) and \( \tau = -Uy + \tau' \) the linearized equations of motion, equivalent to (6.108) with \( \beta = 0 \), are

\[
\frac{\partial}{\partial t} \nabla^2 \psi' = -U \frac{\partial}{\partial x} \nabla^2 \tau', \quad (6.130a)
\]

\[
\frac{\partial}{\partial t} (\nabla^2 - k_d^2) \tau' = -U \frac{\partial}{\partial x} (\nabla^2 + k_d^2) \psi'. \quad (6.130b)
\]

Seeking solutions of the form \((\psi', \tau') = \text{Re}(\tilde{\psi}, \tilde{\tau}) \exp[i(k(x - ct))]\) gives

\[
c \tilde{\psi} - U \tilde{\tau} = 0, \quad (6.131a)
\]

\[
c (K^2 + k_d^2) \tilde{\tau} - U (K^2 - k_d^2) \tilde{\psi} = 0. \quad (6.131b)
\]

These equations have nontrivial solutions if the determinant of the matrix of coefficients is zero, giving the quadratic equation \( c^2(K^2 + k_d^2) - U^2(K^2 - k_d^2) = 0 \). Solving
this gives, reprising (6.118),

\[ c = \pm U \left( \frac{K^2 - k_d^2}{K^2 + k_d^2} \right)^{1/2}. \]  

(6.132)

Instabilities occur for \( K^2 < k_d^2 \), for which \( c = ic_i \); that is, the wave speed is purely imaginary. From (6.131) unstable modes have

\[ \tilde{\tau} = i \frac{c_i}{U} \tilde{\psi} = e^{i\pi/2} \frac{c_i}{U} \tilde{\psi}. \]  

(6.133)

That is, \( \tau \) lags \( \psi \) by \( 90^\circ \) for a growing wave \((c_i > 0)\). Similarly, \( \tau \) leads \( \psi \) by \( 90^\circ \) for a decaying wave. Now, the temperature is proportional to \( \tau \), and in the two-level model is advected by the vertically averaged perturbation meridional velocity, \( \nu \) say (with Fourier amplitude \( \tilde{\nu} \)), where \( \nu = \partial \psi / \partial x \). Thus, for growing or decaying waves,

\[ \tilde{\nu} = \tilde{\tau} \frac{kU}{c_i} \]  

(6.134)

and the meridional velocity is exactly in phase with the temperature for growing modes, and is out of phase with the temperature for decaying modes. That is, for unstable modes, polewards flow is correlated with high temperatures, and for decaying modes polewards flow is correlated with low temperatures. For neutral waves, \( \tilde{\tau} = c_r \tilde{\psi} / U \) and so \( \tilde{\nu} = ik\tilde{\tau}U / c_r \) and the meridional velocity and temperature are \( \pi/2 \) out of phase. Thus, to summarize:

- Growing waves transport heat (or buoyancy) polewards.
- Decaying waves transport heat equatorward.
- Neutral waves do not transport heat.

Further simplifications to the two-layer model

First consider (6.130) for waves much larger than the deformation radius, \( K^2 \ll k_d^2 \); we obtain

\[ \frac{\partial}{\partial t} \nabla^2 \psi = -U \frac{\partial}{\partial x} \nabla^2 \tau, \quad \frac{\partial}{\partial t} \tau = U \frac{\partial}{\partial x} \psi. \]  

(6.135a,b)

for which we obtain [either directly or from (6.132)] \( c = iU \); that is, the flow is unstable. To see the mechanism, suppose that the initial perturbation is barotropic and sinusoidal in \( x \), with no \( y \) variation. Polewards flowing fluid (i.e., \( \partial \psi / \partial x > 0 \)) will, by (6.135b), generate a positive \( \tau \), and the baroclinic flow will be out of phase with the barotropic flow. Then, by (6.135a), the advection of \( \tau \) by the mean shear produces growth of \( \psi \) that is in phase with the original disturbance. Contrast this case with that for very small disturbances, for which \( K^2 \gg k_d^2 \) and (6.130) becomes

\[ \frac{\partial}{\partial t} \nabla^2 \psi = -U \frac{\partial}{\partial x} \nabla^2 \tau, \quad \frac{\partial}{\partial t} \nabla^2 \tau = -U \frac{\partial}{\partial x} \nabla^2 \psi, \]  

(6.136a,b)

or, in terms of the equations for each layer,

\[ \frac{\partial}{\partial t} \nabla^2 \psi_1 = -U \frac{\partial}{\partial x} \nabla^2 \psi_1, \quad \frac{\partial}{\partial t} \nabla^2 \psi_2 = +U \frac{\partial}{\partial x} \nabla^2 \psi_2. \]  

(6.137a,b)

That is, the layers are completely decoupled and no instability can arise. Motivated by this, consider waves that propagate independently in each layer on the potential
vorticity gradient caused by $\beta$ (if non-zero) and shear. Thus in (6.108) we keep the potential vorticity gradients but neglect $k_2^2$ where it appears alongside $\nabla^2$ and find

$$\left[ \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right] \nabla^2 \psi_1 + \frac{\partial \psi_1}{\partial x} \frac{\partial Q_1}{\partial y} = 0,$$  
(6.138a)

$$\left[ \frac{\partial}{\partial t} - U \frac{\partial}{\partial x} \right] \nabla^2 \psi_2 + \frac{\partial \psi_2}{\partial x} \frac{\partial Q_2}{\partial y} = 0.$$  
(6.138b)

where $\partial Q_1/\partial y = \beta + k_2^2 U$ and $\partial Q_2/\partial y = \beta - k_2^2 U$. Seeking solutions of the form (6.109), the phase speeds of the associated waves are

$$c_1 = U - \frac{\partial_y Q_1}{K^2}, \quad c_2 = -U - \frac{\partial_y Q_2}{K^2}.$$  
(6.139a,b)

In the upper layer the phase speed is a combination of an eastward advection and a fast westward wave propagation due to a strong potential vorticity gradient. In the lower layer the phase speed is a combination of a westward advection and a slow eastward wave propagation due to the weak potential vorticity gradient. The two phase speeds are, in general, not equal, but they would need to be so if they are to combine to cause an instability. From (6.139) this occurs when $K^2 = k_2^2$ and $c_1 = c_2 = -\beta/k_2^2$. These conditions are just those occurring at the high-wavenumber cut-off to instability in the two-level model. At higher wavenumbers, the waves are unable to synchronize, whereas at lower wavenumbers they may become inextricably coupled.

Let us suppose that the phase of the wave in the upper layer lags that (i.e., is westward of) that in the lower layer, as illustrated in the top panel Fig. 6.17. The lower panel shows the temperature field, $\tau = (\psi_1 - \psi_2)/2$, and the average meridional velocity, $v = \partial_x (\psi_1 + \psi_2)/2$. In this configuration, the temperature field is in phase with the meridional velocity, meaning that warm fluid is advected polewards. Now, let us allow the waves in the two layers to interact by adding one dynamical equation, the thermodynamic equation, which in its simplest form is

$$\frac{\partial \tau}{\partial t} = -v \frac{\partial \tau}{\partial y} = vU,$$  
(6.140)

where $\tau$ is proportional basic state temperature field. The temperature field, $\tau$, grows in proportion to $v$, which is proportional to $\tau$ if the waves tilt westward with height, and an instability results. This dynamical mechanism is just that which is compactly described by (6.135). It is a straightforward matter to show that if the streamfunction tilts eastward with height, $v$ is out of phase with $\tau$ and the waves decay.

### 6.7.2 Interacting edge waves in the Eady problem

A very similar description applies to the Eady problem. As in the two-layer case, first consider the case in which the bottom and top surfaces are essentially uncoupled. Instead of solutions of (6.82) that have the structure (6.83) (which satisfies both boundary conditions) consider solutions that separately satisfy the bottom and top boundary conditions and that decay into the interior. These are, including the $x$-dependence,

$$\psi_B = \text{Re} \ A_B e^{ik(x-c_B t)} e^{-\mu z/H}, \quad \psi_T = \text{Re} \ A_T e^{ik(x-c_T t)} e^{\mu(z-H)/H}.$$  
(6.141a,b)
for the bottom and top surfaces respectively, and $\phi$ is the phase shift, with $A_B$ and $A_T$ being real constants. The boundary conditions (6.81) then determine the phase speeds of the two systems and we find

$$c_B = \frac{\Lambda H}{\mu}, \quad c_T = \Lambda H \left(1 - \frac{1}{\mu}\right). \quad (6.142a,b)$$

These are the phase speeds of edge waves in the Eady problem; they are real and in general they are unequal. It must therefore be the interaction of the waves on the upper and lower boundaries that is necessary for instability, because the unstable wave has but a single phase speed. This interaction can occur when their phase speeds are equal and from (6.142) this occurs when $\mu = 2$, giving

$$k = \frac{2}{L_d} \quad \text{and} \quad c = \frac{\Lambda H}{2} \quad (6.143a,b)$$

This phase speed is just that of the flow at mid-level, and at the critical wavenumber in the full Eady problem [$k_c = 2.4/L_d$, from (6.91)] the phase speed is purely real and equal to that of (6.143b) — see Fig. 6.10. Thus, (6.143) approximately characterizes the critical wavenumber in the full problem.

To turn this kinematic description into a dynamical instability, suppose that the two rigid surfaces are close enough so that the waves can interact, but still far enough so that their structure is approximately given by (6.141). (Note that if $\mu$ is too large, the waves decay rapidly away from the edges and will not interact.) Specifically, let the buoyancy perturbation at a given boundary be advected by the total meridional velocity perturbation, including that arising from the perturbation at the other boundary, so that at the top and bottom boundaries

$$\frac{\partial b_T'}{\partial t} = -(v_T' + v_T^') \frac{\partial B_T}{\partial y'}, \quad \frac{\partial b_B'}{\partial t} = -(v_B' + v_T^') \frac{\partial B_B}{\partial y'} \quad (6.144)$$

The waves will reinforce each other if $v_T'$ is in phase with $b_T'$ at the lower boundary, and if $v_B'$ is in phase with $b_T'$ at the upper boundary. Now, using (6.141), the velocity
Figure 6.18  Interacting edge waves in the Eady model. The upper panel shows waves on the top surface, and the lower panel shows waves on the bottom. If the streamfunction tilts westward with height, then the temperature on the top (bottom) is correlated with the meridional velocity on the bottom (top), the waves can reinforce each other. See also Fig. 6.12.

\[ b_B = -k N A_B e^{-\mu z/H}, \quad b_T = k N A_T e^{i\phi} e^{\mu(z-H)/H}, \]  
\[ v_B = i k A_B e^{-\mu z/H}, \quad v_T = i k A_T e^{i\phi} e^{\mu(z-H)/H}. \]  

(6.145a)  

(6.145b)  

The fields \( b_T \) and \( v_T \), and \( b_B \) and \( v_B \), will be positively correlated if \( 0 < \phi < \pi \), and will be exactly in phase if \( \phi = \pi/2 \), and this case is illustrated in Fig. 6.18. Just as in the two-layer case, this phase corresponds to a westward tilt with height, and this, in conjunction with geostrophic and hydrostatic balance, that allows warm fluid to move poleward and available potential energy to be released. From (6.144), the perturbation will grow and an instability will result. The analogy between baroclinic instability and barotropic instability should be evident from the similarity of this description and that of section 6.2.4, with \( z \) in the baroclinic problem playing the role of \( y \) in the barotropic problem, and \( b \) the role of \( v \). However, the analogy is not perfect because the boundary condition that \( \omega = 0 \) does not have an exact correspondence in the barotropic problem. Also, the nonlinear development of the baroclinic problem, discussed in chapter 9, is generally three-dimensional.

### 6.8 * THE ENERGETICS OF LINEAR BAROCLINIC INSTABILITY*

In baroclinic instability, warm parcels move poleward and cold parcels move equatorward. This motion draws on the available potential energy of the mean state, because warm light parcels move upward, and cold dense parcels downward and the height of the mean center of gravity of the fluid falls, and the loss of potential energy is converted to kinetic energy of the perturbation. However, because the instability is growing, the energy of the perturbation is of course not conserved, and both the kinetic energy and the available potential energy of the perturbation will grow. However, we still expect a conversion of potential energy to kinetic, and the purpose of this section is to demonstrate that explicitly. For simplicity, we restrict attention to the flat-bottomed two-level model with \( \beta = 0 \).

As in section 5.6, the energy may be partitioned into kinetic energy and available
The kinetic energy in a three-dimensional quasi-geostrophic flow is given by:

\[ KE = \frac{1}{2} \int (\nabla \psi)^2 \, dV \]  

(6.146)

which, in the case of the two-layer model becomes:

\[ KE = \frac{1}{2} \int (\nabla \psi_1)^2 + (\nabla \psi_2)^2 \, dA = \int (\nabla \psi)^2 + (\nabla \tau)^2 \, dA. \]  

(6.147)

Restricting attention to a single Fourier mode this becomes:

\[ KE = k^2 \tilde{\psi}^2 + k^2 \tilde{\tau}^2. \]  

(6.148)

The available potential energy in the continuous case is given by:

\[ APE = \frac{1}{2} \int \left( \frac{f_0 N}{N} \right)^2 (\partial \psi / \partial z)^2 \, dV. \]  

(6.149)

For a single Fourier mode in a two layer model this becomes:

\[ APE = k^2 \tilde{\tau}^2. \]  

(6.150)

Now, the nonlinear vorticity equations for each level are:

\[ \frac{\partial}{\partial t} \nabla^2 \psi_1 + J(\psi_1, \nabla^2 \psi_1) = -2 \frac{f_0 w}{H}, \]  

(6.151a)

\[ \frac{\partial}{\partial t} \nabla^2 \psi_2 + J(\psi_2, \nabla^2 \psi_2) = 2 \frac{f_0 w}{H}, \]  

(6.151b)

where \( w \) is the vertical velocity between the levels. (These equations are the two-level analogs of the continuous vorticity equation, with the right-hand sides being finite difference versions of \( f_0 \partial w / \partial z \).) Multiplying the two equations of (6.151) by \( -\psi_1 \) and \( -\psi_2 \), respectively, and adding we readily find:

\[ \frac{d}{dt} KE = \frac{4f_0}{H} \int w \tau \, dA. \]  

(6.152a)

For a single Fourier mode this becomes:

\[ \frac{d}{dt} KE = \text{Re} \left( \frac{4f_0}{H} \tilde{w} \tilde{\tau}^* \right), \]  

(6.152b)

where \( w = \tilde{w} \exp[i(kx - ct)] + \text{c.c.} \), and the asterisk denotes complex conjugacy.

The continuous thermodynamic equation is:

\[ \frac{Db}{Dt} + w N^2 = 0. \]  

(6.153)

Using \( b = f_0 \partial \psi / \partial z \) and finite-differencing [with \( \partial \psi / \partial z \rightarrow (\psi_1 - \psi_2)/(H/2) = 4 \tau / H \)], we obtain the two-level thermodynamic equation:

\[ \frac{\partial \tau}{\partial t} + J(\psi, \tau) \frac{w N^2 H}{4f_0} = 0. \]  

(6.154)
The change of available potential energy is obtained from this by multiplying by \( k_d^2 \tau \) and integrating, giving

\[
\int \left( \frac{1}{2} \frac{d}{dt} k_d^2 \tau^2 + \tau u \frac{2f_0}{H} \right) dA = 0
\]

(6.155)

or

\[
\frac{d}{dt} APE = -4f_0 \int w \tau dA
\]

(6.156a)

or, for a single Fourier mode,

\[
\frac{d}{dt} APE = -Re \frac{4f_0}{H} \tilde{w} \tilde{\tau}^*.
\]

(6.156b)

From (6.152) and (6.156) it is clear that in the nonlinear equations the sum of the kinetic energy and the available potential energy is conserved.

We now specialize by obtaining \( w \) from the linear baroclinic instability problem. Using this in (6.152) and (6.156) will give us the conversion between kinetic energy and potential energy in the growing baroclinic wave. It is important to realize that the total energy of the disturbance will not be conserved — both the potential and kinetic energy are growing, exponentially in this problem, because they are extracting energy from the mean state. To calculate \( w \) we use the linearized thermodynamic equation. From (6.154) this is

\[
\frac{\partial \tau}{\partial t} - U \frac{\partial \psi}{\partial x} + \frac{H \omega N^2}{4f_0} = 0,
\]

(6.157)

omitting the primes on perturbation quantities. For a single Fourier mode, this gives

\[
\frac{4N^2}{4f_0} \tilde{w} = i k \tilde{\tau} + U \tilde{\psi}.
\]

(6.158)

But, from (6.131), \( c \tilde{\psi} = U \tilde{\tau} \) in two-layer \( f \)-plane baroclinic instability and so

\[
\frac{4N^2}{4f_0} \tilde{w} = i c \tilde{\tau} \left( 1 + \frac{U^2}{c^2} \right) = i k c \tilde{\tau} \left( \frac{2K^2}{K^2 - k_d^2} \right).
\]

(6.159)

using (6.132). For stable waves, \( K^2 > k_d^2 \) and \( c = c_r \) and in that case the vertical velocity is \( \pi/2 \) out of phase with the temperature, and there is no conversion of APE to KE. For unstable waves \( c = ic_i \) and \( K^2 < k_d^2 \), and the vertical velocity is in phase with the temperature. That is, warm air is rising and so there is a conversion of APE to KE. To see this more formally, recall that the conversion from APE to KE is given by \( 4\tilde{w} \tilde{\tau}^* f_0 / H \). Thus, using (6.159),

\[
\frac{d}{dt} \left( APE \rightarrow KE \right) = Re 2ic \frac{k_d^2}{K^2 - k_d^2} \tilde{\tau}^2,
\]

(6.160)

using also the definition of \( k_d \) given in (6.104). If the wave is growing, then \( K^2 < k_d^2 \) and \( c = ic_i \) and the right-hand side is real and positive. For neutral waves, If \( c = c_r \) the right-hand side of (6.160) is pure imaginary, and so the conversion is zero. This completes our demonstration that baroclinic instability converts potential energy into kinetic energy.
6.9 * BETA, SHEAR AND STRATIFICATION IN A CONTINUOUS MODEL

The two-layer model of section 6.6 indicates that $\beta$ has a number of important effects of $\beta$ on baroclinic instability. Do these carry over to the continuously stratified case? The answer by-and-large is yes, but with some important qualifications that generally concern weak or shallow instabilities. In particular, we will find that there is no short-wave cut-off in the continuous model with non-zero beta, and that the instability determines its own depth scale. We will illustrate these properties first by way of scaling arguments, and then by way of numerical calculations.

6.9.1 Scaling arguments for growth rates, scales and depth

With finite density scale height and non-zero $\beta$, the quasi-geostrophic potential vorticity equation, linearized about a mean zonal velocity $U(z)$, is

$$
\left( \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) q' + \frac{\partial q'}{\partial y} \frac{\partial q}{\partial y} = 0, \tag{6.161}
$$

where

$$
q' = \nabla^2 q' + \frac{f_0^2}{\rho} \frac{\partial}{\partial z} \left( \rho \frac{\partial U}{\partial z} \right), \tag{6.162}
$$

$$
\frac{\partial q}{\partial y} = \beta - \frac{f_0^2}{\rho} \frac{\partial}{\partial z} \left( \rho \frac{\partial U}{\partial z} \right), \tag{6.163}
$$

and $\rho$ is a specified density profile. If we assume that $U = \Lambda z$ where $\Lambda$ is constant and that $N$ is constant, and let $H_{\rho}^{-1} = -\rho^{-1} \partial \rho / \partial z$, then

$$
\frac{\partial q}{\partial y} = \beta + \frac{f_0^2 \Lambda}{N^2 H_{\rho}} = \beta(1 + \alpha), \tag{6.164}
$$

where

$$
\alpha = \left( \frac{f_0^2 \Lambda}{\beta N^2 H_{\rho}} \right), \tag{6.165}
$$

The boundary conditions on (6.161) are

$$
\left( \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \frac{\partial \psi'}{\partial z} - \frac{\partial \psi'}{\partial x} \frac{\partial \psi}{\partial z} = 0, \quad \text{at } z = 0 \tag{6.166}
$$

and that $\psi \to 0$ as $z \to \infty$. The problem we have defined essentially constitutes the Charney problem. We can reduce this to the Eady problem by setting $\beta = 0$ and $H_{\rho} = \infty$, and providing a lid some finite height above the ground.

As in the Eady problem, we seek solutions of the form

$$
\psi = \text{Re} \tilde{\psi}(z) e^{i(kx + ly - kct)}, \tag{6.167}
$$

and substituting into (6.161) gives

$$
\left( \frac{f_0^2}{H_{\rho}^2 N^2} \right) \left( H_{\rho}^2 \frac{d^2 \tilde{\psi}}{dz^2} - H_{\rho} \frac{d \tilde{\psi}}{dz} \right) - \left( K^2 - \frac{\beta + f_0^2/(N^2 H_{\rho})}{\Lambda z - c} \right) \tilde{\psi} = 0. \tag{6.168}
$$
The Boussinesq version of this expression for a fluid contained between two horizontal surfaces is obtained by letting \( H = \infty \), giving
\[
\left( \frac{f_0^2}{N^2} \right) \frac{d^2 \tilde{\psi}}{d\hat{z}^2} - \left( K^2 - \frac{\beta}{\Lambda - \hat{c}} \right) \tilde{\psi} = 0. \tag{6.169}
\]

It seems natural to nondimensionalize (6.168) using:
\[
z = H\rho \hat{z}, \quad c = \Lambda H\rho \hat{c}, \quad K = \left( \frac{f_0}{NH\rho} \right) \hat{K},
\]
whence it becomes
\[
\frac{d^2 \tilde{\psi}}{d\hat{z}^2} - \left( \hat{K}^2 - \frac{\gamma + 1}{\hat{z} - \hat{c}} \right) \tilde{\psi} = 0 \tag{6.170}
\]
where
\[
\gamma = \alpha^{-1} = \frac{\beta N^2 H\rho}{f_0^2 \Lambda} = \frac{\beta L_d}{H\rho \Lambda} = \frac{H\rho}{h}.
\tag{6.171}
\]
where \( h \equiv \Lambda f_0^2 / (\beta N^2) \). The non-dimensional parameter \( \gamma \) is known as the Charney-Green number.\(^{15} \)

The Boussinesq version, (6.169), may be non-dimensionalized using \( H_D \) in place of \( H\rho \), where \( H_D \) is the depth of the fluid between two rigid surfaces. In that case
\[
\frac{d^2 \tilde{\psi}}{d\hat{z}^2} - \left( \hat{K}^2 - \frac{1}{\hat{z} - \hat{c}} \right) \tilde{\psi} = 0 \tag{6.172}
\]
where here the non-dimensional variables are scaled with \( H_D \).

Now, suppose that \( \gamma \) is large, for example if \( \beta \) or the static stability are large or the shear is weak. Eq. (6.171) admits of no non-trivial balance, suggesting that we rescale the variables using \( h \) instead of \( H\rho \) as the vertical scale in (6.170). The rescaled version of (6.171) is then
\[
\frac{d^2 \tilde{\psi}}{d\hat{z}^2} - \frac{1}{\gamma} \frac{d\tilde{\psi}}{d\hat{z}} - \left( \hat{K}^2 - \frac{1 + \gamma^{-1}}{\hat{z} - \hat{c}} \right) \tilde{\psi} = 0 \tag{6.173}
\]
or, approximately,
\[
\frac{d^2 \tilde{\psi}}{d\hat{z}^2} - \left( \hat{K}^2 - \frac{1}{\hat{z} - \hat{c}} \right) \tilde{\psi} = 0 \tag{6.174}
\]
This is exactly the same equation as results from a similar rescaling of the Boussinesq system, (6.173), as we might have expected because now the dynamical vertical scale, \( h \), is much smaller than the scale height \( H\rho \) (or \( H_D \)) and the system is essentially Boussinesq. Thus, noting that (6.175) has the same nondimensional form as (6.173) save that \( \gamma \) is replaced by unity, and that (6.175) with \( \gamma = 1 \) must produce the same scales and growth rates as in the Eady problem, we may deduce that:

(i) The wavelength of the instability is \( O(Nh/f_0) \).

(ii) The growth rate of the instability is \( O(Kc) = O(f_0\Lambda/N) \).

(iii) The vertical scale of the instability is \( O(h) = O((f_0^2 \Lambda)/(\beta N^2)) \).

These are the same as for the Eady problem, except with the dynamical height \( h \) replacing the geometric or scale height \( H_D \). Effectively, the dynamics has determined
its own vertical scale, $h$, that is much less than the scale height or geometric height, producing 'shallow modes'.

In the limit $\gamma \ll 1$ (strong shear, weak $\beta$), the Boussinesq and compressible problems differ. The Boussinesq problem reduces to the Eady problem, considered previously, whereas (6.171) becomes, approximately,

$$
\frac{d^2 \tilde{\psi}}{dz^2} - \frac{d \tilde{\psi}}{dz} - \left( \tilde{K}^2 - \frac{1}{z - \tilde{c}} \right) \tilde{\psi} = 0,
$$

and in this limit the appropriate vertical scale is the density scale height $H_{\rho}$. Because $H_{\rho} \gg h$ these are 'deep modes', occupying the entire vertical extent of the domain.

The scale $h$ does not arise in the two-level model, but there is a connection between it and the critical shear for instability in the two-level model. The condition $\gamma \ll 1$, or $h \gg H$, may be written as

$$
H \Lambda \gg \beta \left( \frac{NH}{f_0} \right)^2.
$$

Compare this with the necessary condition for instability in a two-level model, (6.124), namely

$$
(U_1 - U_2) > \beta \left( \frac{NH_{\Delta}}{f_0} \right)^2
$$

where $H_{\Delta}$ is the vertical distance between the two levels. Thus, essentially the same condition governs the onset of instability in the two-level model as governs the production of deep modes in the continuous model. This correspondence is a natural one, because in the two-level model all modes are 'deep', and the model fails (as it should) to capture the shallow modes of the continuous system. For similar reasons, there is a high-wavenumber cut-off in the two-level model: in the continuous model these modes are shallow and so cannot be captured by two-level dynamics. Somewhat counter-intuitively, for these modes the $\beta$-effect must be important, even though the modes have small horizontal scale: when $\beta = 0$ the instability arises via an interaction between edge waves at the top and bottom of the domain, whereas the shallow instability arises via an interaction of the edge waves at the surface with Rossby waves just above the surface.

### 6.9.2 Some numerical calculations

**Adding $\beta$ to the Eady model**

Our first step is add the $\beta$-effect to the Eady problem. That is, we suppose a Boussinesq fluid with uniform stratification, that the shear is zonal and constant, and that the entire problem is sandwiched between two rigid surfaces. Growth rates and phase speeds of such an instability calculation are illustrated in Fig. 6.19 and the vertical structure is shown in Fig. 6.20. As in the two-layer problem, there is a low-wavenumber cut-off to the main instability, although there is now an additional weak instability at very large-scales. These so-called Green modes have no counterpart in the two-layer model — they are deep, slowly growing modes that will be dominated by faster growing modes in most real situations. (Also, the fact
Fig. 6.19 Growth rates and wave speeds for the two-layer (solid) and continuous (dashed) models, with the same values of the Charney-Green number, $\gamma$, and uniform shear and stratification. (In the two-layer case $\gamma = \beta L_2^2 / [2(U_1 - U_2)] = 0.5$, and in the continuous case $\gamma = \beta L_2^2 / (H \Lambda) = 0.5.$) In the continuous case only the wave speed associated with the unstable mode is shown. In the two-layer case there are two real wave speeds which coalesce in the unstable region. The two-layer model has an abrupt short-wave and long-wave cut-off, whereas the growth rate of the continuous model tails off gradually at small wavelengths, and has a weak instability (the ‘Green modes’) at large wavelengths.

that the Green modes have a scale much larger than the deformation scale suggests that a degree of caution in the accuracy of the quasi-geostrophic calculation is warranted.) At high wavenumbers is no cut-off to the instability in the continuous problem in the case of non-zero beta; the high-wavenumber modes are shallow and unstable via an interaction between edge waves at the lower boundary and Rossby waves in the lower atmosphere, and so have no counterpart in either the the two-layer problem (where the modes are deep) or the Eady problem (which has no Rossby waves).

**Effects of nonuniform shear and stratification**

If the shear or stratification is non-uniform an analytic treatment is, even in problems without $\beta$, usually impossible and the resulting equations must be solved numerically. However, if we restrict attention to a discontinuity in the shear or the stratification, then resulting problem is very similar to the problem with rigid boundaries, and this property provides much of the justification for using the Eady problem to model instabilities in the earth’s atmosphere: in the troposphere the stratification is (approximately) constant, and the rapid increase in stratification in the stratosphere can be approximated by a lid at tropopause. Heuristically, we can see this from the form of the thermodynamic equation, namely

$$\frac{Dh}{Dt} + N^2 w = 0. \quad (6.179)$$

If $N^2$ is high this suggests $w$ will be small, and a lid is the limiting case of this. The oceanic problem is rather more involved, because although both the stratification and shear are concentrated in the upper ocean, they vary relatively smoothly;
Fig. 6.20 Vertical structure of the most unstable modes in a continuously stratified instability calculation with $\beta = 0$ (dashed lines, the Eady problem) and $\beta \neq 0$ (solid lines), as in Fig. 6.19. The effect of beta is to depress the height of maximum amplitude of the instability.

Furthermore, the shear is high where the stratification is high, and the two have opposing effects.

To go one step further, consider the Boussinesq potential vorticity equation, linearized about a zonally uniform state $\Psi(y,z)$, with a rigid surface at $z = 0$. The normal-mode evolution equations are similar to (6.70), namely

\[
(U - c) \left[ \frac{\partial^2}{\partial y^2} - k^2 + \frac{\partial}{\partial z} \left( F \frac{\partial}{\partial z} \right) \right] \tilde{\psi} + \frac{\partial Q}{\partial y} \tilde{\psi} = 0, \quad z > 0, \tag{6.180a}
\]

\[
(U - c) \frac{\partial \tilde{\psi}}{\partial z} - \frac{\partial U}{\partial z} \tilde{\psi} = 0, \quad \text{at } z = 0. \tag{6.180b}
\]

where $\partial_y Q = \beta - \partial_{yy} U - \partial_z (F \partial_z U)$. Now suppose that there is a discontinuity in the shear and/or the stratification in the interior of the fluid, at some level $z = z_c$.

Integrating (6.180a) across the discontinuity, noting that $\tilde{\psi}$ is continuous in $z$, gives

\[
(U - c) \left[ F \frac{\partial \tilde{\psi}}{\partial z} \right]_{z_c^+}^{z_c^-} - \tilde{\psi}(y,z_c) \left[ F \frac{\partial U}{\partial z} \right]_{z_c^-}^{z_c^+} = 0. \tag{6.181}
\]

which has similar form to (6.180b). This construction is evocative of the equivalence of a delta-function sheet of potential vorticity at a rigid boundary, except that now a discontinuity in the potential vorticity in the interior has a similarity with a rigid boundary.

We can illustrate the effects of an interior discontinuity that crudely represents the tropopause by numerically solving the linear eigenvalue problem. For simplicity, we pose the problem on the $f$-plane, in a horizontally doubly-periodic domain, with no horizontal variation of shear, and between two horizontal rigid lids. The eigenvalue problem is defined by (6.70), and the numerical procedure then solves for the complex eigenvalue $c$ and eigenfunction $\tilde{\psi}(z)$; various results are illustrated in Fig. 6.21. To parse this rather complex figure, first look at the solid curves in all the panels. These arise when the problem is solved with a uniform shear and a uniform stratification, with a lid at $z = 0$ and $z = 1$, so simply giving the Eady
Fig. 6.21 The effect of a stratosphere on baroclinic instability. (a) the given profiles of shear and stratification; (b) The growth rate of the instabilities; (c) amplitude of the most unstable mode as a function of height; (d) phase of the most unstable mode. The instability problem is numerically solved with various profiles of stratification and shear. In each profile, in the idealized troposphere ($z < 1$) the shear and stratification are uniform and the same in each case. We consider four idealized stratospheres ($z \geq 1$): 1, A lid at $z = 1$, i.e., no stratosphere, so Eady problem itself (profiles A+D, solid lines); 2, Stratospheric stratification same as the troposphere, but zero shear (profiles a+c, dashed); 3, Stratospheric shear same as troposphere, but stratification ($N^2$) four times the tropospheric value (b+d, dot-dashed); 4, Zero shear and high stratification in the stratosphere (b+c, dotted). In the troposphere the amplitude and structure of the instability is similar in all cases, illustrating the similarity of a rigid-lid and abrupt changes in shear or stratification. Either a high stratification or a low shear (or both) will result in weak stratospheric instability.

The familiar growth rates and vertical structure of the solution are given by the solid curves in panels (b), (c) and (d), and these are just the same as in Fig. 6.10. The various dotted and dashed curves show the results when the lid at $z = 1$ is replaced by stratosphere stretching from $1 < z < 2$ either with high stratification, zero shear, or both, an in all of these cases the stratosphere acts in the same qualitative way as a rigid lid. The vertical structure of the solution in the troposphere in all cases is quite similar, and the amplitude decays rapidly above the idealized tropopause, consistent with the almost uniform phase of the disturbance illustrated in panel (d) — recall that a tilting of the disturbance with height is
Fig. 6.22 The baroclinic instability in an idealized ocean, with four different profiles of shear or stratification. The panels are: (a) The profiles of velocity and density (and so $N^2$) used; (b) the growth rates of the various cases; (c) the vertical structure of the amplitude of the most unstable models; (d) the phase in the vertical of the most unstable modes. The instability is numerically calculated with four combinations of shear and stratification: 1, Uniform stratification and shear i.e. the Eady problem, (profiles b+d, solid lines). 2, Uniform shear, upper-ocean enhanced stratification (a+d, dashed); 3, Uniform stratification, upper ocean enhanced shear (b+c, dot-dashed); 4, Both stratification and shear enhanced in upper ocean (a+c, dotted). Case 2 (a+d, dashed) is really more like to an atmosphere with a stratosphere (see Fig. 6.21), and the amplitude of the disturbance falls off, rather unrealistically, in the upper ocean. Case 4 (a+c, dotted) is the most oceanically relevant.

necessary for instability. It is these properties that make the Eady problem, or more generally any baroclinic instability problem that is posed between two rigid lids, of more general applicability to the earth’s atmosphere than might be first thought: the high stratification above the tropopause and consequent decay of the instability is mimicked by the imposition of a rigid lid. (Of course, the $\beta$ effect is still absent in the Eady problem.)

In the ocean, the stratification is highest in the upper ocean where the shear is also strongest, and numerical calculations of the structure and growth rate of idealized profiles illustrated in Fig. 6.22. The solid curve shows the Eady problem, and the various dashed curves show the phase speeds, growth rates and phase with combinations of the profiles illustrated in panel (a). Much of the ocean is
characterized by having both a higher shear and a higher stratification in the upper 1 km or so, and this case is the one with the dotted line in Fig. 6.22. In this case the amplitude of the instability is also largely confined to the upper ocean, and unlike the Eady problem it does not arise through the interaction of edge waves at the top and bottom: the potential vorticity changes sign because of the interior variations due to the nonuniform shear, mainly in the upper ocean. Consistently, the phase of the baroclinic waves is nearly constant in the lower ocean in the two cases in which the shear is confined to the upper ocean. The ocean itself is still more complicated, because the most unstable regions near intense western boundary currents are often also barotropically unstable, and the mean flow itself may be meridionally directed. Nevertheless, the result that linear baroclinic instability is primarily an upper ocean phenomenon is quite robust. However, we will find in chapter 9 that the nonlinear evolution of baroclinic instability leads to eddies throughout the water column.

Notes

1. Thomson (1871), Helmholtz (1868). The more general case, considered by Thomson (later Lord Kelvin), allows the fluid’s density to vary.
2. See Drazin and Reid (1981) or Chandrasekhar (1961) for more detail.
3. This is Squire’s theorem, which states that for every three-dimensional disturbance to a plane-parallel flow there corresponds a more unstable two-dimensional one. This means there is no need to consider three dimensional effects to determine whether such a flow is unstable.
4. Rayleigh, Lord (1880).
5. First obtained by Rayleigh, Lord (1894).
6. The solution of Fig. 6.6 is obtained with a gridpoint code with 400 × 400 equally spaced gridpoints. This kind of problem is also well suited to contour dynamics approach, as in Dritschel (1989).
7. Rayleigh, Lord (1880) and, for the case with β, I. Kuo (1949).
10. Eady (1949), Charney (1947). Eric Eady (1915–1966) is best remembered today as the author of the iconic ‘Eady model’ of baroclinic instability, which describes the fundamental hydrodynamic instability mechanism that gives rise to weather systems. After an undergraduate education in mathematics he joined the U. K. Meteorological Office in 1937, becoming a forecaster and upper air analyst, in which capacity he served throughout the war. In 1946 he joined the Department of Mathematics at Imperial College, presenting his Ph.D. thesis in 1948 on ‘The theory of development in dynamical meteorology’, subsequently summarized in Tellus (Eady 1949). This work, masterly in its combination of austerity and relevance, provides a mathematical description of the essential aspects of cyclone development that stands to this day as a canonical model in the field. It also includes, rather obliquely, a derivation of the stratified quasi-geostrophic equations, albeit in a special form. The impact of the work was immediate and it led to visits to Bergen (in 1947 with J. Bjerknes), Stockholm (in 1952 with C.-G. Rossby) and Princeton (in 1953 with J. von Neumann.
and Charney). Eady followed his baroclinic instability work with prescient discussions of the general circulation of the atmosphere (Eady 1950, Eady and Sawyer 1951, Eady 1954). A perfectionist who sought to understand it all, Eady’s subsequent published output was small and he later turned his attention to fundamental problems in other areas of fluid mechanics, the dynamics of the sun and the earth’s interior, and biochemistry. He finally took his own life. There is little published about him, save for the obituary by Charnock et al. (1966).

Jule Charney (1917–1981) played a defining role in dynamical meteorology in the second half of the 20th century. He made seminal contributions in many areas including: the theory of baroclinic instability (Charney 1947); a systematic scaling theory for large-scale atmospheric motions and the derivation of the quasi-geostrophic equations (Charney 1948); a theory of stationary waves in the atmosphere (Charney and Eliassen 1949); the demonstration of the feasibility of numerical weather forecasts (Charney et al. 1950); planetary wave propagation into the stratosphere (Charney and Drazin 1961); a criterion for baroclinic instability (Charney and Stern 1962); a theory for hurricane growth (Charney and Eliassen 1964); and the concept of geostrophic turbulence (Charney 1971). His Ph.D. is from UCLA in 1946 and this, entitled ‘Dynamics of long waves in a baroclinic westerly current’, became his well-known 1947 paper. After this he spent a year at Chicago and another at Oslo, and in 1948 joined the Institute of Advanced Study in Princeton where he stayed until 1956 (and where Eady visited for a while). He spent most of his subsequent career at MIT, interspersed with many visits to Europe, especially Norway. For a more complete picture of Charney, see Lindzen et al. (1990) and a brief biography by N. Phillips available at http://www.nap.edu/readingroom/books/biomems/jcharney.html.

11 At least I find it so. My treatment of the Eady problem draws from unpublished notes by J. S. A. Green, as well as Eady (1949) itself.

12 If $c$ is purely real (and so the waves are neutral), then there exists the possibility that $\Lambda z - c = 0$, and the equation for $\Phi$ is

$$\frac{d^2\Phi}{dz^2} - \mu^2\Phi = C\delta(z - z_c), \quad z_c = c/\Lambda. \quad (6.182)$$

where $C$ is a constant. Because $z_c$ is continuous in the interval $[0,1]$ so is $c$, and these solutions have a continuous spectrum of eigenvalues. The associated eigenfunctions provide formal completeness to the normal modes, enabling any function to be represented as their superposition.

13 Our nondimensionalization of the two-layer system is such as to be in correspondence with that for the continuous system. Thus we choose $H$ to be the total depth of the domain. This choice produces growth rates and wavenumbers that are equivalent to those in the Eady problem.


15 After Charney (1947), in whose problem it appears, and Green (1960), who appreciated its importance.

16 Our numerical procedure is to assume a wavelike solution in the horizontal of the form $\tilde{\psi}\exp[i(kx + ly - \omega t)]$, and to finite difference the equations in the vertical. The resulting eigenvalue equations are solved by standard matrix methods, for each horizontal wavenumber. See Smith and Vallis (1998).

17 Gill et al. (1974) and Robinson and McWilliams (1974) were among the first to look at baroclinic instability in the ocean.
Further Reading


A standard text on hydrodynamic instability theory. It discusses nearly all the classic cases in a straightforward and clear fashion. It includes a more extensive discussion of the linear instability of parallel shear flow than is contained here, although the treatment of baroclinic instability is rather brief.


A classic text discussing many forms of instability but not, alas, baroclinic instability.

Pierrehumbert and Swanson (1995) review many aspects of baroclinic instability.

Problems

6.1 Derive the jump condition (6.29) without directly considering the motion of the interface. In particular, from the momentum equation along the interface show that

\[ \frac{\partial}{\partial y} \left( \frac{\tilde{\psi}}{U - c} \right) = - \frac{\tilde{p}}{(U - c)^2} \]  

and show that (6.29) follows. Be explicit about the conditions under which the right-hand side vanishes when integrated across the interface. (For help see Drazin and Reid 1981).

6.2 By applying the matching conditions (6.23) and (6.29) at \( y = \pm a \) to Rayleigh’s equation, explicitly derive the dispersion relationship (6.42).

6.3 Show that for very long waves, or as the shear layer becomes thinner, the growth rate given by (6.42) reduces to that of Kelvin-Helmholz instability of a vortex sheet.

6.4 Obtain the stability properties of the triangular jet, with a basic state velocity given by

\[ U(y) = \begin{cases} 
0 & \text{for } z \geq 1 \\
1 - |y| & \text{for } -1 \leq y \leq 1 \\
0 & \text{for } z \leq -1 
\end{cases} \]

(P6.2)

In particular, obtain the eigenfunctions and eigenvalues of the problem, and show that each eigenfunction is either even or odd. Perturbations with even \( \psi' \) are known as ‘sinuous modes’ and those with odd \( \psi' \) are ‘varicose modes’. Show that sinuous waves are unstable for sufficiently long wavelengths in the \( z \)-direction, but that all varicose modes are stable.

6.5 Consider the incompressible piecewise linear shear flow below:

The flow is two-dimensional, and \( A \) and \( B \) are constants with \( B > 0 \).

(a) Find the two normal mode frequencies as a function of zonal wavenumber \( k \).

(b) Find the stability boundaries in terms of \( k \) and \( A \) and provide a physical interpretation. If \( A = B \) is the flow stable or unstable? Why?
(If the algebra defeats you, explain carefully the method for doing the problem.)

6.6 Show numerically or analytically that, in the Eady problem:
   (a) Instability occurs for \( \mu < 2.399 \).
   (b) The wavenumber at which the instability is greatest is \( \mu = 1.61 \).
   (c) The nondimensional growth rate at that wavenumber is 0.31.

6.7 ♦ Consider the vertical modes of continuously stratified problems:
   (a) When solving the continuous form of the eigenvalue stability problem (as in the Eady problem, for example) the differential equation typically seems to have just one pair of eigenvalues. However, if the equation is solved on a vertical grid with \( N \) levels, the resulting difference equation has \( N \) roots. Does this mean that \( N - 2 \) roots are spurious, and if so how might the ‘correct’ eigenvalues be identified? Alternatively, are there corresponding additional roots in the continuously stratified problem?
   (b) The \( N \)-level problem is equivalent to a physically realizable \( N \)-layer system, in which there are presumably \( N \) physically meaningful eigenvalues. As \( N \) becomes large, with the density differences and thicknesses of each layer chosen to become smaller in a consistent way, the equations describing the layered system presumably converge to those describing the continuous system, yet there are \( N \) eigenvalues in the former. What is the physical nature of the eigenvalues in the layered system, and how do they relate to those of the continuous system?

6.8 Show, using the two-layer model (or otherwise) that the presence of \( \beta \) reduces the efficiency of baroclinic instability. For example, show that it makes the meridional velocity slightly out of phase with the temperature.

6.9 ♦ Consider the baroclinic instability problem with a discontinuity in the stratification, but a uniform shear. For example, suppose the shear is uniform for \( z \in (0,1) \) with an abrupt change in stratification at \( z = 0.5 \). How does the amplitude of the instability vary on either side of the discontinuity? Your answer may be an analytical or a numerical calculation, or both.