Topics in fluid mechanics

A. C. Fowler
Mathematical Institute, Oxford University

October 30, 2018

Contents

Preface

These notes have been produced to accompany the section C course 'Topics in fluid mechanics', introduced into the Oxford curriculum for the first time in Michaelmas Term 2007. The aim of the course is to show how fluid mechanics is used in real applications. The notes describe five topics of current interest, and the course treats each of these in turn. The separate chapters of the notes follow the scheme of the lectures. The present edition of these notes should be treated as a draft; they have been cobbled together from other sources (of mine), and are not yet fully edited or indeed finished.

A. C. Fowler

Oxford, October 30, 2018

Chapter 1

Thin film flows

1.1 Lubrication theory

Lubrication theory refers to a class of approximations of the Navier–Stokes equations which are based on a large *aspect ratio* of the flow. The aspect ratio is the ratio of two different directional length scales of the flow, as for example the depth and the width. Typical examples of flows where the aspect ratio is large (or small, depending on which length is in the numerator) are lakes, rivers, atmospheric winds, waterfalls, lava flows, and in an industrial setting, oil flows in bearings (whence the term lubrication theory). Lubrication theory forms a basic constituent of a viscous flow course and will not be dwelt on here.

In brief the Navier–Stokes equations for an incompressible take the form

$$\begin{aligned} \boldsymbol{\nabla}.\mathbf{u} &= 0, \\ \rho[\mathbf{u}_t + (\mathbf{u}.\boldsymbol{\nabla})\mathbf{u}] &= -\boldsymbol{\nabla}p + \mu\nabla^2\mathbf{u}, \end{aligned} \tag{1.1}$$

at least in Cartesian coordinates. It should be recalled that the actual definition of $\nabla^2 \equiv \boldsymbol{\nabla}\,\boldsymbol{\nabla}. - \boldsymbol{\nabla}\times\boldsymbol{\nabla}\times$, and the components of $\nabla^2\mathbf{u} = \nabla^2 u_i\mathbf{e}_i$ (we use the summation convention) is only applicable in Cartesian coordinates. For other systems, one can for example consult the appendix in Batchelor (1967).

We begin by non-dimensionalising the equations by choosing scales

$$\mathbf{x} \sim l, \quad t \sim \frac{l}{U}, \quad \mathbf{u} \sim U, \quad p - p_a \sim \frac{\mu U}{l}; \tag{1.2}$$

this is the usual way to scale the equations, except that we have chosen to balance the pressure with the viscous terms. The pressure p_a is an ambient pressure, commonly atmospheric pressure. The resulting dimensionless equations are

$$\begin{aligned} \boldsymbol{\nabla}.\mathbf{u} &= 0, \\ Re\,\dot{\mathbf{u}} \equiv Re\,[\mathbf{u}_t + (\mathbf{u}.\boldsymbol{\nabla})\mathbf{u}] &= -\boldsymbol{\nabla}p + \nabla^2\mathbf{u}, \end{aligned} \tag{1.3}$$

where

$$Re = \frac{\rho U l}{\mu} \tag{1.4}$$

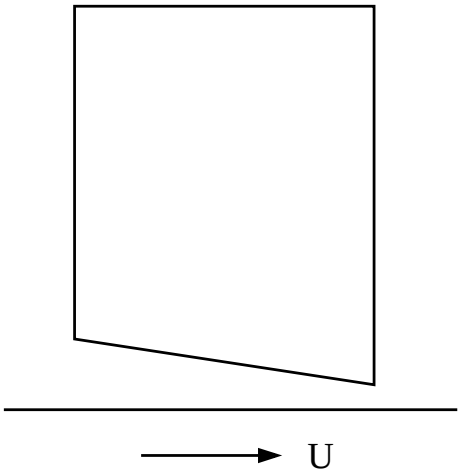

Figure 1.1: A slider bearing.

is the Reynolds number; the overdot denotes the material derivative. For $Re \ll 1$ we have Stokes flow, where the inertial terms can be neglected, and for $Re \gg 1$, boundary layers generally occur (and the pressure would be rescaled to balance the inertia terms, thus $p \sim Re$).

Lubrication theory describes a situation where the geometry of the flow allows the neglect of the inertial terms, even if the Reynolds number is not small. Suppose for example that l measures the extent of the flow in the x direction, but the fluid thickness in the (say) z direction is small. A simple example is the slider bearing, shown in figure 1.1, in which the fluid is confined between two surfaces, which we might take to be $z = 0$ and $z = h(x)$, and one of the surfaces moves at speed U relative to the other. To be specific, we assume a two-dimensional flow in which the coordinates are (x, z), the velocity components are (u, w), the bearing ($z = h$) is of finite length l and lies above a flat surface $z = 0$ which moves at speed U; the bearing is open to the atmosphere at each end, and the gap width $h \sim d \ll l$. We define the small parameter

$$\varepsilon = \frac{d}{l}, \tag{1.5}$$

so that in non-dimensional terms, the bearing is at $z = \varepsilon h(x)$ (where we scaled the dimensional h with d, so that the dimensionless h is $O(1)$). It is then appropriate to rescale the variables as follows:

$$z \sim \varepsilon, \quad w \sim \varepsilon, \quad p \sim \frac{1}{\varepsilon^2}, \tag{1.6}$$

and the equations then take the form

$$\begin{aligned} u_x + w_z &= 0, \\ \varepsilon^2 Re\, \dot{u} &= -p_x + u_{zz} + \varepsilon^2 u_{xx}, \\ \varepsilon^4 Re\, \dot{w} &= -p_z + \varepsilon^2 (w_{zz} + \varepsilon^2 w_{xx}), \end{aligned} \tag{1.7}$$

with boundary conditions

$$\begin{aligned} u = 1, w = 0 \quad &\text{at} \quad z = 0, \\ u = w = 0 \quad &\text{at} \quad z = h, \\ p = 0 \quad &\text{at} \quad x = 0, 1. \end{aligned} \tag{1.8}$$

At leading order we then have $p = p(x, t)$, and thus, integrating, we obtain

$$u = 1 - \frac{z}{h} - \tfrac{1}{2} p_x (hz - z^2). \tag{1.9}$$

The final part of the solution comes from integrating the mass conservation equation from $z = 0$ to $z = h$. This gives

$$0 = -[w]_0^h = -\int_0^h w_z \, dz = \int_0^h u_x \, dz = \frac{\partial}{\partial x} \int_0^h u \, dz, \tag{1.10}$$

where we can take the differentiation outside the integral because u is zero at $z = h$. In fact we can write down (1.10) directly since it is an expression of conservation of mass across the layer; and this applies more generally, even if the base is not flat, and indeed even if both surfaces depend on time, and the result can be extended to three dimensions; see question 1.2. Calculating the flux from (1.9), we obtain

$$\int_b^s u \, dz = \tfrac{1}{2} h - \tfrac{1}{12} h^3 p_x = K \tag{1.11}$$

is constant. Given h, the solution for p can be found as a quadrature, and is

$$p = 6 \left[f_2(x) - \frac{f_2(1) f_3(x)}{f_3(1)} \right], \quad f_n(x) = \int_0^x \frac{dx}{h^n}. \tag{1.12}$$

In three dimensions, exactly the same procedure leads to the equation

$$\tfrac{1}{12} \boldsymbol{\nabla}_H . (h^3 \boldsymbol{\nabla}_H p) = \tfrac{1}{2} h_x, \tag{1.13}$$

where the plate flow direction is taken along the x axis; derivation of this is left as an exercise.

1.2 Droplet dynamics

When one of the surfaces is a free surface (meaning it is free to deform), such as a droplet of liquid resting on a surface, or a rivulet flowing down a window pane, there are two differences which must be accounted for in formulating the problem. One is that the free surface is usually a material surface, so that a kinematic condition is appropriate. In three dimensions, this takes the form

$$w = s_t + u s_x + v s_y - a. \tag{1.14}$$

Here, $z = s$ is the free surface, and (u, v, w) is the velocity; the term a is normally absent, but a non-zero value describes surface accumulation (which might for example be due to condensation); if $a < 0$ it describes ablation due for example to evaporation.

The other difference is that the boundary conditions at the free surface are generally not ones of prescribed velocity but of prescribed stress. In the common case of a droplet of liquid with air above, these conditions take the form

$$\sigma_{nn} = -p_a, \quad \sigma_{nt} = 0, \tag{1.15}$$

representing the fact that the atmosphere exerts a constant pressure on the surface, and no shear stress. To unravel these conditions, we will consider the case of a two-dimensional incompressible flow. In this case, the components of the stress tensor are

$$\sigma_{11} = -p + \tau_1, \quad \sigma_{13} = \sigma_{31} = \tau_3, \quad \sigma_{33} = -p - \tau_1, \tag{1.16}$$

where

$$\tau_1 = 2\mu u_x, \quad \tau_3 = \mu(u_z + w_x), \tag{1.17}$$

and then with

$$\mathbf{n} = \frac{(-s_x, 1)}{(1 + s_x^2)^{1/2}}, \quad \mathbf{t} = \frac{(1, s_x)}{(1 + s_x^2)^{1/2}}, \tag{1.18}$$

we have

$$\begin{aligned} \sigma_{nn} = \sigma_{ij} n_i n_j &= -p - \frac{[\tau_1(1 - s_x^2) + 2\tau_3 s_x]}{1 + s_x^2}, \\ \sigma_{nt} = \sigma_{ij} n_i t_j &= \frac{[\tau_3(1 - s_x^2) - 2\tau_1 s_x]}{1 + s_x^2}. \end{aligned} \tag{1.19}$$

The dimensionless equations are virtually the same, as we initially scale $p - p_a, \tau_1$ and τ_3 with $\mu U/l$, and then when the rescaling in (1.6) is done (note that consequently we rescale $\tau_3 \sim 1/\varepsilon$), the surface boundary conditions become

$$\begin{aligned} p + \frac{\varepsilon^2[\tau_1(1 - \varepsilon^2 s_x^2) + 2\tau_3 s_x]}{1 + \varepsilon^2 s_x^2} &= 0, \\ \tau_3(1 - \varepsilon^2 s_x^2) - 2\varepsilon^2 \tau_1 s_x &= 0, \end{aligned} \tag{1.20}$$

where

$$\tau_1 = 2u_x, \quad \tau_3 = u_z + \varepsilon^2 w_x. \tag{1.21}$$

Putting $\varepsilon = 0$, we thus obtain the leading order conditions

$$p = \tau_3 = 0 \quad \text{on} \quad z = s. \tag{1.22}$$

We can then integrate $u_{zz} = p_x$, assuming also a no slip base at $z = b$, to obtain an expression for the flux

$$\int_b^s u \, dz = -\tfrac{1}{3} h^3 p_x, \tag{1.23}$$

and the conservation of mass equation then integrates (see question 1.2) to give the evolution equation for $h = s - b$ in the form

$$h_t = \tfrac{1}{3} \frac{\partial}{\partial x}[h^3 p_x]. \tag{1.24}$$

1.2.1 Gravity

The astute reader will notice that something is missing. Unlike the slider bearing, nothing is driving the flow! Indeed, since $p = p(x,t)$ and $p = 0$ at $z = s$, $p = 0$ everywhere. Related to this is the fact that there is nothing to determine the velocity scale U. Commonly such droplet flows are driven by gravity. If we include gravity in the z momentum equation, then it takes the dimensional form $\ldots = -p_z - \rho g \ldots$, and since in the rescaled model all the other terms are negligible, the pressure will be hydrostatic, $p \approx p_a + \rho g(s - z)$, and this gives a natural scale for $p - p_a \sim \rho g d$, and equating this with the eventual pressure scale $\mu U l/d^2$ determines the velocity scale as

$$U = \frac{\rho g d^3}{\mu l}. \tag{1.25}$$

The dimensionless pressure then becomes $p = s - z$, so that $p_x = s_x$, and (1.24) now takes the form of a nonlinear diffusion equation,

$$h_t = \tfrac{1}{3}\frac{\partial}{\partial x}[h^3 s_x]. \tag{1.26}$$

One might wonder how the length scales l and d should be chosen; the answer to this, at least if the base is flat, is that it can be taken from the initial condition for s. The reason for this is that, since (1.26) is a diffusion equation, the drop will simply continue to spread out: there is no natural length scale in the model. Associated with this is the consequent fact that for an initial concentration of liquid at the origin (again on a flat base), the solution takes the form of a similarity solution (see question 1.5). On the other hand, if b is variable, then it provides a natural length scale. Indeed, for a basin shaped b (for example x^2, dimensionlessly), the initial volume (or cross-sectional area) determines the eventual steady state as a lake with s constant, and both d and l prescribed.

1.2.2 Surface tension

Another way in which a natural length scale can occur in the model is through the introduction of surface tension at the interface. Let us digress for a moment to consider how surface tension arises. Surface tension is a property of interfaces, whereby they have an apparent strength. This is most simply manifested by the ability of small objects which are themselves heavier than water to float on the interface. The experiment is relatively easily done using a paper clip, and certain insects (water striders) have the ability to stay on the surface of a pond.

The simplest way to think about surface tension is mechanically. The interface between two fluids has an associated tension, such that if one draws a line in the interface of length l, then there is a force of magnitude γl which acts along this line: γ is the surface tension, and is a force per unit length. The presence of a surface tension causes an imbalance in the normal stress across the interface, as is indicated in figure 1.2, which also provides a means of calculating it. Taking ds as a short

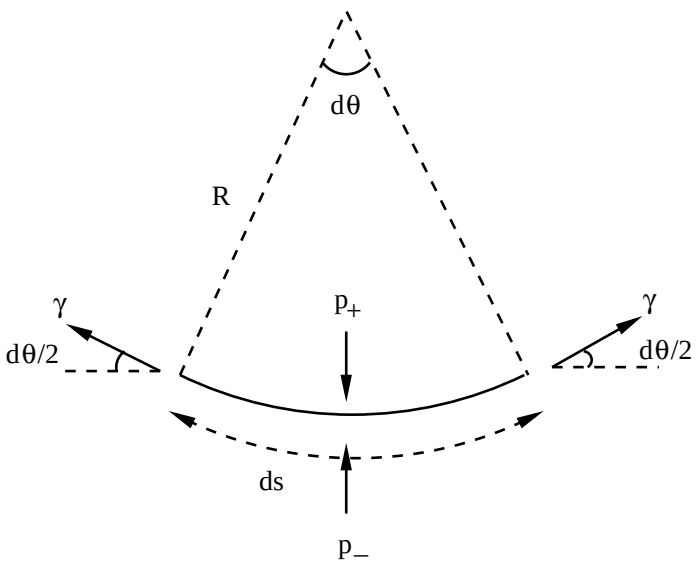

Figure 1.2: The simple mechanical interpretation of surface tension.

line segment in an interface subtending an angle $d\theta$ at its centre of curvature, a force balance normal to the interface leads to the condition

$$p_+ - p_- = \frac{\gamma}{R}, \tag{1.27}$$

where

$$R = \frac{ds}{d\theta} \tag{1.28}$$

is the *radius of curvature*, and its inverse $1/R$ is the curvature.

For a two-dimensional surface, the curvature is described by two *principal radii of curvature* R_1 and R_2, the mean curvature is defined by

$$\kappa = \tfrac{1}{2}\left(\frac{1}{R_1} + \frac{1}{R_2}\right), \tag{1.29}$$

and the pressure jump condition is

$$p_+ - p_- = 2\gamma\kappa = \gamma\left(\frac{1}{R_1} + \frac{1}{R_2}\right), \tag{1.30}$$

although this is not much use to us unless we have a way of calculating the curvature of a surface. This leads us off into the subject of differential geometry, and we do not want to go there. A better way lies along the following path.

The sceptical reader will in any case wonder what this surface tension actually is. It manifests itself as a force, but along a *line*? And what is its physical origin? The answer to this question veers towards the philosophical. We think we understand force, after all it pops up in Newton's second law, but how do we measure it? Pressure, for example, we conceive of as being due to the collision of molecules with a surface,

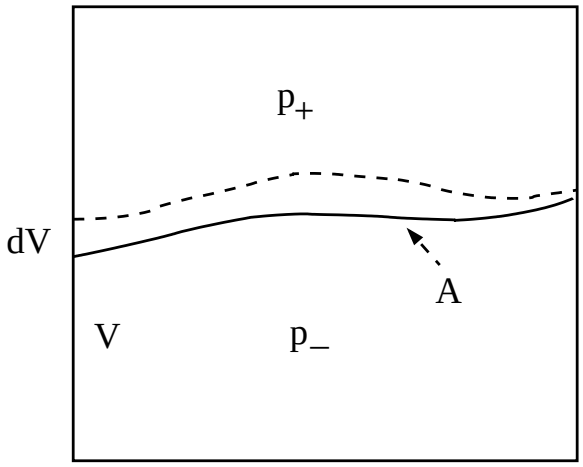

Figure 1.3: The energetic basis of surface tension.

and the measure of the force they exert is due to the momentum exchange at the surface. We pull on a rope, exerting a force, but the measure of the force is in the extension of the rope via Hooke's law. Force is apparently something we measure via its effect on momentum exchange, or on mechanical displacement; we can actually define force through these laws.

The more basic quantity is energy, which has a direct interpretation, whether as kinetic energy or internal energy (the vibration of molecules). And in fact Newton's second law for a particle is equivalent to the statement that the rate of change of energy is equal to the rate of doing work, and this might be taken as the fundamental law.

The meaning of surface tension actually arises through the property of an interface, which has a surface energy γ with units of energy per unit area. The interfacial condition then arises through the (thermodynamic) statement that in equilibrium the energy of the system is minimised.

To be specific, consider the situation in figure 1.3, where two fluids at pressures p_- and p_+ are separated by an interface with area A. Consider a displacement of the interface causing a change of volume dV as shown. Evidently the work done on the upper fluid is $p_+\,dV$, which is thus its change of energy, and correspondingly the change for the lower fluid is $-p_-\,dV$. If the change of interfacial surface area is dA, then the total change of energy[1] is

$$dF = (p_+ - p_-)\,dV + \gamma\,dA, \tag{1.31}$$

and at equilibrium this must be zero (since F is minimised). The equilibrium interfacial boundary condition is therefore

$$p_+ - p_- = -\gamma\frac{\partial A}{\partial V}, \tag{1.32}$$

[1]This energy is the *Helmholtz free energy*.

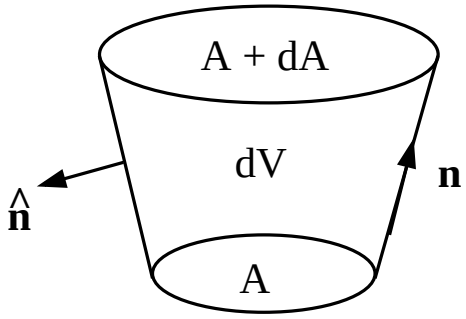

Figure 1.4: Calculation of $\dfrac{\partial A}{\partial V}$.

which, it turns out, is equivalent to (1.30).

Computation of $\dfrac{\partial A}{\partial V}$ can be done as follows. We consider a displacement of the interface as shown in figure 1.4. An element of surface A is displaced to $A + dA$, and we can form a connecting volume dV such that the normal $\mathbf{n}$ to the interface is always parallel to the connecting surface between the end faces A and $A + dA$. We need to distinguish between the normal $\hat{\mathbf{n}}$ to the surface of the connecting volume and the normal to the interfacial surface. Evidently we have $\mathbf{n} = \hat{\mathbf{n}}$ at the end faces, but $\mathbf{n}.\hat{\mathbf{n}} = 0$ on the connecting cylindrical surface.

Applying the divergence theorem, we see that the change in area is

$$dA = \int_{\partial(dV)} \mathbf{n}.\hat{\mathbf{n}}\, dS = \int_{dV} \boldsymbol{\nabla}.\,\mathbf{n}\, dV, \tag{1.33}$$

and thus

$$\frac{\partial A}{\partial V} = \boldsymbol{\nabla}.\,\mathbf{n}. \tag{1.34}$$

For example, if the interface is represented as $z = s(x, y, t)$, then

$$\boldsymbol{\nabla}.\,\mathbf{n} = -\boldsymbol{\nabla}.\left[\frac{\boldsymbol{\nabla} s}{(1 + |\boldsymbol{\nabla} s|^2)^{1/2}}\right], \tag{1.35}$$

where on the right hand side $\boldsymbol{\nabla} = \boldsymbol{\nabla}_H = \left(\dfrac{\partial}{\partial x}, \dfrac{\partial}{\partial y}\right)$, and for small interfacial displacement, this may be linearised to obtain

$$2\kappa = -\frac{\partial A}{\partial V} = -\boldsymbol{\nabla}.\,\mathbf{n} = \boldsymbol{\nabla}.\left[\frac{\boldsymbol{\nabla} s}{(1 + |\boldsymbol{\nabla} s|^2)^{1/2}}\right] \approx \nabla^2 s. \tag{1.36}$$

1.2.3 The capillary droplet

Now we use this in the droplet equation. Again we restrict attention to two-dimensional droplets. For three-dimensional droplets, see question 1.6. The surface boundary condition is now approximately $p - p_a = -\gamma s_{xx}$, and non-dimensionally

$$p = -\frac{1}{B} s_{xx} \quad \text{on} \quad z = s, \tag{1.37}$$

where B (commonly also written Bo) is the Bond number, given by

$$B = \frac{\rho g l^2}{\gamma}. \tag{1.38}$$

This gives a natural length scale for the droplet, by choosing $B = 1$, thus

$$l = \left(\frac{\gamma}{\rho g}\right)^{1/2}; \tag{1.39}$$

in this case the dimensionless pressure is $p = s - z - s_{xx}$, and thus mass conservation leads to

$$h_t = \tfrac{1}{3}\frac{\partial}{\partial x}\left[h^3(s_x - s_{xxx})\right], \tag{1.40}$$

and the surface tension term acts as a further stabilising term.[2]

Surface tension acts to limit the spread of a droplet. Indeed there is a steady state of (1.40) which is easily found. Suppose the base is flat, so $s = h$. We prescribe the cross-sectional area of the drop, A. In dimensionless terms, we thus require

$$\int h\,dx = 2\alpha = \left(\frac{\rho g}{\gamma}\right)^{1/2}\frac{A}{d}. \tag{1.41}$$

Let us choose d so that the maximum depth is one (note that the value of d remains to be determined). We can suppose that the drop is symmetric about the origin, and that its dimensionless half-width is λ, also to be determined. Thus

$$h(\pm\lambda) = 0, \quad h(0) = 1, \tag{1.42}$$

as well as (1.41), and both α and λ are to be determined.

A further condition is necessary at the margins. This is the prescription of a contact angle, which can be construed as arising through a balance of the surface tension forces at the three interfaces at the contact line: gas/liquid, liquid/solid, and solid/gas. All three interfaces have a surface energy, and minimisation of this corresponds to prescription of a contact angle. Specifically, if θ is the angle between the gas/liquid and liquid/solid interfaces, then resolution of the surface tension tangential to the wall leads to

$$\gamma_{\mathrm{SL}} + \gamma\cos\theta = \gamma_{\mathrm{SG}}, \tag{1.43}$$

where γ_{SL} is the solid/liquid surface energy, and γ_{SG} is the solid/gas surface energy. Defining $S = l\tan\theta/d$, this implies that

$$h_x = \mp S \quad \text{at} \quad x = \pm\lambda. \tag{1.44}$$

[2]This can be seen by considering small perturbations about a uniform solution $h = s = 1$ (with a flat base), for which the linearised equation has normal mode solutions $\propto \exp(\sigma t + ikx)$, with $\sigma = -\frac{1}{3}(k^2 + k^4)$.

The steady state of (1.40) is easily found. The flux is zero, so $h_x - h_{xxx}$ is zero, and integration of this leads to

$$h = 1 - \left(\frac{\cosh x - 1}{\cosh \lambda - 1} \right), \tag{1.45}$$

and then (1.41) and (1.44) yield

$$\alpha = \frac{\lambda \cosh \lambda - \sinh \lambda}{\cosh \lambda - 1}, \quad \frac{\sinh \lambda}{\cosh \lambda - 1} = S. \tag{1.46}$$

$S(\lambda)$ is a monotonically decreasing function of λ (why?), and tends to one as $\lambda \to \infty$, and therefore the second relation determines λ providing $S > 1$. It seems there is a problem if $S < 1$, but this is illusory since both α and S depend on the unknown d, so it is best to solve

$$\frac{\alpha}{S} = \frac{A}{2l^2 \tan \theta} = \frac{\lambda \cosh \lambda - \sinh \lambda}{\sinh \lambda}; \tag{1.47}$$

the right hand side increases monotonically from 0 to ∞ as λ increases, and therefore provides a unique solution for λ for any values of A and θ; d is then determined by either expression in (1.46).

It is of interest to see when the assumption $d \ll l$ is then valid. From (1.46),

$$\varepsilon = \tan \theta \left(\frac{\cosh \lambda - 1}{\sinh \lambda} \right). \tag{1.48}$$

The expression in λ increases monotonically from 0 to 1 as λ increases. Thus $\varepsilon \ll 1$ if either $\theta \ll 1$, or (if $\tan \theta \sim O(1)$) $\lambda \ll 1$. From (1.47), this is the case provided $A \ll l^2$, i. e., $\dfrac{\rho g A}{\gamma} \ll 1$. For air and water, this implies $A \ll 7$ mm^2.

1.2.4 Advance and retreat

When a droplet is of finite extent, it is possible to describe the behaviour near the margins by a local expansion. Typically the surface approaches the base with local power law behaviour, and this depends on whether the droplet is advancing or retreating. Consider, for example, the gravity-driven droplet with an accumulation or ablation term:

$$h_t = \tfrac{1}{3} \left(h^3 h_x \right)_x + a, \tag{1.49}$$

where $a > 0$ for accumulation, and $a < 0$ for ablation. (1.49) represents a simple model for the motion of an ice sheet such as Antarctica, where $a > 0$ represents accumulation due to snowfall. If we suppose that near the margin $x = x_s$ in a two-dimensional motion, $h \sim C(x_s - x)^\nu$, then a local expansion shows that if the front is advancing, $\dot{x}_s > 0$, then $\nu = \frac{1}{3}$ and $\dot{x}_s \sim \frac{1}{9} C^3$; in advance the front is therefore steep. On the other hand, if the front is retreating, then this can only occur if $a < 0$ (as is in fact obvious), and in that case $\nu = 1$ and $\dot{x}_s \sim -|a|/C$. The fact that the front slope is infinite in advance and finite in retreat is associated with 'waiting time' behaviour, which occurs when the front has to 'fatten up' before it can advance.

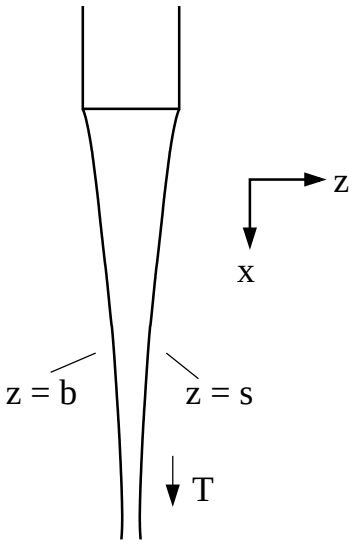

Figure 1.5: An elongational film flow.

1.3 Elongational flows

A different application of lubrication theory occurs in a falling sheet of fluid, such as occurs when a tap is switched on. At low velocities, the flow is continuous and laminar (though at very low flow rates it breaks up into droplets), and is also thin, but is distinguished from surface droplets or bearing flows by the fact that *both* surfaces of the fluid have zero stress acting on them.

To be specific, we consider the situation shown in figure 1.5. We consider flow from an orifice, and we take the flow to be two-dimensional, with the x direction in the direction of flow and z transverse to it. To begin with we ignore gravity and suppose that the flow is driven by an applied tension T (force per unit width in the y direction) at ∞; this is like drawing honey out of a jar with a spoon.

The basic equations are those as scaled in (1.3), and can be written in the form

$$\begin{aligned} u_x + w_z &= 0, \\ Re\,\dot{u} &= -p_x + \tau_{1x} + \tau_{3z}, \\ Re\,\dot{w} &= -p_z + \tau_{3x} - \tau_{1z}, \end{aligned} \tag{1.50}$$

where

$$\tau_1 = 2u_x, \quad \tau_3 = u_z + w_x. \tag{1.51}$$

If the two free surfaces are $z = s$ and $z = b$, then the boundary conditions on both surfaces are $\sigma_{nn} = \sigma_{nt} = 0$ (we subtract off the ambient pressure), or in other words $\sigma_{ni} = \sigma_{ij} n_j = 0$, and for $z = s$, this gives

$$\begin{aligned} (p - \tau_1)s_x + \tau_3 &= 0, \\ -\tau_3 s_x - p - \tau_1 &= 0. \end{aligned} \tag{1.52}$$

(These are actually equivalent to (1.19).)

Now we rescale the variables to account for the large aspect ratio. The difference with the earlier approach is that shear stresses are uniformly small, and so we also rescale τ_3 to be small. Thus we rescale the variables as

$$z \sim \varepsilon, \quad w \sim \varepsilon, \quad \tau_3 \sim \varepsilon, \tag{1.53}$$

and this leads to the rescaled equations

$$\begin{aligned} u_x + w_z &= 0, \\ Re\,\dot{u} &= -p_x + \tau_{1x} + \tau_{3z}, \\ \varepsilon^2 Re\,\dot{w} &= -p_z + \varepsilon^2\tau_{3x} - \tau_{1z}, \end{aligned} \tag{1.54}$$

where

$$\tau_1 = 2u_x, \quad \varepsilon^2\tau_3 = u_z + \varepsilon^2 w_x, \tag{1.55}$$

and on the free surfaces (e. g., $z = s$)

$$\begin{aligned} (p - \tau_1)s_x + \tau_3 &= 0, \\ -\varepsilon^2\tau_3 s_x - p - \tau_1 &= 0. \end{aligned} \tag{1.56}$$

At leading order, we have $u = u(x,t)$, $p + \tau_1 = 0$, $p = -2u_x$, whence we find

$$\tau_{3z} = Re\,\dot{u} - 4u_{xx}, \tag{1.57}$$

with

$$\tau_3 = 4u_x s_x \quad \text{on} \quad z = s, \; \tau_3 = 4u_x b_x \quad \text{on} \quad z = b,$$

and from these we deduce

$$\begin{aligned} Re\,h(u_t + uu_x) &= 4(hu_x)_x, \\ h_t + (hu)_x &= 0, \end{aligned} \tag{1.58}$$

where the second equation is derived as usual to represent conservation of mass. Note in this derivation that the inertial terms are not necessarily small; nevertheless the asymptotic procedure works in the usual way.

1.3.1 Steady flow

For a long filament such as that shown in figure 1.5, it is appropriate to prescribe inlet conditions, and these can be taken to be

$$h = u = 1 \quad \text{at} \quad x = 0, \tag{1.59}$$

by appropriate choice of U and d. In addition, we prescribe the force (per unit width in the third dimension) to be T, and this leads to

$$hu_x \to 1 \quad \text{as} \quad x \to \infty, \tag{1.60}$$

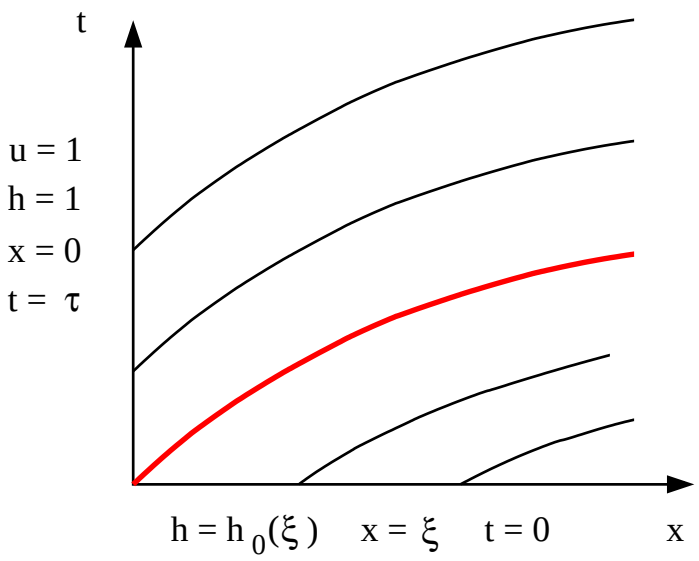

Figure 1.6: Characteristics for (1.58). The dividing characteristic from the origin is shown in red.

where the constant is set to one by choice of the length scale as

$$l = \frac{2\mu d U}{T}; \tag{1.61}$$

thus the aspect ratio is small ($d \ll l$) if $T \ll \mu U$.

If we consider a slow, steady flow in which the inertial terms can be ignored ($Re \to 0$), it is easy to solve the equations. We have $hu = 1$ and $hu_x = 1$, and thus

$$u = e^x, \quad h = e^{-x}. \tag{1.62}$$

As a matter of curiosity, one can actually solve the time-dependent problem (1.58), at least when $Re = 0$. We write the equations in the form

$$\begin{aligned} h_t + u h_x &= -1, \\ h u_x &= 1, \end{aligned} \tag{1.63}$$

with the boundary and initial conditions as shown in figure 1.6. The characteristic form of the first equation is

$$x_t = u[x(\xi, t), t], \quad h_t = -1, \tag{1.64}$$

where the partial derivatives are holding ξ fixed, i. e., we consider $x = x(\xi, t)$, $h = h(\xi, t)$. The dividing characteristic from the origin (which we define to be $t = t_d(x)$) divides the quadrant into two regions, in which the initial data is parameterised differently. For the lower region $t < t_d(x)$, we have

$$h = h_0(\xi) - t. \tag{1.65}$$

We take the first equation in (1.64), and differentiate with respect to ξ. Using the definition of u_x from (1.63), we find

$$x_{\xi t} = \frac{x_\xi}{h_0(\xi) - t}. \tag{1.66}$$

We can integrate this with respect to t, *holding* ξ *constant*, that is, the integral with respect to t is along a characteristic. It follows that

$$x_\xi = \frac{h_0(\xi)}{h_0(\xi) - t}, \tag{1.67}$$

in which we have applied the initial condition $x_\xi = 1$ at $t = 0$.

Next we integrate with respect to ξ *holding* t *constant*; since (1.67) only holds for $t < t_d(x)$, we integrate back to this, but note that this corresponds to the value $\xi = 0$; we then have

$$x = x_d(t) + \int_0^\xi \frac{h_0(s)\, ds}{h_0(s) - t}, \tag{1.68}$$

where x_d is the inverse of $t_d(x)$: to calculate this we need to solve for the upper region $t > t_d$.

To do this, we can proceed as above, but it is quicker to note that since the boundary conditions on $x = 0$ are constant, the solution is just the steady state solution (1.62). In particular, the characteristics are $e^{-x} = 1 - (t - \tau)$, and the dividing characteristic is that with $\tau = 0$, thus

$$t_d = 1 - e^{-x}, \quad x_d = -\ln(1 - t). \tag{1.69}$$

The solution in $t < t_d$ is thus

$$x = -\ln(1 - t) + \int_0^\xi \frac{h_0(s)\, ds}{h_0(s) - t}, \tag{1.70}$$

but the transient is of little interest since it disappears after finite time, $t = 1$. As a check, notice that if $h_0 = e^{-\xi}$, the steady state solution is regained everywhere.

The steady solution can be extended to positive Reynolds number. In steady flow we then find

$$u_x = Ku + \tfrac{1}{4} Re\, u^2 \tag{1.71}$$

for some constant K, and we see that there is no solution in which the filament can be drawn to ∞, as pinch-off always occurs. This is in keeping with experience.

1.3.2 Capillary effects

As for the shear-driven droplet flows, one can add gravity to the model, and this is done in question 1.3. In this section we consider the modification to the equations

which occurs when capillary effects are included. The normal stress conditions are modified to

$$\begin{aligned} -\sigma_{nn} &= -\frac{\gamma s_{xx}}{(1+s_x^2)^{3/2}} \quad \text{on} \quad z=s, \\ \sigma_{nn} &= -\frac{\gamma b_{xx}}{(1+b_x^2)^{3/2}} \quad \text{on} \quad z=b. \end{aligned} \tag{1.72}$$

The definition of σ_{nn} is in (1.19), and with the basic scaling (all lengths scaled with l, etc.) this leads to

$$-p - \frac{2\tau_3 s_x}{1+s_x^2} - \frac{\tau_1(1-s_x^2)}{1+s_x^2} = \frac{1}{Ca}\frac{\gamma s_{xx}}{(1+s_x^2)^{3/2}} \quad \text{on} \quad z=s, \tag{1.73}$$

where

$$Ca = \frac{\mu U}{\gamma} \tag{1.74}$$

is the capillary number; a similar expression applies on $z=b$, with the opposite sign on the right hand side. When the equations are re-scaled ($z \sim \varepsilon$, etc.), then these take the approximate form

$$\begin{aligned} p+\tau_1 &\approx -\frac{1}{C}s_{xx} \quad \text{on} \quad z=s, \\ p+\tau_1 &\approx \frac{1}{C}b_{xx} \quad \text{on} \quad z=b, \end{aligned} \tag{1.75}$$

where we write

$$Ca = \varepsilon C. \tag{1.76}$$

Now the normal stress is constant across the filament, thus

$$p+\tau_1 \approx -\frac{1}{C}s_{xx} \tag{1.77}$$

everywhere, and this forces symmetry of the filament, $s_{xx} = -b_{xx}$. The rest of the derivation proceeds as before, except that (1.57) gains an extra term $-s_{xxx}/C$ on the right hand side; integrating this and applying the boundary conditions leads to the modification of (1.58) as (bearing in mind that $h = s - b$ and thus $h_{xx} = 2s_{xx}$)

$$\begin{aligned} h_t + (hu)_x &= 0, \\ Re\, h(u_t + uu_x) &= \frac{1}{2C}hh_{xxx} + 4(hu_x)_x. \end{aligned} \tag{1.78}$$

Steady flow

The extra derivatives for h require, apparently, two extra boundary conditions. If we suppose the pressure becomes atmospheric at ∞, then we might apply

$$h_{xx} \to 0 \quad \text{as} \quad x \to \infty. \tag{1.79}$$

Since this also implies $h_x \to 0$, it may be sufficient. On the other hand, if $h \to 0$ at ∞, the multiplication of the third derivative term by h may render an extra boundary condition unnecessary.

Again we can consider the steady state. Then $hu = 1$, and (1.78) has a first integral

$$K + \frac{Re}{h} = \frac{1}{2C}\left[hh_{xx} - \tfrac{1}{2}h_x^2\right] - \frac{4h_x}{h}, \tag{1.80}$$

where K is constant. Evidently there is no solution if $Re > 0$, as pinch-off must again occur. For the case of slow flow, taking $Re = 0$, we have $K = 4$ due to the far field stress condition, and

$$h^2 h_{xx} - \tfrac{1}{2}hh_x^2 - 8C(h_x + h) = 0. \tag{1.81}$$

We seek a solution of this with $h(0) = 1$ and $h(\infty) = 0$. Phase plane analysis shows that there is a unique such solution: see question 1.7.

1.4 Foam drainage

Exercises

1.1 A thin incompressible liquid film flows in two dimensions (x, z) between a solid base $z = 0$ where the horizontal (x) component of the velocity is $U(t)$, and may depend on time, and a stationary upper solid surface $z = h(x)$, where a no slip condition applies. The upper surface is of horizontal length l, and is open to the atmosphere at the ends. Write down the equations and boundary conditions describing the flow, and non-dimensionalise them assuming that $U(t) \sim U_0$. (You may neglect gravity.)

Assuming $\varepsilon = d/l$ is sufficiently small, where d is a measure of the gap width, rescale the variables suitably, and derive an approximate equation for the pressure p. Hence derive a formal solution if the block is of finite length l, and the pressure is atmospheric at each end, and obtain an expression involving integrals of powers of h for the horizontal fluid flux, $q(t) = \int_0^h u\,dz$.

1.2 A two-dimensional incompressible fluid flow is contained between two surfaces $z = b(x, t)$ and $z = s(x, t)$, on which kinematic conditions hold:

$$w = s_t + us_x \quad \text{at} \quad z = s,$$

$$w = b_t + ub_x \quad \text{at} \quad z = b.$$

By integrating the equation of conservation of mass, show that the fluid thickness $h = s - b$ satisfies the conservation law

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}\int_b^s u\,dx = 0.$$

Extend the result to three dimensions to show that

$$h_t + \boldsymbol{\nabla}_H . \left[\int_b^s \mathbf{u}_H \, dz \right] = 0,$$

where $\mathbf{u}_H = (u, v)$ is the horizontal velocity, and $\boldsymbol{\nabla}_H = \left(\dfrac{\partial}{\partial x}, \dfrac{\partial}{\partial y} \right)$ is the horizontal gradient operator.

1.3 An incompressible two-dimensional flow from a slit of width d falls vertically under gravity. Define *vertical* and *horizontal* coordinates x and z, with corresponding velocity components u and w. The stream is symmetric with free interfaces at $z = \pm s$, on which no stress conditions apply. Write down the equations and boundary conditions in terms of the deviatoric stress components $\tau_1 = \tau_{11} = -\tau_{33}$ and $\tau_3 = \tau_{13} = \tau_{31}$, and by scaling lengths with l, velocities with the inlet velocity U, and choosing suitable scales for time t and the pressure and stresses, show that the equations take the form

$$u_x + w_z = 0,$$

$$Re\, \dot{u} = -p_x + \tau_{1x} + \tau_{3z} + 1,$$

$$Re\, \dot{w} = -p_z + \tau_{3x} - \tau_{1z},$$

where you should define $\dot{u}$, the Reynolds number Re, and write down expressions for τ_1 and τ_3.

Now define $\varepsilon = \dfrac{d}{l}$, and assume it is small. Find a suitable rescaling of the equations, and show that the vertical momentum equation takes the approximate form

$$h[Re\, \dot{u} - 1] = 4(h u_x)_x,$$

where $u = u(x, t)$ and h is the stream width.

Show also that

$$h_t + (hu)_x = 0.$$

Explain why suitable boundary conditions are

$$h = u = 1 \quad \text{at} \quad x = 0, \quad h u_x \to 0 \quad \text{as} \quad x \to \infty.$$

Write down a single second order equation for u in steady flow. If $Re = 0$, find the solution.

If $Re > 0$, find a pair of first order equations for $v = \ln u$ and $w = v_x$. (*Note: w here is no longer the horizontal velocity.*) Show that $(\infty, 0)$ is a saddle point, and that a unique solution satisfying the boundary conditions exists. If $Re \gg 1$ (but still $\varepsilon^2 Re \ll 1$), show (by rescaling $w = W/Re$ and $x = Re\, X$) that the required trajectory hugs the W–nullcline, and thus show that in this case

$$u \approx \left(1 + \frac{2x}{Re} \right)^{1/2}.$$

1.4 A (two-dimensional) droplet rests on a rough surface $z = b$ and is subject to gravity g and surface tension γ. Write down the equations and boundary conditions which govern its motion, non-dimensionalise them, and assuming the depth at the summit d is much less than the half-width l, derive an approximate equation for the evolution in time of the depth h. Show that the horizontal velocity scale is

$$U = \frac{\rho g d^3}{\mu l},$$

and derive an approximate set of equations assuming

$$\varepsilon = \frac{d}{l} \ll 1, \quad F = \frac{U}{\sqrt{gd}} \ll 1.$$

Hence show that

$$h_t = \frac{\partial}{\partial x}\left[\tfrac{1}{3}h^3\left(s_x - \frac{1}{B}s_{xxx}\right)\right],$$

where you should define the Bond number B.

Find a steady state solution of this equation for the case of a flat base, assuming that the droplet area A and a contact angle $\theta = \varepsilon\phi$ are prescribed, with $\phi \sim O(1)$, and show that it is unique. Explain how the solution chooses the unknowns d and l.

1.5 A droplet of thickness h satisfies the equation

$$h_t = \frac{\partial}{\partial x}\left[\tfrac{1}{3}h^3 h_x\right].$$

Find a similarity solution of this equation which describes the spread of a drop of area one which is initially concentrated at the origin (i. e., $h(x, 0) = \delta(x)$).

1.6 Three-dimensional droplet

1.7 A film of fluid is drawn downwards under the action of a tensile force. A model for the dimensionless thickness h and dimensionless downwards velocity u of the film is

$$h_t + (hu)_x = 0,$$
$$Re\, h(u_t + uu_x) = \frac{1}{2C}hh_{xxx} + 4(hu_x)_x,$$

with

$$h = u = 1 \quad \text{on} \quad x = 0, \quad hu_x \to 1 \quad \text{as} \quad x \to \infty.$$

Show that a steady state solution in which $h \to 0$ as $x \to \infty$ can only occur if $Re = 0$. In that case, determine a second order differential equation satisfied by h, and by writing $h = \frac{1}{2}U^2$ and $V = U' = U_x$, write the equation as a pair of first order equations for U and V. Show that the origin is a (degenerate) saddle, and therefore show that a solution exists which satisfies the boundary conditions.

Chapter 2

Porous media

Groundwater is water which is stored in the soil and rock beneath the surface of the Earth. It forms a fundamental constituent reservoir of the hydrological system, and it is important because of its massive and long lived storage capacity. It is the resource which provides drinking and irrigation water for crops, and increasingly in recent decades it has become an unwilling recipient of toxic industrial and agricultural waste. For all these reasons, the movement of groundwater is an important subject of study.

Soil consists of very small grains of organic and inorganic matter, ranging in size from millimetres to microns. Differently sized particles have different names. Particularly, we distinguish clay particles (size < 2 microns) from silt particles (2–60 microns) and sand (60 microns to 1 mm). Coarser particles still are termed gravel.

Viewed at the large scale, soil thus forms a continuum which is granular at the small scale, and which contains a certain fraction of pore space, as shown in figure 2.1. The volume fraction of the soil (or sediment, or rock) which is occupied by the pore space (or void space, or voidage) is called the *porosity*, and is commonly denoted by the symbol ϕ; sometimes other symbols are used, for example n, as in chapter **??**.

As we described in chapter **??**, soils are formed by the weathering of rocks, and are specifically referred to as soils when they contain organic matter formed by the rotting of plants and animals. There are two main types of rock: igneous, formed by the crystallisation of molten lava, and sedimentary, formed by the cementation of sediments under conditions of great temperature and pressure as they are buried at depth.[1] Sedimentary rocks, such as sandstone, chalk, shale, thus have their porosity built in, because of the pre-existing granular structure. With increasing pressure, the grains are compacted, thus reducing their porosity, and eventually intergranular cements bond the grains into a rock. Sediment compaction is described in section 2.6.

Igneous rock tends to be porous also, for a different reason. It is typically the case for any rock that it is fractured. Most simply, rock at the surface of the Earth is

[1] There are also *metamorphic* rocks, which form from pre-existing rocks through chemical changes induced by burial at high temperatures and pressures; for example, marble is a metamorphic form of limestone.

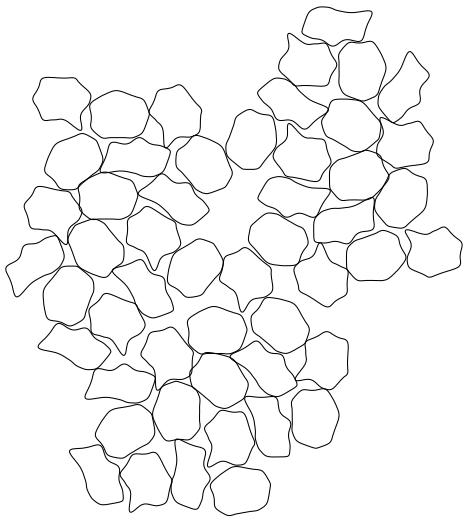

Figure 2.1: A granular porous medium.

subjected to enormous tectonic stresses, which cause folding and fracturing of rock. Thus, even if the rock *matrix* itself is not porous, there are commonly faults and fractures within the rock which act as channels through which fluids may flow, and which act on the large scale as an effective porosity. If the matrix is porous at the grain scale also, then one refers to the rock as having a dual porosity, and the corresponding flow models are called double porosity models.

In the subsurface, whether it be soil, underlying regolith, a sedimentary basin, or oceanic lithosphere, the pore space contains liquid. At sufficient depth, the pore space will be saturated with fluid, normally water. At greater depths, other fluids may be present. For example, oil may be found in the pore space of the rocks of sedimentary basins. In the near surface, both air and water will be present in the pore space, and this (unsaturated) region is called the unsaturated zone, or the vadose zone. The surface separating the two is called the piezometric surface, the phreatic surface, or more simply the water table. Commonly it lies tens of metres below the ground surface.

2.1 Darcy's law

Groundwater is fed by surface rainfall, and as with surface water it moves under a pressure gradient driven by the slope of the piezometric surface. In order to characterise the flow of a liquid in a porous medium, we must therefore relate the flow rate to the pressure gradient. An idealised case is to consider that the pores consist of uniform cylindrical tubes of radius a; initially we will suppose that these are all aligned in one direction. If a is small enough that the flow in the tubes is laminar (this will be the case if the associated Reynolds number is $\lesssim 1000$), then Poiseuille

flow in each tube leads to a volume flux in each tube of $q = \dfrac{\pi a^4}{8\mu}|\boldsymbol{\nabla} p|$, where μ is the liquid viscosity, and $\boldsymbol{\nabla} p$ is the pressure gradient along the tube. A more realistic porous medium is *isotropic*, which is to say that if the pores have this tubular shape, the tubules will be arranged randomly, and form an interconnected network. However, between nodes of this network, Poiseuille flow will still be appropriate, and an appropriate generalisation is to suppose that the volume flux vector is given by

$$\mathbf{q} \approx -\frac{a^4}{\mu X}\boldsymbol{\nabla} p, \tag{2.1}$$

where the approximation takes account of small interactions at the nodes; the numerical tortuosity factor $X \gtrsim 1$ takes some account of the arrangement of the pipes.

To relate this to macroscopic variables, and in particular the porosity ϕ, we observe that $\phi \sim a^2/d_p^2$, where d_p is a representative particle or grain size so that $\mathbf{q}/d_p^2 \sim -\left(\dfrac{\phi^2 d_p^2}{\mu X}\right)\boldsymbol{\nabla} p$. We define the volume flux per unit area (having units of velocity) as the discharge $\mathbf{u}$. Darcy's law then relates this to an applied pressure gradient by the relation

$$\mathbf{u} = -\frac{k}{\mu}\boldsymbol{\nabla} p, \tag{2.2}$$

where k is an empirically determined parameter called the *permeability*, having units of length squared. The discussion above suggests that we can write

$$k = \frac{d_p^2\phi^2}{X}; \tag{2.3}$$

the numerical factor X may typically be of the order of 10^3.

To check whether the pore flow is indeed laminar, we calculate the (particle) Reynolds number for the porous flow. If $\mathbf{v}$ is the (average) fluid velocity in the pore space, then

$$\mathbf{v} = \frac{\mathbf{u}}{\phi}; \tag{2.4}$$

If a is the pore radius, then we define a particle Reynolds number based on grain size as

$$Re_p = \frac{2\rho v a}{\mu} \sim \frac{\rho|\mathbf{u}|d_p}{\mu\sqrt{\phi}}, \tag{2.5}$$

since $\phi \sim a/d_p$. Suppose (2.3) gives the permeability, and we use the gravitational pressure gradient ρg to define (via Darcy's law) a velocity scale[2]; then

$$Re_p \sim \frac{\phi^{3/2}}{X}\left(\frac{\rho\sqrt{g d_p}\, d_p}{\mu}\right)^2 \sim 10[d_p]^3, \tag{2.6}$$

[2]This scale is thus the hydraulic conductivity, defined below in (2.9).

where $d_p = [d_p]$ mm, and using $\phi^{3/2}/X = 10^{-3}$, $g = 10$ m s^{-2}, $\mu/\rho = 10^{-6}$ m^2 s^{-2}. Thus the flow is laminar for $d < 5$ mm, corresponding to a gravel. Only for free flow through very coarse gravel could the flow become turbulent, but for water percolation in rocks and soils, we invariably have slow, laminar flow.

In other situations, and notably for forced gas stream flow in fluidised beds or in packed catalyst reactor beds, the flow can be rapid and turbulent. In this case, the Poiseuille flow balance $-\boldsymbol{\nabla} p = \mu \mathbf{u}/k$ can be replaced by the *Ergun equation*

$$-\boldsymbol{\nabla} p = \frac{\rho |\mathbf{u}| \mathbf{u}}{k'}; \tag{2.7}$$

more generally, the right hand side will a sum of the two (laminar and turbulent) interfacial resistances. The Ergun equation reflects the fact that turbulent flow in a pipe is resisted by *Reynolds stresses*, which are generated by the fluctuation of the inertial terms in the momentum equation. Just as for the laminar case, the parameter k', having units of length, depends both on the grain size d_p and on ϕ. Evidently, we will have

$$k' = d_p E(\phi), \tag{2.8}$$

with the numerical factor $E \to 0$ as $\phi \to 0$.

Hydraulic conductivity

Another measure of flow rate in porous soil or rock relates specifically to the passage of water through a porous medium under gravity. For free flow, the pressure gradient downwards due to gravity is just ρg, where ρ is the density of water and g is the gravitational acceleration; thus the water flux per unit area in this case is just

$$K = \frac{k \rho g}{\mu}, \tag{2.9}$$

and this quantity is called the *hydraulic conductivity*. It has units of velocity. A hydraulic conductivity of $K = 10^{-5}$ m s^{-1} (about 300 m y^{-1}) corresponds to a permeability of $k = 10^{-12}$ m^2, this latter unit also being called the *darcy*.

2.1.1 Homogenisation

The 'derivation' of Darcy's law can be carried out in a more formal way using the method of homogenisation. This is essentially an application of the method of multiple (space) scales to problems with microstructure. Usually (for analytic reasons) one assumes that the microstructure is periodic, although this is probably not strictly necessary (so long as local averages can be defined).

Consider the Stokes flow equations for a viscous fluid in a medium of macroscopic length l, subject to a pressure gradient of order $\Delta p/l$. If the microscopic (e. g., grain size) length scale is d_p, and $\varepsilon = d_p/l$, then if we scale velocity with $d_p^2 \Delta p/l\mu$

(appropriate for local Poiseuille-type flow), length with l, and pressure with Δp, the Navier-Stokes equations can be written in the dimensionless form

$$\begin{aligned} \boldsymbol{\nabla}.\mathbf{u} &= 0, \\ 0 &= -\boldsymbol{\nabla} p + \varepsilon^2 \nabla^2 \mathbf{u}, \end{aligned} \tag{2.10}$$

together with the no-slip boundary condition,

$$\mathbf{u} = 0 \text{ on } S : f(\mathbf{x}/\varepsilon) = 0, \tag{2.11}$$

where S is the interfacial surface. We put $\mathbf{x} = \varepsilon \boldsymbol{\xi}$ and seek solutions in the form

$$\begin{aligned} \mathbf{u} &= \mathbf{u}^{(0)}(\mathbf{x}, \boldsymbol{\xi}) + \varepsilon \mathbf{u}^{(1)}(\mathbf{x}, \boldsymbol{\xi}) \ldots \\ p &= p^{(0)}(\mathbf{x}, \boldsymbol{\xi}) + \varepsilon p^{(1)}(\mathbf{x}, \boldsymbol{\xi}) \ldots . \end{aligned} \tag{2.12}$$

Expanding the equations in powers of ε and equating terms leads to $p^{(0)} = p^{(0)}(\mathbf{x})$, and $\mathbf{u}^{(0)}$ satisfies

$$\begin{aligned} \boldsymbol{\nabla}_\xi.\mathbf{u}^{(0)} &= 0, \\ 0 &= -\boldsymbol{\nabla}_\xi p^{(1)} + \nabla_\xi^2 \mathbf{u}^{(0)} - \boldsymbol{\nabla}_x p^{(0)}, \end{aligned} \tag{2.13}$$

equivalent to Stokes' equations for $\mathbf{u}^{(0)}$ with a forcing term $-\boldsymbol{\nabla}_x p^{(0)}$. If $\mathbf{w}^j$ is the velocity field which (uniquely) solves

$$\begin{aligned} \boldsymbol{\nabla}_\xi.\mathbf{w}^j &= 0, \\ 0 &= -\boldsymbol{\nabla}_\xi P + \nabla_\xi^2 \mathbf{w}^j + \mathbf{e}_j, \end{aligned} \tag{2.14}$$

with periodic (in $\boldsymbol{\xi}$) boundary conditions and $\mathbf{u} = 0$ on $f(\boldsymbol{\xi}) = 0$, where $\mathbf{e}_j$ is the unit-vector in the ξ_j direction, then (since the equation is linear) we have (summing over j)[3]

$$\mathbf{u}^{(0)} = -\frac{\partial p^{(0)}}{\partial x_j} \mathbf{w}^j. \tag{2.15}$$

We define the average flux

$$\langle \mathbf{u} \rangle = \frac{1}{V} \int_V \mathbf{u}^{(0)} dV, \tag{2.16}$$

where V is the volume over which S is periodic.[4] Averaging (2.15) then gives

$$\langle \mathbf{u} \rangle = -\mathbf{k}^*.\boldsymbol{\nabla} p, \tag{2.17}$$

where the (dimensionless) permeability tensor is defined by

$$k_{ij}^* = \langle w_i^j \rangle. \tag{2.18}$$

[3]In other words, we employ the summation convention which states that summation is implied over repeated suffixes, see for example Jeffreys and Jeffreys (1953).

[4]Specifically, we take V to be the soil volume, but the integral is only over the pore space volume, where $\mathbf{u}$ is defined. In that case, the average $\langle \mathbf{u} \rangle$ is in fact the Darcy flux (i. e., volume fluid flux per unit area).

k (m^2)	material
10^{-8}	gravel
10^{-10}	sand
10^{-12}	fractured igneous rock
10^{-13}	sandstone
10^{-14}	silt
10^{-18}	clay
10^{-20}	granite

Table 2.1: Different grain size materials and their typical permeabilities.

Recollecting the scales for velocity, length and pressure, we find that the dimensional version of (2.17) is

$$\langle \mathbf{u} \rangle = -\frac{\mathbf{k}}{\mu}.\boldsymbol{\nabla} p, \tag{2.19}$$

where

$$\mathbf{k} = \mathbf{k}^* d_p^2, \tag{2.20}$$

so that $\mathbf{k}^*$ is the equivalent in homogenisation theory of the quantity ϕ^2/X in (2.3).

2.1.2 Empirical measures

While the validity of Darcy's law can be motivated theoretically, it ultimately relies on experimental measurements for its accuracy. The permeability k has dimensions of (length)^2, which as we have seen is related to the mean 'grain size'. If we write $k = d_p^2 C$, then the number C depends on the pore configuration. For a tubular network (in three dimensions), one finds $C \approx \phi^2/72\pi$ (as long as ϕ is relatively small). A different and often used relation is that of Carman and Kozeny, which applies to pseudo-spherical grains (for example sand grains); this is

$$C \approx \frac{\phi^3}{180(1-\phi)^2}. \tag{2.21}$$

The factor $(1-\phi)^2$ takes some account of the fact that as ϕ increases towards one, the resistance to motion becomes negligible. In fact, for media consisting of uncemented (i. e., separate) grains, there is a critical value of ϕ beyond which the medium as a whole will deform like a fluid. Depending on the grain size distribution, this value is about 0.5 to 0.6. When the medium deforms in this way, the description of the intergranular fluid flow can still be taken to be given by Darcy's law, but this now constitutes a particular choice of the interactive drag term in a two-phase flow model. At lower porosities, deformation can still occur, but it is elastic not viscous (on short time scales), and given by the theory of consolidation or compaction, which we discuss later.

In the case of soils or sediments, empirical power laws of the form

$$C \sim \phi^m \tag{2.22}$$

are often used, with much higher values of the exponent (e.g. $m = 8$). Such behaviour reflects the (chemically-derived) ability of clay-rich soils to retain a high fraction of water, thus making flow difficult. Table 2.1 gives typical values of the permeability of several common rock and soil types, ranging from coarse gravel and sand to finer silt and clay.

An explicit formula of Carman-Kozeny type for the turbulent Ergun equation expresses the 'turbulent' permeability k', defined in (2.7), as

$$k' = \frac{\phi^3 d_p}{175(1-\phi)}. \tag{2.23}$$

2.2 Basic groundwater flow

Darcy's equation is supplemented by an equation for the conservation of the fluid phase (or phases, for example in oil recovery, where these may be oil and water). For a single phase, this equation is of the simple conservation form

$$\frac{\partial}{\partial t}(\rho\phi) + \boldsymbol{\nabla}.(\rho\mathbf{u}) = 0, \tag{2.24}$$

supposing there are no sources or sinks within the medium. In this equation, ρ is the material density, that is, mass per unit volume *of the fluid*. A term ϕ is not present in the divergence term, since $\mathbf{u}$ has already been written as a volume flux (i.e., the ϕ has already been included in it: *cf.* (2.4)).

Eliminating $\mathbf{u}$, we have the parabolic equation

$$\frac{\partial}{\partial t}(\rho\phi) = \boldsymbol{\nabla}.\left[\frac{k}{\mu}\rho\boldsymbol{\nabla}p\right], \tag{2.25}$$

and we need a further equation of state (or two) to complete the model. The simplest assumption corresponds to incompressible groundwater flowing through a rigid porous medium. In this case, ρ and ϕ are constant, and the governing equation reduces (if also k is constant) to Laplace's equation

$$\nabla^2 p = 0. \tag{2.26}$$

This simple equation forms the basis for the following development. Before pursuing this, we briefly mention one variant, and that is when there is a compressible pore fluid (e. g., a gas) in a non-deformable medium. Then ϕ is constant (so k is constant), but ρ is determined by pressure and temperature. If we can ignore the effects of temperature, then we can assume $p = p(\rho)$ with $p'(\rho) > 0$, and

$$\rho_t = \frac{k}{\mu\phi}\boldsymbol{\nabla}.[\rho p'(\rho)\boldsymbol{\nabla}\rho], \tag{2.27}$$

which is a nonlinear diffusion equation for ρ, sometimes called the *porous medium equation*. If $p \propto \rho^\gamma$, $\gamma > 0$, this is degenerate when $\rho = 0$, and the solutions display the typical feature of finite spreading rate of compactly supported initial data.

2.2.1 Boundary conditions

The Laplace equation (2.26) in a domain D requires boundary data to be prescribed on the boundary ∂D of the spatial domain. Typical conditions which apply are a no flow through condition at an impermeable boundary, $\mathbf{u}.\mathbf{n} = 0$, whence

$$\frac{\partial p}{\partial n} = 0 \quad \text{on} \quad \partial D, \tag{2.28}$$

or a permeable surface condition

$$p = p_a \quad \text{on} \quad \partial D, \tag{2.29}$$

where for example p_a would be atmospheric pressure at the ground surface. Another example of such a condition would be the prescription of oceanic pressure at the interface with the oceanic crust.

A more common application of the condition (2.29) is in the consideration of flow in the saturated zone below the water table (which demarcates the upper limit of the saturated zone). At the water table, the pressure is in equilibrium with the air in the unsaturated zone, and (2.29) applies. The water table is a free surface, and an extra kinematic condition is prescribed to locate it. This condition says that the phreatic surface is also a material surface for the underlying groundwater flow, so that its velocity is equal to the average fluid velocity (*not* the flux): bearing in mind (2.4), we have

$$\frac{\partial F}{\partial t} + \frac{\mathbf{u}}{\phi}.\boldsymbol{\nabla} F = 0 \quad \text{on} \quad \partial D, \tag{2.30}$$

if the free surface ∂D is defined by $F(\mathbf{x}, t) = 0$.

2.2.2 Dupuit approximation

One of the principally obvious features of mature topography is that it is relatively flat. A slope of 0.1 is very steep, for example. As a consequence of this, it is typically also the case that gradients of the free groundwater (phreatic) surface are also small, and a consequence of this is that we can make an approximation to the equations of groundwater flow which is analogous to that used in shallow water theory or the lubrication approximation, i. e., we can take advantage of the large aspect ratio of the flow. This approximation is called the Dupuit, or Dupuit–Forchheimer, approximation.

To be specific, suppose that we have to solve

$$\nabla^2 p = 0 \quad \text{in} \quad 0 < z < h(x, y, t), \tag{2.31}$$

where z is the vertical coordinate, $z = h$ is the phreatic surface, and $z = 0$ is an impermeable basement. We let $\mathbf{u}$ denote the horizontal (vector) component of the Darcy flux, and w the vertical component. In addition, we now denote by $\boldsymbol{\nabla} = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$ the horizontal component of the gradient vector. The boundary conditions are then

$$\begin{aligned} p = 0, \quad \phi h_t + \mathbf{u}.\boldsymbol{\nabla} h = w \quad \text{on} \quad z = h, \\ \frac{\partial p}{\partial z} + \rho g = 0 \quad \text{on} \quad z = 0; \end{aligned} \tag{2.32}$$

here we take (gauge) pressure measured relative to atmospheric pressure. The condition at $z = 0$ is that of no normal flux, allowing for gravity.

Let us suppose that a horizontal length scale of relevance is l, and that the corresponding variation in h is of order d, thus

$$\varepsilon = \frac{d}{l} \tag{2.33}$$

is the size of the phreatic gradient, and is small. We non-dimensionalise the variables by scaling as follows:

$$\begin{aligned} x, y \sim l, \quad z \sim d, \quad p \sim \rho g d, \\ \mathbf{u} \sim \frac{k\rho g d}{\mu l}, \quad w \sim \frac{k\rho g d^2}{\mu l^2}, \quad t \sim \frac{\phi \mu l^2}{k\rho g d}. \end{aligned} \tag{2.34}$$

The choice of scales is motivated by the same ideas as lubrication theory. The pressure is nearly hydrostatic, and the flow is nearly horizontal.

The dimensionless equations are

$$\begin{aligned} \mathbf{u} = -\boldsymbol{\nabla} p, \quad \varepsilon^2 w = -(p_z + 1), \\ \boldsymbol{\nabla}.\,\mathbf{u} + w_z = 0, \end{aligned} \tag{2.35}$$

with

$$\begin{aligned} p_z = -1 \quad \text{on} \quad z = 0, \\ p = 0, \quad h_t = w + \boldsymbol{\nabla} p.\boldsymbol{\nabla} h \quad \text{on} \quad z = h. \end{aligned} \tag{2.36}$$

At leading order as $\varepsilon \to 0$, the pressure is hydrostatic:

$$p = h - z + O(\varepsilon^2). \tag{2.37}$$

More precisely, if we put

$$p = h - z + \varepsilon^2 p_1 + \ldots, \tag{2.38}$$

then (2.35) implies

$$p_{1zz} = -\nabla^2 h, \tag{2.39}$$

with boundary conditions, from (2.36),

$$p_{1z} = 0 \quad \text{on} \quad z = 0,$$
$$p_{1z} = -h_t + |\boldsymbol{\nabla} h|^2 \quad \text{on} \quad z = h. \tag{2.40}$$

Integrating (2.39) from $z = 0$ to $z = h$ thus yields the evolution equation for h in the form

$$h_t = \boldsymbol{\nabla}.[h\boldsymbol{\nabla} h], \tag{2.41}$$

which is a nonlinear diffusion equation of degenerate type when $h = 0$.

This is easily solved numerically, and there are various exact solutions which are indicated in the exercises. In particular, steady solutions are found by solving Laplace's equation for $\frac{1}{2}h^2$, and there are various kinds of similarity solution. (2.41) is a second order equation requiring two boundary conditions. A typical situation in a river catchment is where there is drainage from a watershed to a river. A suitable problem in two dimensions is

$$h_t = (hh_x)_x + r, \tag{2.42}$$

where the source term r represents recharge due to rainfall. It is given by

$$r = \frac{r_D}{\varepsilon^2 K}, \tag{2.43}$$

where r_D is the rainfall rate and $K = k\rho g/\mu$ is the hydraulic conductivity. At the divide (say, $x = 0$), we have $h_x = 0$, whereas at the river (say, $x = 1$), the elevation is prescribed, $h = 1$ for example. The steady solution is

$$h = \left[1 + r - rx^2\right]^{1/2}, \tag{2.44}$$

and perturbations to this decay exponentially. If this value of the elevation of the water table exceeds that of the land surface, then a seepage face occurs, where water seeps from below and flows over the surface. This can sometimes be seen in steep mountainous terrain, or on beaches, when the tide is going out.

The Dupuit approximation is not uniformly valid at $x = 1$, where conditions of symmetry at the base of a valley would imply that $u = 0$, and thus $p_x = 0$. There is therefore a boundary layer near $x = 1$, where we rescale the variables by writing

$$x = 1 - \varepsilon X, \quad w = \frac{W}{\varepsilon}, \quad h = 1 + \varepsilon H, \quad p = 1 - z + \varepsilon P. \tag{2.45}$$

Substituting these into the two-dimensional version of (2.35) and (2.36), we find

$$u = P_X, \quad W = -P_z, \quad \nabla^2 P = 0 \quad \text{in} \quad 0 < z < 1 + \varepsilon H, \quad 0 < X < \infty, \tag{2.46}$$

with boundary conditions

$$P = H, \quad \varepsilon H_t + P_X H_X = \frac{W}{\varepsilon} + r \quad \text{on} \quad z = 1 + \varepsilon H,$$
$$P_X = 0 \quad \text{on} \quad X = 0,$$
$$P_z = 0 \quad \text{on} \quad z = 0,$$
$$P \sim H \sim rX \quad \text{as} \quad X \to \infty. \tag{2.47}$$

At leading order in ε, this is simply

$$\begin{aligned} \nabla^2 P = 0 \quad &\text{in} \quad 0 < z < 1, \quad 0 < X < \infty, \\ P_z = 0 \quad &\text{on} \quad z = 0, 1, \\ P_X = 0 \quad &\text{on} \quad X = 0, \\ P \sim rX \quad &\text{as} \quad X \to \infty. \end{aligned} \tag{2.48}$$

Evidently, this has no solution unless we allow the incoming groundwater flux r from infinity to drain to the river at $X = 0$, $z = 1$. We do this by having a singularity in the form of a sink at the river,

$$P \sim \frac{r}{\pi} \ln \left\{ X^2 + (1 - z)^2 \right\} \quad \text{near} \quad X = 0, \quad z = 1. \tag{2.49}$$

The solution to (2.48) can be obtained by using complex variables and the method of images, by placing sinks at $z = \pm(2n + 1)$, for integral values of n. Making use of the infinite product formula (Jeffrey 2004, p. 72)

$$\prod_1^\infty \left(1 + \frac{\zeta^2}{(2n+1)^2} \right) = \cosh \frac{\pi \zeta}{2}, \tag{2.50}$$

where $\zeta = X + iz$, we find the solution to be

$$P = \frac{r}{\pi} \ln \left[\cosh^2 \frac{\pi X}{2} \cos^2 \frac{\pi z}{2} + \sinh^2 \frac{\pi X}{2} \sin^2 \frac{\pi z}{2} \right]. \tag{2.51}$$

The complex variable form of the solution is

$$\phi = P + i\psi = \frac{2r}{\pi} \ln \cosh \frac{\pi \zeta}{2}, \tag{2.52}$$

which is convenient for plotting. The streamlines of the flow are the lines $\psi =$ constant, and these are shown in figure 2.2.

This figure illustrates an important point, which is that although the flow towards a drainage point may be more or less horizontal, near the river the groundwater seeps upwards from depth. Drainage is not simply a matter of near surface recharge and drainage. This means that contaminants which enter the deep groundwater may reside there for a very long time.

A related point concerns the recharge parameter r defined in (2.43). According to table 2.1, a typical permeability for sand is 10^{-10} m^2, corresponding to a hydraulic conductivity of $K = 10^{-3}$ m s^{-1}, or 3×10^4 m y^{-1}. Even for phreatic slopes as low as $\varepsilon = 10^{-2}$, the recharge parameter $r \lesssim O(1)$, and shallow aquifer drainage is feasible.

However, finer-grained sediments are less permeable, and the calculation of r for a silt with permeability of 10^{-14} m^2 ($K = 10^{-7}$ m s^{-1} $= 3$ m y^{-1} suggests that $r \sim 1/\varepsilon^2 \gg 1$, so that if the Dupuit approximation applied, the groundwater surface would lie above the Earth's surface everywhere. This simply points out the obvious

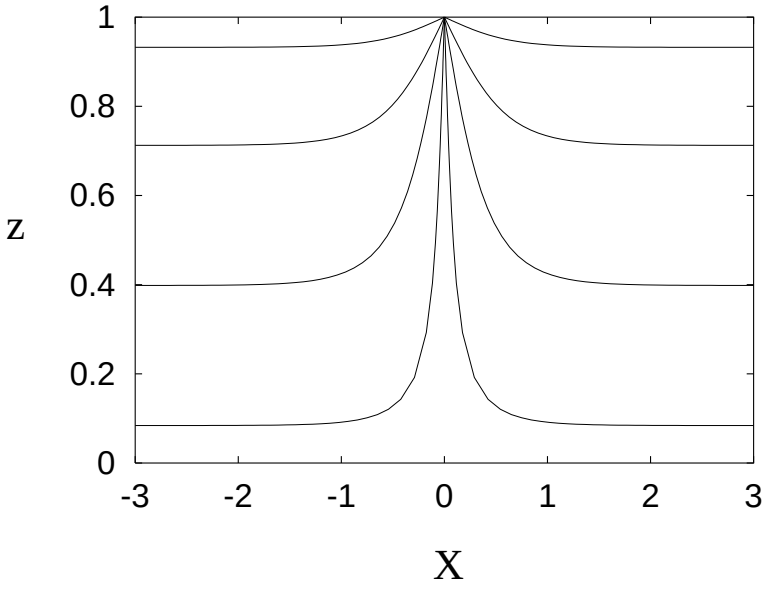

Figure 2.2: Groundwater flow lines towards a river at $X = 0$, $z = 1$.

fact that if the groundmass is insufficiently permeable, drainage cannot occur through it but water will accumulate at the surface and drain by overland flow. The fact that usually the water table is below but quite near the surface suggests that the long term response of landscape to recharge is to form topographic gradients and sufficiently deep sedimentary basins so that this *status quo* can be maintained.

2.3 Unsaturated soils

Let us now consider flow in the unsaturated zone. Above the water table, water and air occupy the pore space. If the porosity is ϕ and the water volume fraction per unit volume of soil is W, then the ratio $S = W/\phi$ is called the *relative saturation*. If $S = 1$, the soil is saturated, and if $S < 1$ it is unsaturated. The pore space of an unsaturated soil is configured as shown in figure 2.3. In particular, the air/water interface is curved, and in an equilibrium configuration the curvature of this interface will be constant throughout the pore space. The value of the curvature depends on the amount of liquid present. The less liquid there is (i. e., the smaller the value of S), then the smaller the pores where the liquid is found, and thus the higher the curvature. Associated with the curvature is a suction effect due to surface tension across the air/water interface. The upshot of all this is that the air and water pressures are related by a *capillary suction characteristic* function which expresses the difference between the pressures as a function of mean curvature, and hence, directly, S:

$$p_a - p_w = f(S). \tag{2.53}$$

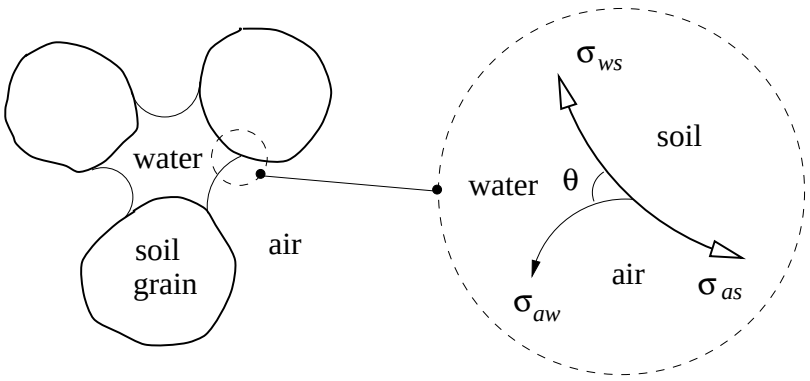

Figure 2.3: Configuration of air and water in pore space. The contact angle θ measured through the water is acute, so that water is the *wetting* phase. σ_{ws}, σ_{as} and σ_{aw} are the surface energies of the three interfaces.

The suction characteristic $f(S)$ is equal to $2\sigma\kappa$, where κ is the mean interfacial curvature: σ is the surface tension. For air and water in soil, f is positive as water is the *wetting phase*, that is, the *contact angle* at the contact line between air, water and soil grain is acute, measured through the water (see figure 2.3). The resulting form of $f(S)$ displays hysteresis as indicated in figure 2.4, with different curves depending on whether drying or wetting is taking place.

2.3.1 The Richards equation

To model the flow, we have the conservation of mass equation in the form

$$\frac{\partial(\phi S)}{\partial t} + \boldsymbol{\nabla}.\mathbf{u} = 0, \tag{2.54}$$

where we take ϕ as constant. Darcy's law for an unsaturated flow has the form, now with gravitational acceleration included,

$$\mathbf{u} = -\frac{k(S)}{\mu}[\boldsymbol{\nabla} p + \rho g \hat{\mathbf{k}}], \tag{2.55}$$

where $\hat{\mathbf{k}}$ is a unit vector upwards, and the permeability k depends on S. If $k(1) = k_0$ (the saturated permeability), then one commonly writes $k = k_0 k_{rw}(S)$, where k_{rw} is the *relative permeability*. The most obvious assumption would be $k_{rw} = S$, but this is rarely appropriate, and a better representation is a convex function, such as $k_{rw} = S^3$. An even better representation is a function such as $k_{rw} = \left(\frac{S - S_0}{1 - S_0}\right)_+^3$, where S_0 is known as the residual saturation. It represents the fact that in fine-grained soils, there is usually some minimal water fraction which cannot be removed. It is naturally associated with a capillary suction characteristic function $p_a - p = f(S)$ which tends to infinity as $S \to S_0+$, also appropriate for fine-grained soils.

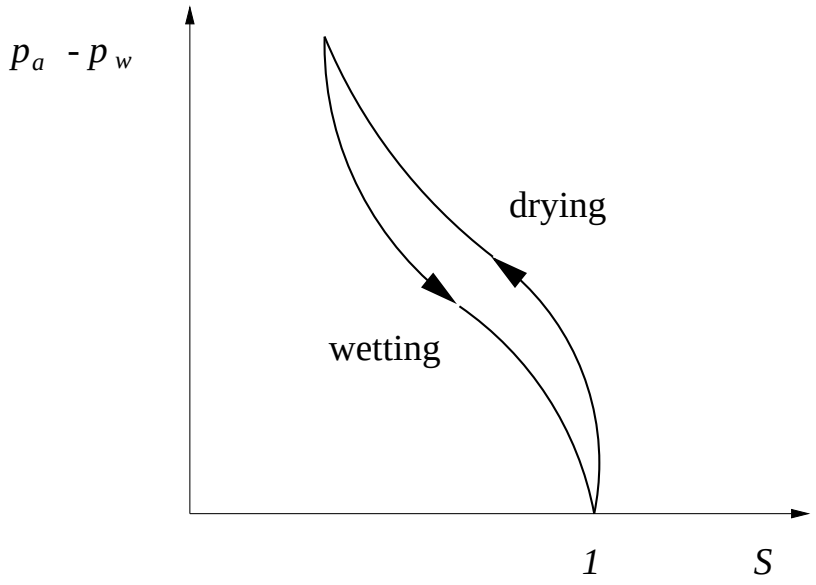

Figure 2.4: Capillary suction characteristic. It displays hysteresis in wetting and drying.

In one dimension, and if we take the vertical coordinate z to point *downwards*, we obtain the *Richards equation*

$$\phi \frac{\partial S}{\partial t} = -\frac{\partial}{\partial z}\left[\frac{k_0}{\mu} k_{rw}(S)\left\{\frac{\partial f}{\partial z} + \rho g\right\}\right]. \tag{2.56}$$

We are assuming $p_a =$ constant (and also that the soil matrix is incompressible).

2.3.2 Non-dimensionalisation

We choose scales for the variables as follows:

$$f = \frac{\sigma}{d_p}\psi, \quad z \sim \frac{\sigma}{\rho g d_p}, \quad t \sim \frac{\phi \mu z}{\rho g k_0}, \tag{2.57}$$

where d_p is grain size and σ is the surface tension, assumed constant. The Richards equation then becomes, in dimensionless variables,

$$\frac{\partial S}{\partial t} = -\frac{\partial}{\partial z}\left[k_{rw}\left(\frac{\partial \psi}{\partial z} + 1\right)\right]. \tag{2.58}$$

To be specific, we consider the case of soil wetting due to surface infiltration: of rainfall, for example. Suitable boundary conditions for infiltration are

$$S = 1 \ \text{ at } \ z = 0 \tag{2.59}$$

if surface water is ponded, or

$$k_{rw}\left(\frac{\partial\psi}{\partial z}+1\right)=u^*=\frac{\mu u_0}{k_0\rho_w g}=\frac{\mu u_0}{K_0}, \tag{2.60}$$

if there is a prescribed downward flux u_0; K_0 is the saturated hydraulic conductivity. In a dry soil we would have $S\to 0$ as $z\to\infty$, or if there is a water table at $z=z_p$, $S=1$ there.[5] For silt with $k_0=10^{-14}$ m^2, the hydraulic conductivity $K_0\sim 10^{-7}$ m s^{-1} or 3 m y^{-1}, while average rainfall in England, for example, is ≤ 1 m y^{-1}. Thus *on average* $u^*\leq 1$, but during storms we can expect $u^*\gg 1$. For large values of u^*, the desired solution may have $S>1$ at $z=0$; in this case ponding occurs (as one observes), and (2.60) is replaced by (2.59), with the pond depth being determined by the balance between accumulation, infiltration, and surface run-off.

2.3.3 Snow melting

An application of the unsaturated flow model occurs in the study of melting snow. In particular, it is found that pollutants which may be uniformly distributed in snow (e. g. SO_2 from sulphur emissions via acid rain) can be concentrated in melt water run-off, with a consequent enhanced detrimental effect on stream pollution. The question then arises, why this should be so. We shall find that uniform surface melting of a dry snowpack can lead to a meltwater spike at depth.

Suppose we have a snow pack of depth d. Snow is a porous aggregate of ice crystals, and meltwater formed at the surface can percolate through the snow pack to the base, where run-off occurs. (We ignore effects of re-freezing of meltwater). The model (2.58) is appropriate, but the relevant length scale is d. Therefore we define a parameter

$$\kappa=\frac{\sigma}{\rho g d d_p}, \tag{2.61}$$

and we rescale the variables as $z\sim 1/\kappa$, $t\sim 1/\kappa$. To be specific, we will also take

$$k_{rw}=S^3, \tag{2.62}$$

and

$$\psi(S)=\frac{1}{S}-S, \tag{2.63}$$

based on typical experimental results.

Suitable boundary conditions in a melting event might be to prescribe the melt flux u_0 at the surface, thus

$$k_{rw}\left(\frac{\partial\psi}{\partial z}+1\right)=u^*=\frac{u_0}{K_0}\quad\text{at}\quad z=0. \tag{2.64}$$

[5]With constant air pressure, continuity of S follows from continuity of pore water pressure.

If the base is impermeable, then

$$k_{rw}\left(\frac{\partial\psi}{\partial z}+1\right)=0 \quad \text{at} \quad z=h. \tag{2.65}$$

This is certainly not realistic if S reaches 1 at the base, since then ponding must occur and presumably melt drainage will occur via a channelised flow, but we examine the initial stages of the flow using (2.65). Finally, we suppose $S=0$ at $t=0$. Again, this is not realistic in the model (it implies infinite capillary suction) but it is a feasible approximation to make.

Simplification of this model now leads to the dimensionless Darcy-Richards equation in the form

$$\frac{\partial S}{\partial t}+3S^2\frac{\partial S}{\partial z}=\kappa\frac{\partial}{\partial z}\left[S(1+S^2)\frac{\partial S}{\partial z}\right]. \tag{2.66}$$

If we choose $\sigma=70$ mN m^{-1}, $d_p=0.1$ mm, $\rho=10^3$ kg m^{-3}, $g=10$ m s^{-2}, $d=1$ m, then $\kappa=0.07$. it follows that (2.66) has a propensity to form shocks, these being diffused by the term in κ over a distance $O(\kappa)$ (by analogy with the shock structure for the Burgers equation, see chapter 3).

We want to solve (2.66) with the initial condition

$$S=0 \quad \text{at} \quad t=0, \tag{2.67}$$

and the boundary conditions

$$S^3-\kappa S(1+S^2)\frac{\partial S}{\partial z}=u^* \quad \text{on} \quad z=0, \tag{2.68}$$

and

$$S^3-\kappa S(1+S^2)\frac{\partial S}{\partial z}=0 \quad \text{at} \quad z=1. \tag{2.69}$$

Roughly, for $\kappa\ll 1$, these are

$$\begin{aligned} S&=S_0 \quad \text{at} \quad z=0,\\ S&=0 \quad \text{at} \quad z=1, \end{aligned} \tag{2.70}$$

where $S_0=u^{*1/3}$, which we initially take to be $O(1)$ (and <1, so that surface ponding does not occur).

Neglecting κ, the solution is the step function

$$\begin{aligned} S&=S_0, \quad z<z_f,\\ S&=0, \quad z>z_f, \end{aligned} \tag{2.71}$$

and the shock front at z_f advances at a rate $\dot{z}_f$ given by the jump condition

$$\dot{z}_f=\frac{[S^3]^+_-}{[S]^+_-}=S_0^2. \tag{2.72}$$

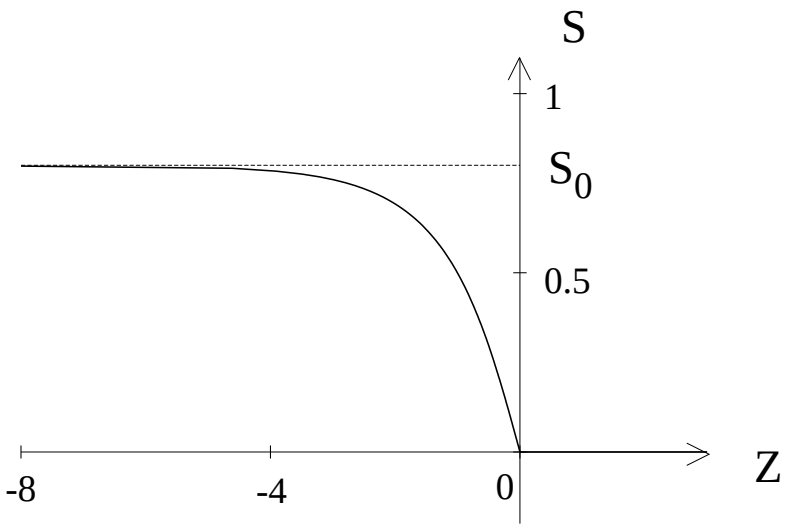

Figure 2.5: $S(Z)$ given by (2.78); the shock front terminates at the origin.

In dimensional terms, the shock front moves at speed $u_0/\phi S_0$, which is in fact obvious (given that it has constant S behind it).

The shock structure is similar to that of Burgers' equation. We put

$$z = z_f + \kappa Z, \tag{2.73}$$

and S rapidly approaches the quasi-steady solution $S(Z)$ of

$$-VS' + 3S^2S' = [S(1+S^2)S']', \tag{2.74}$$

where $V = \dot{z}_f$; hence

$$S(1+S^2)S' = -S(S_0^2 - S^2), \tag{2.75}$$

in order that $S \to S_0$ as $Z \to -\infty$, and where we have chosen

$$V = S_0^2, \tag{2.76}$$

(as $S_+ = 0$), thus reproducing (2.72). The solution is a quadrature,

$$\int^S \frac{(1+S^2)\,dS}{(S_0^2 - S^2)} = -Z, \tag{2.77}$$

with an arbitrary added constant (amounting to an origin shift for Z). Hence

$$S - \frac{(1+S_0^2)}{2S_0} \ln\left[\frac{S_0+S}{S_0-S}\right] = Z. \tag{2.78}$$

The shock structure is shown in figure 2.5; the profile terminates where $S = 0$ at $Z = 0$. In fact, (2.75) implies that $S = 0$ or (2.78) applies. Thus when S given

by (2.78) reaches zero, the solution switches to $S = 0$. The fact that $\partial S/\partial Z$ is discontinuous is not a problem because the diffusivity $S(1 + S^2)$ goes to zero when $S = 0$. This degeneracy of the equation is a signpost for fronts with discontinuous derivatives: essentially, the profile can maintain discontinuous gradients at $S = 0$ because the diffusivity is zero there, and there is no mechanism to smooth the jump away.

Suppose now that $k_0 = 10^{-10}$ m^2 and $\mu/\rho = 10^{-6}$ m^2 s^{-1}; then the saturated hydraulic conductivity $K_0 = k_0\rho g/\mu = 10^{-3}$ m s^{-1}. On the other hand, if a metre thick snow pack melts in ten days, this implies $u_0 \sim 10^{-6}$ m s^{-1}. Thus $S_0^3 = u_0/K_0 \sim 10^{-3}$, and the approximation $S \approx S_0$ looks less realistic. With

$$S^3 - \kappa S(1 + S^2)\frac{\partial S}{\partial z} = S_0^3, \tag{2.79}$$

and $S_0 \sim 10^{-1}$ and $\kappa \sim 10^{-1}$, it seems that one should assume $S \ll 1$. We define

$$S = \left(\frac{S_0^3}{\kappa}\right)^{1/2} s; \tag{2.80}$$

(2.79) becomes

$$\beta s^3 - s\left[1 + \frac{S_0^3}{\kappa}s^2\right]\frac{\partial s}{\partial z} = 1 \quad \text{on} \quad z = 0, \tag{2.81}$$

and we have $S_0^3/\kappa \sim 10^{-2}$, $\beta = (S_0/\kappa)^{3/2} \sim 1$.

We neglect the term in S_0^3/κ, so that

$$\beta s^3 - s\frac{\partial s}{\partial z} \approx 1 \quad \text{on} \quad z = 0, \tag{2.82}$$

and substituting (2.80) into (2.66) leads to

$$\frac{\partial s}{\partial \tau} + 3\beta s^2\frac{\partial s}{\partial z} \approx \frac{\partial}{\partial z}\left[s\frac{\partial s}{\partial z}\right], \tag{2.83}$$

if we define $t = \tau/\left(\kappa S_0^3\right)^{1/2}$. A simple analytic solution is no longer possible, but the development of the solution will be similar. The flux condition (2.82) at $z = 0$ allows the surface saturation to build up gradually, and a shock will only form if $\beta \gg 1$ (when the preceding solution becomes valid).

2.3.4 Similarity solutions

If, on the other hand, $\beta \ll 1$, then the saturation profile approximately satisfies

$$\begin{aligned} \frac{\partial s}{\partial \tau} &= \frac{\partial}{\partial z}\left[s\frac{\partial s}{\partial z}\right], \\ -s\frac{\partial s}{\partial z} &= \begin{cases} 1 & \text{on} \quad z = 0, \\ 0 & \text{on} \quad z = 1. \end{cases} \end{aligned} \tag{2.84}$$

At least for small times, the model admits a similarity solution of the form

$$s = \tau^\alpha f(\eta), \quad \eta = z/\tau^\beta, \tag{2.85}$$

where satisfaction of the equations and boundary conditions requires $2\alpha = \beta$ and $2\beta = 1 = \alpha$, whence $\alpha = \frac{1}{3}$, $\beta = \frac{2}{3}$, and f satisfies

$$(ff')' - \tfrac{1}{3}(f - 2\eta f') = 0, \tag{2.86}$$

with the condition at $z = 0$ becoming

$$-ff' = 1 \quad \text{at} \quad \eta = 0. \tag{2.87}$$

The condition at $z = 1$ can be satisfied for small enough τ, as we shall see, because the equation (2.86) is degenerate, and f reaches zero in a finite distance, η_0, say, and $f = 0$ for $\eta > \eta_0$. As $\eta = 1/\tau^{2/3}$ at $z = 1$, then this solution will satisfy the no flux condition at $z = 1$ as long as $\tau < \eta_0^{-3/2}$, when the advancing front will reach $z = 1$.

To see why f behaves in this way, integrate once to find

$$f(f' + \tfrac{2}{3}\eta) = -1 + \int_0^\eta f \, d\eta. \tag{2.88}$$

For small η, the right hand side is negative, and f is positive (to make physical sense), so f decreases (and in fact $f' < -\frac{2}{3}\eta$). For sufficiently small $f(0) = f_0$, f will reach zero at a finite distance $\eta = \eta_0$, and the solution must terminate. On the other hand, for sufficiently large f_0, $\displaystyle\int_0^\eta f \, d\eta$ reaches 1 at $\eta = \eta_1$ while f is still positive (and $f' = -\frac{2}{3}\eta_1$ there). For $\eta > \eta_1$, then f remains positive and $f' > -\frac{2}{3}\eta$ (f cannot reach zero for $\eta > \eta_1$ since $\displaystyle\int_0^\eta f \, d\eta > 1$ for $\eta > \eta_1$). Eventually f must have a minimum and thereafter increase with η. This is also unphysical, so we require f to reach zero at $\eta = \eta_0$. This will occur for a range of f_0, and we have to select f_0 in order that

$$\int_0^{\eta_0} f \, d\eta = 1, \tag{2.89}$$

which in fact represents global conservation of mass. Figure 2.6 shows the schematic form of solution both for $\beta \gg 1$ and $\beta \ll 1$. Evidently the solution for $\beta \sim 1$ will have a profile with a travelling front between these two end cases.

2.4 Immiscible two-phase flows: the Buckley-Leverett equation

In some circumstances, the flow of more than one phase in a porous medium is important. The type example is the flow of oil and gas, or oil and water (or all three!) in a sedimentary basin, such as that beneath the North Sea. Suppose there

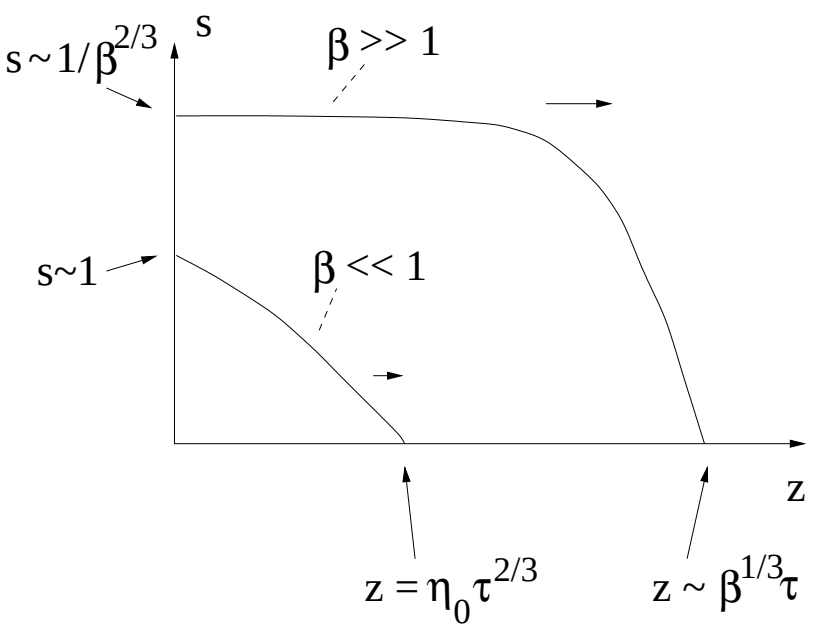

Figure 2.6: Schematic representation of the evolution of s in (2.83) for both large and small β.

are two phases; denote the phases by subscripts 1 and 2, with fluid 2 being the *wetting fluid*, and S is its saturation. Then the capillary suction characteristic is

$$p_1 - p_2 = p_c(S), \tag{2.90}$$

with the capillary suction p_c being a positive, monotonically decreasing function of saturation S; mass conservation takes the form

$$\begin{aligned} -\phi\frac{\partial S}{\partial t} + \boldsymbol{\nabla}.\mathbf{u}_1 &= 0, \\ \phi\frac{\partial S}{\partial t} + \boldsymbol{\nabla}.\mathbf{u}_2 &= 0, \end{aligned} \tag{2.91}$$

where ϕ is (constant) porosity, and Darcy's law for each phase is

$$\begin{aligned} \mathbf{u}_1 &= -\frac{k_0}{\mu_1}k_{r1}[\boldsymbol{\nabla}p_1 + \rho_1 g\hat{\mathbf{k}}], \\ \mathbf{u}_2 &= -\frac{k_0}{\mu_2}k_{r2}[\boldsymbol{\nabla}p_2 + \rho_2 g\hat{\mathbf{k}}], \end{aligned} \tag{2.92}$$

with k_{ri} being the relative permeability of fluid i.

For example, if we consider a one-dimensional flow, with z pointing upwards, then we can integrate (2.91) to yield the total flux

$$u_1 + u_2 = q(t). \tag{2.93}$$

If we define the mobilities of each fluid as

$$M_i = \frac{k_0}{\mu_i} k_{ri}, \tag{2.94}$$

then it is straightforward to derive the equation for S,

$$\phi \frac{\partial S}{\partial t} = -\frac{\partial}{\partial z}\left[M_{\text{eff}} \left\{ \frac{q}{M_1} + \frac{\partial p_c}{\partial z} + (\rho_1 - \rho_2) g \right\} \right], \tag{2.95}$$

where the effective mobility is determined by

$$M_{\text{eff}} = \left(\frac{1}{M_1} + \frac{1}{M_2} \right)^{-1}. \tag{2.96}$$

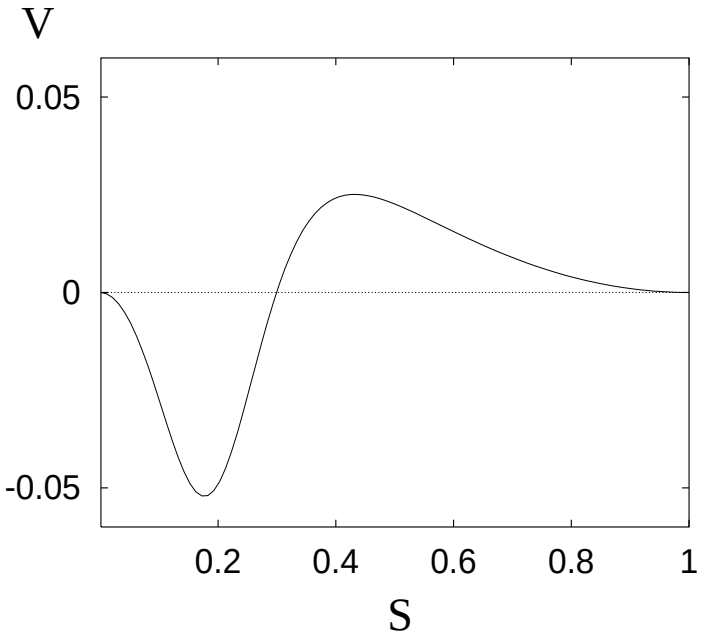

Figure 2.7: Graph of dimensionless wave speed $V(S)$ as a function of wetting fluid saturation, indicating the speed and direction of wave motion ($V > 0$ means waves move upwards) if the wetting fluid is more dense. The viscosity ratio μ_r (see (2.100)) is taken to be 30.

This is a convective-diffusion equation for S. If suction is very small, we obtain the *Buckley-Leverett equation*

$$\phi \frac{\partial S}{\partial t} + \frac{\partial}{\partial z}\left[M_{\text{eff}} \left\{ \frac{q}{M_1} + (\rho_1 - \rho_2) g \right\} \right] = 0, \tag{2.97}$$

which is a nonlinear hyperbolic wave equation. As a typical situation, suppose $q = 0$, and $k_{r2} = S^3$, $k_{r1} = (1 - S)^3$. Then

$$M_{\text{eff}} = \frac{k_0 S^3 (1 - S)^3}{\mu_1 S^3 + \mu_2 (1 - S)^3}, \tag{2.98}$$

and the wave speed $v(S)$ is given by

$$v = -(\rho_2 - \rho_1) g M'_{\text{eff}}(S) = v_0 V(S), \tag{2.99}$$

where

$$v_0 = \frac{(\rho_2 - \rho_1) g k_0}{\mu_2}, \qquad V(S) = \frac{\chi'(S)}{\chi(S)^2},$$
$$\chi(S) = \frac{\mu_r}{(1-S)^3} + \frac{1}{S^3}, \qquad \mu_r = \frac{\mu_1}{\mu_2}. \tag{2.100}$$

The variation of V with S is shown in figure 2.7. For $\rho_2 > \rho_1$ (as for oil and water, where water is the wetting phase), waves move upwards at low water saturation and downwards at high saturation.

Shocks will form, but these are smoothed by the diffusion term $-\dfrac{\partial}{\partial z}\left[M_{\text{eff}} p'_c \dfrac{\partial S}{\partial z}\right]$, in which the diffusion coefficient is

$$D = -M_{\text{eff}}\, p'_c. \tag{2.101}$$

As a typical example, take

$$p_c = \frac{p_0 (1-S)^{\lambda_1}}{S^{\lambda_2}} \tag{2.102}$$

with $\lambda_i > 0$. Then we find

$$D = k_0 p_0 S^{2-\lambda_2} (1-S)^{2+\lambda_1} \left[\frac{\lambda_1 S + \lambda_2 (1-S)}{\mu_1 S^3 + \mu_2 (1-S)^3}\right], \tag{2.103}$$

and we see that D is typically degenerate at $S = 0$. In particular, if $\lambda_2 < 2$, then infiltration of the wetting phase into the non-wetting phase proceeds at a finite rate, and this always occurs for infiltration of the non-wetting phase into the wetting phase.

A particular limiting case is when one phase is much less dense than the other, the usual situation being that of gas and liquid. This is exemplified by the problem of snow-melt run-off considered earlier. In that case, water is the wetting phase, thus $\rho_2 - \rho_1 = \rho_w - \rho_a$ is positive, and also $\mu_w \approx 10^{-3}$ Pa s, $\mu_a \approx 10^{-5}$ Pa s, whence $\mu_a \ll \mu_w$ ($\mu_r \ll 1$), so that, from (2.98),

$$M_{\text{eff}} \approx \frac{k_0 S^3}{\mu_w}, \tag{2.104}$$

at least for saturations not close to unity. Shocks form and propagate downwards (since $\rho_2 > \rho_1$). The presence of non-zero flux $q < 0$ does not affect this statement. Interestingly, the approximation (2.104) will always break down at sufficiently high saturation. Inspection of $V(S)$ for $\mu_r = 0.01$ (as for air and water) indicates that (2.104) is an excellent approximation for $S \lesssim 0.5$, but not for $S \gtrsim 0.6$; for $S \gtrsim 0.76$, V is positive and waves move upwards. As $\mu_r \to 0$, the right hand hump in figure 2.7 moves towards $S = 1$, but does not disappear; indeed the value of the maximum increases, and is $V \sim \mu_r^{-1/3}$. Thus the single phase approximation for unsaturated flow is a singular approximation when $\mu_r \ll 1$ and $1 - S \ll 1$.

2.5 Consolidation

Consolidation refers to the ability of a granular porous medium such as a soil to compact under its own weight, or by the imposition of an overburden pressure. The grains of the medium rearrange themselves under the pressure, thus reducing the porosity and in the process pore fluid is expelled. Since the porosity is no longer constant, we have to postulate a relation between the porosity ϕ and the pore pressure p. In practice, it is found that soils, when compressed, obey a (non-reversible) relation between ϕ and the *effective pressure*

$$p_e = P - p, \tag{2.105}$$

where P is the overburden pressure.

The concept of effective pressure, or more generally effective stress, is an extremely important one. The idea is that the total imposed pressure (e. g., the overburden pressure due to the weight of the rock or soil) is borne by both the pore fluid and the porous medium. The pore fluid is typically at a lower pressure than the overburden, and the extra stress (the effective stress) is that which is applied through grain to grain contacts. Thus the effective pressure is that which is transmitted through the porous medium, and it is in consequence of this that the medium responds to the effective stress; in particular, the characteristic relation between ϕ and p_e represents the nonlinear pseudo-elastic effect of compression.

As p_e increases, so ϕ decreases, thus we can write (ignoring irreversibility)

$$p_e = p_e(\phi), \quad p_e'(\phi) < 0. \tag{2.106}$$

Taking fluid density ρ to be constant, we obtain from the conservation of mass equation the nonlinear diffusion equation

$$\phi_t = \boldsymbol{\nabla}.\left[\frac{k(\phi)}{\mu}\,|\,p_e'(\phi)\,|\,\boldsymbol{\nabla}\phi\right], \tag{2.107}$$

assuming Darcy's law with a permeability k, ignoring gravity, and taking P as constant. This is essentially the same as the Richards equation for unsaturated soils.

The dependence of the effective pressure on porosity is non-trivial and involves hysteresis, as indicated in figure 2.8. Specifically, a soil follows the *normal consolidation line* providing consolidation is occurring, i.e $\dot{p}_e > 0$. However, if at some point the effective pressure is reduced, only a partial recovery of ϕ takes place. When p_e is increased again, ϕ more or less retraces its (overconsolidated) path to the normal consolidation line, and then resumes its normal consolidation path. Here we will ignore effects of hysteresis, as in (3.90).

When modelling groundwater flow in a consolidating medium, we must take account also of deformation of the medium itself. In turn, this requires prescription of a constitutive rheology for the deformable matrix. This is often a complex matter, but luckily in one dimension, the issue does not arise, and a one-dimensional model is often what is of practical interest. We take z to point vertically upwards, and let

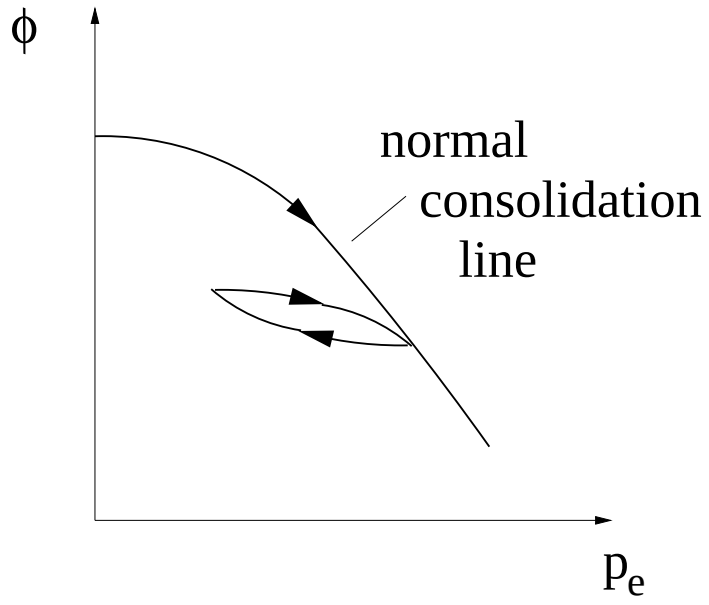

Figure 2.8: Form of the relationship between porosity and effective pressure. A hysteretic decompression-reconsolidation loop is indicated. In soil mechanics this relationship is often written in terms of the *void ratio* $e = \phi/(1-\phi)$, and specifically $e = e_0 - C_c \log p_e$, where C_c is the *compression index*.

v^l and v^s be the linear (or *phase-averaged*) velocities of liquid and solid, respectively. Then $u^l = \phi v^l$ and $u^s = (1-\phi)v^s$ are the respective fluxes, and conservation of mass of each phase requires

$$\begin{aligned} \frac{\partial \phi}{\partial t} + \frac{\partial (\phi v^l)}{\partial z} &= 0, \\ -\frac{\partial \phi}{\partial t} + \frac{\partial \{(1-\phi)v^s\}}{\partial z} &= 0; \end{aligned} \tag{2.108}$$

Darcy's law is then

$$\phi(v^l - v^s) = -\frac{k}{\mu}\left[\frac{\partial p}{\partial z} + \rho_l g\right], \tag{2.109}$$

while the overburden pressure is

$$P = P_0 + [\rho_s(1-\phi) + \rho_l \phi]g(h-z); \tag{2.110}$$

here $z = h$ represents the ground surface and P_0 is the applied load. (2.110) assumes variations of ϕ are small. More generally, we would have $\partial P/\partial z = -[\rho_s(1-\phi) + \rho_l\phi]g$. The effective pressure is then just $p_e = P - p$.

We suppose these equations apply in a vertical column $0 < z < h$, for which suitable boundary conditions are

$$\begin{aligned} v^l = v^s = 0 \quad &\text{at} \quad z = 0, \\ p = 0,\ \dot{h} = v^s \quad &\text{at} \quad z = h, \end{aligned} \tag{2.111}$$

and with an initial condition for p (or ϕ).

The two mass conservation equations imply

$$v^s = -\frac{\phi v^l}{1-\phi}. \tag{2.112}$$

Substituting this into (2.109), we derive, using (2.108),

$$\frac{\partial \phi}{\partial t} = \frac{\partial}{\partial z}\left[\frac{k}{\mu}(1-\phi)\left\{\frac{\partial p}{\partial z} + \rho_l g\right\}\right]. \tag{2.113}$$

If we assume the normal consolidation line takes the commonly assumed form (see figure 2.8)

$$\frac{\phi}{1-\phi} = e_0 - C_c \ln(p_e/p_e^0), \tag{2.114}$$

then we derive the consolidation equation

$$\frac{\partial p_e}{\partial t} = \frac{p_e}{C_c(1-\phi)^2}\frac{\partial}{\partial z}\left[\frac{k}{\mu}(1-\phi)\left\{\frac{\partial p_e}{\partial z} + \Delta\rho(1-\phi)g\right\}\right], \tag{2.115}$$

where $\Delta\rho = \rho_s - \rho_l$.

If C_c is small (and typical values are in the range $C_c \leq 0.1$) then ϕ varies little, and the consolidation equation takes the simpler form

$$\frac{\partial p_e}{\partial t} = c_v \frac{\partial^2 p_e}{\partial z^2}, \tag{2.116}$$

where

$$c_v = \frac{k}{\mu}\frac{p_e}{C_c(1-\phi)} \tag{2.117}$$

is the *coefficient of consolidation*.

Suitable boundary conditions are

$$\begin{aligned} \frac{\partial p_e}{\partial z} + \Delta\rho(1-\phi)g = 0 \quad \text{at} \quad z = 0, \\ p_e = P_0 \quad \text{at} \quad z = h, \end{aligned} \tag{2.118}$$

and if the load is applied at $t = 0$, the initial condition is

$$p_e = \Delta\rho(1-\phi)g(h-z) \text{ at } t = 0. \tag{2.119}$$

The equation is trivially solved. The consolidation time is

$$t_c \sim \frac{h^2}{c_v} = \frac{\mu C_c(1-\phi)h^2}{k p_e}, \tag{2.120}$$

and depends primarily on the permeability k. If we take $k \sim 10^{-14}$ m^2 (silt), $C_c = 0.1$, $\phi = 0.3$, $\mu = 10^{-3}$ Pa s, $P_0 = 10^5$ Pa (a small house), then $c_v \sim 10^{-5}$ m^2 s^{-1}, and $t_e \sim 1$ year for $h \sim 10$ m.

2.6 Compaction

Compaction is the same process as consolidation, but on a larger scale. Other mechanisms can cause compaction apart from the rearrangement of sediments: pressure solution in sedimentary basins, grain creep in partially molten mantle (see chapter **??**). The compaction of sedimentary basins is a problem which has practical consequences in oil-drilling operations, since the occurrence of abnormal pore pressures can lead to blow-out and collapse of the borehole wall. Such abnormal pore pressures (i.e., above hydrostatic) can occur for a variety of reasons, and part of the purpose of modelling the system is to determine which of these are likely to be realistic causes. A further distinction from smaller scale consolidation is that the variation in porosity (and, particularly, permeability) is large.

The situation we consider was shown in figure **??**. Sediments, both organic and inorganic, are deposited at the ocean bottom and accumulate. As they do so, they compact under their weight, thus expelling pore water. If the compaction is fast (i.e., the rate of sedimentation is greater than the hydraulic conductivity of the sediments) then excess pore pressure will occur.

Sedimentary basins, such as the North Sea or the Gulf of Mexico, are typically hundreds of kilometres in extent and several kilometres deep. It is thus appropriate to model the compacting system as one-dimensional. A typical sedimentation rate is 10^{-11} m s^{-1}, or 300 m My^{-1}, so that a 10 kilometre deep basin may accumulate in 30 My (30 million years). On such long time scales, tectonic processes are important, and in general accumulation is not a monotonic process. If tectonic uplift occurs so that the surface of the basin rises above sea level, then erosion leads to denudation and a negative sedimentation rate. Indeed, one purpose of studying basin porosity and pore pressure profiles is to try and infer what the previous subsidence history was — an inverse problem.

The basic mathematical model is that of slow two-phase flow, where the phases are solid and liquid, and is the same as that of consolidation theory. The effective pressure p_e is related, in an elastic medium, to the porosity by a function $p_e = p_e(\phi)$. In a soil, or for sediments near the surface up to depths of perhaps 500 m, the relation is elastic and hysteretic. At greater depths, more than a kilometre, pressure solution becomes important, and an effective viscous relationship becomes appropriate, as described below. At greater depths still, cementation occurs and a stiffer elastic rheology should apply.[6] In addition, the permeability is a function $k = k(\phi)$ of porosity, with k decreasing to zero fairly rapidly as ϕ decreases to zero.

Let us suppose the basin overlies an impermeable basement at $z = 0$, and that its surface is at $z = h$; then suitable boundary conditions are

$$v^s = v^l = 0 \quad \text{at} \quad z = 0,$$

$$p_e = 0, \quad \dot{h} = \dot{m}_s + v^s \quad \text{at} \quad z = h, \tag{2.121}$$

[6]Except that at elevated temperatures, creep deformation will start to occur.

where v^s and v^l are solid and liquid average velocities, and $\dot{m}_s$ is the prescribed sedimentation rate, which we take for simplicity to be constant.

If we assume a specific elastic compactive rheology of the form

$$p_e = p_0\{\ln(\phi_0/\phi) - (\phi_0 - \phi)\}, \tag{2.122}$$

then non-dimensionalisation (using a depth scale $d = \dfrac{p_0}{(\rho_s - \rho_l)g}$ and a time scale $\dfrac{d}{\dot{m}_s}$) and simplification of the model leads to the nonlinear diffusion equation, analogous to (2.113),

$$\frac{\partial \phi}{\partial t} = \lambda \frac{\partial}{\partial z}\left\{\tilde{k}(1-\phi)^2\left[\frac{1}{\phi}\frac{\partial \phi}{\partial z} - 1\right]\right\}, \tag{2.123}$$

where the permeability is defined to be

$$k = k_0\tilde{k}(\phi), \tag{2.124}$$

k_0 being a suitable scale for k.

The dimensionless parameter λ is given by

$$\lambda = \frac{K_0}{\dot{m}_s}, \tag{2.125}$$

where $K_0 = k_0(\rho_s - \rho_l)g/\mu$ is essentially the surface hydraulic conductivity, and we can then distinguish between slow compaction ($\lambda \ll 1$) and fast compaction ($\lambda \gg 1$). Typical values of λ depend primarily on the sediment type. For $\dot{m}_s = 10^{-11}$ m s^{-1}, we have $\lambda \approx 0.1$ for the finest clay, $\lambda \approx 10^9$ for coarse sands. In general, therefore, we can expect large values of λ. The associated boundary conditions for the model become

$$\begin{gathered}\phi_z - \phi = 0 \text{ at } z = 0, \\ \phi = \phi_0, \;\; \dot{h} = 1 + \lambda\tilde{k}(1-\phi)\left[\frac{1}{\phi}\frac{\partial \phi}{\partial z} - 1\right] \text{ at } z = h.\end{gathered} \tag{2.126}$$

Slow compaction, $\lambda \ll 1$

When λ is small, overpressuring occurs. A boundary layer analysis is easy to do, and shows that $\phi \approx \phi_0$ in the bulk of the (uncompacted) sediment, while a compacting boundary layer of thickness $\sqrt{\lambda t}$ exists at the base.

Fast compaction, $\lambda \gg 1$

The more realistic case of fast compaction is also the more mathematically interesting. Most simply, the solution when $\lambda \gg 1$ is the equilibrium profile

$$\phi = \phi_0 \exp[h - z]; \tag{2.127}$$

the exponential decline of porosity with depth is sometimes called an Athy profile, but it only applies while $\lambda\tilde{k} \gg 1$. If we assume a power law for the dimensionless permeability of the form

$$\tilde{k} = (\phi/\phi_0)^m, \tag{2.128}$$

then we find that $\lambda\tilde{k}$ reaches one when ϕ decreases to a value

$$\phi^* = \phi_0 \exp\left[-\frac{1}{m}\ln\lambda\right], \tag{2.129}$$

and this occurs at a dimensionless depth

$$\Pi = \frac{1}{m}\ln\lambda \tag{2.130}$$

and time

$$t^* = \frac{\Pi - \phi_0(1 - e^{-\Pi})}{1 - \phi_0}. \tag{2.131}$$

Typical values $m = 8$, $\lambda = 100$, $\phi_0 = 0.5$, give values $\phi^* = 0.28$, $\Pi = 0.58$, $t^* = 0.71$. In particular, for a reasonable depth scale of 1 km (corresponding to $p_0 = 2 \times 10^7$ Pa = 200 bars), this would correspond to a depth of 580 m. Below this, the profile is not equilibrated, and the pore pressure is elevated. Figure 2.9 shows the resulting difference in the porosity profiles at $t = t^*$ and $t > t^*$, and figure 2.10 shows the effect on the pore pressure, whose gradient changes abruptly from hydrostatic to lithostatic at the critical depth.

If we take $\phi^* = O(1)$ and $\lambda \gg 1$, then formally $m \gg 1$, and it is possible to analyse the profile below the critical depth. One finds that

$$\phi = \phi^* \exp\left[-\frac{1}{m}\{\ln m + O(1)\}\right], \tag{2.132}$$

which can explain the flattening of the porosity profile evident in figure 2.9, and which is also seen in field data.

Viscous compaction

Below a depth of perhaps a kilometre, pressure solution at intergranular contacts becomes important, and the resulting dissolution and local reprecipitation leads to an effective creep of the grains (and hence of the bulk medium) in a manner analogous to regelation in ice. For such viscous compaction, the constitutive relation for the effective pressure becomes

$$p_e = -\xi \boldsymbol{\nabla}.\mathbf{u}^s. \tag{2.133}$$

In one dimension, the resulting dimensionless model is

$$-\frac{\partial\phi}{\partial t} + \frac{\partial}{\partial z}\left[(1-\phi)u\right] = 0,$$

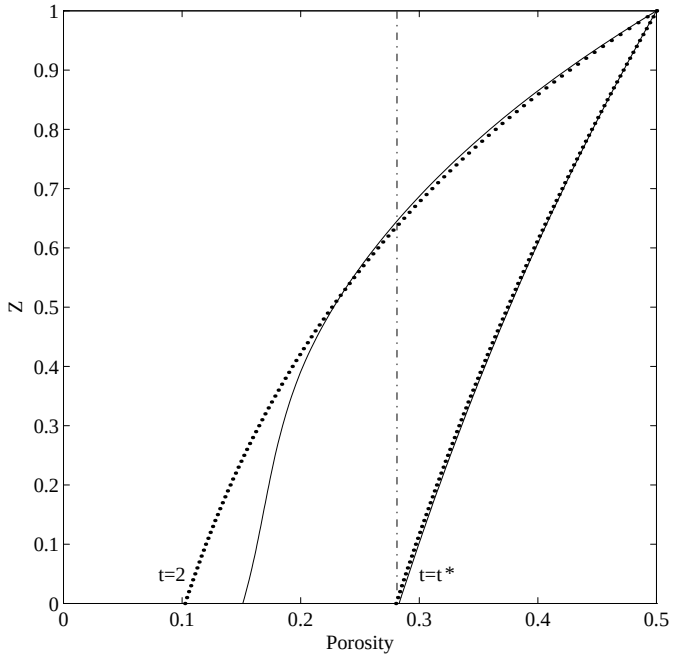

Figure 2.9: Solution of (2.123) for $\lambda = 100$ at times $t = t^* \approx 0.71$ and at $t = 2$. The porosity (horizontal axis) is plotted as a function of the scaled vertical height $z/h(t)$. The solid lines are numerical solutions, whereas the dotted lines are the large λ equilibrium profiles. There is a clear divergence at depth for $t > t^*$.

$$u = -\lambda\tilde{k}\left[\frac{\partial p}{\partial z} + 1 - \phi\right],$$

$$p = -\Xi\frac{\partial u}{\partial z}, \tag{2.134}$$

where p is the scaled effective pressure. The compaction parameter is the same as before, and the extra parameter Ξ can be taken to be of $O(1)$ for typical basin depths of kilometres. Boundary conditions for (2.134) are

$$u = 0 \quad \text{on} \quad z = 0,$$

$$p = 0, \quad \phi = \phi_0, \quad \dot{h} = 1 + u \quad \text{at} \quad z = h. \tag{2.135}$$

This system can also be studied asymptotically. When $\lambda \ll 1$, compaction is slow and a basal compaction layer again forms. When $\lambda \gg 1$, explicit solutions can again be obtained. There is an upper layer at equilibrium, but now the porosity decreases

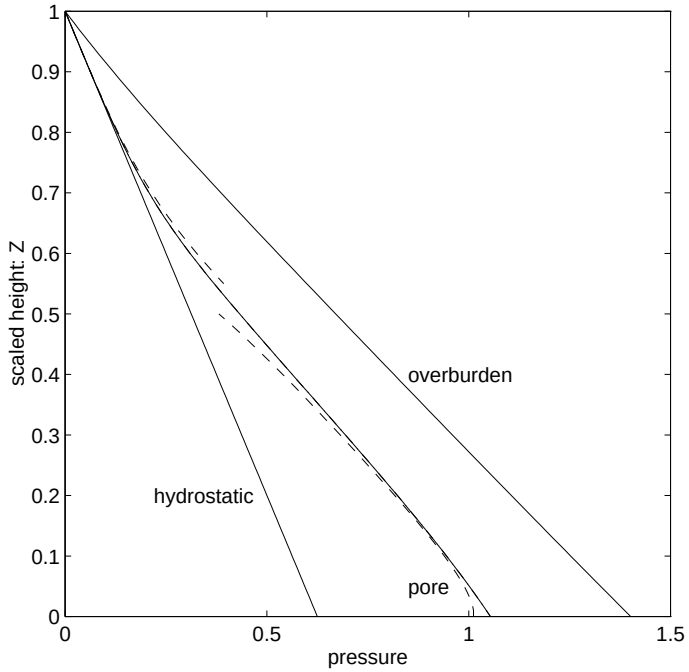

Figure 2.10: Hydrostatic, overburden (lithostatic) and pore pressures at $t = 5$ and $\lambda = 100$, as functions of the scaled height $z/h(t)$. The transition from equilibrium to non-equilibrium compaction at the critical depth is associated with a transition from normal to abnormal pore pressures. The dashed lines represent two distinct approximations to the pore pressure profile, respectively valid above and below the transition region.

concavely with depth.[7] As before, there is a transition when $\phi = \phi^*$, and below this

$$\phi = \phi^* \exp\left[-\frac{2}{m}\{\ln m + O(1)\}\right], \tag{2.136}$$

similar to (2.132).

The main distinction between viscous and elastic compaction is thus in the form of the rapidly compacted equilibrium profile near the surface (figure 2.11). The concave profile is not consistent with observations, but we need not expect it to be, as the viscous behaviour of pressure solution only becomes appropriate at reasonable depths. A more general relation which allows for this is a viscoelastic compaction law of the form

$$\nabla.\mathbf{u}^s = -\frac{1}{K_e}\frac{dp_e}{dt_s} - \frac{p_e}{\xi}. \tag{2.137}$$

[7]In view of chapter **??**, we need to be careful here. The function is mathematically concave, i. e., the rate of decrease of porosity with depth increases as depth increases.

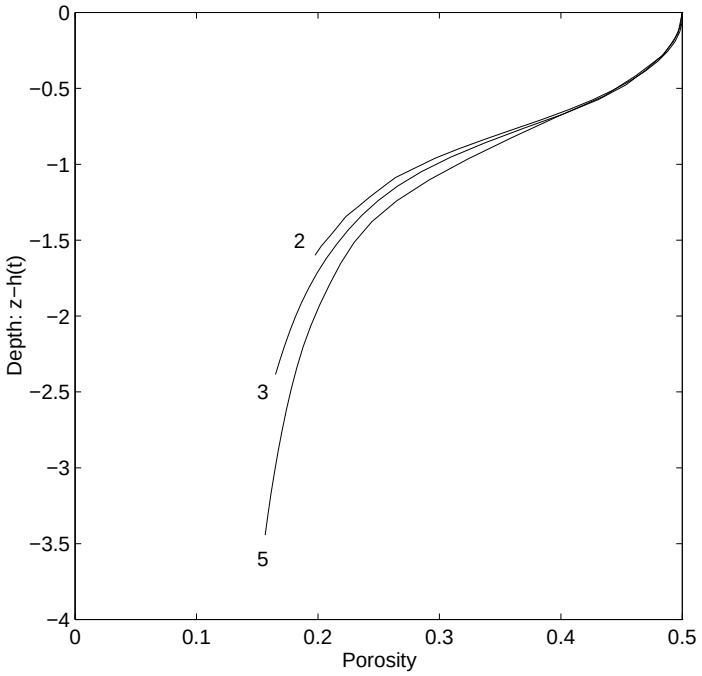

Figure 2.11: Evolution of the porosity as a function of depth $h - z$, with a viscous rheology, at $\lambda = 100$. The upper concave part is in equilibrium, while overpressuring occurs where the profile is flatter below this.

2.7 Notes and references

Flow in porous media is described in the books by Bear (1972) and Dullien (1979). More recent versions, for example by Bear and Bachmat (1990) have developed a taste for more theoretical, deductive treatments based on homogenisation (see below) or averaging, with a concomitant loss of readability. The classic geologists' book on groundwater is by Freeze and Cherry (1979) and the classic engineering text is by Polubarinova-Kochina (1962). A short introduction, of geographical style, is by Price (1985). A more mathematical survey, with a variety of applications, is by Bear and Verruijt (1987). The book edited by Cushman (1990) contains a wealth of articles on topics of varied and current interest, including dispersion, homogenisation, averaging, dual porosity models, multigrid methods and heterogeneous porous media. Further information on the concepts of soil mechanics can be found in Lambe and Whitman (1979).

Homogenisation

The technique of homogenisation is no more than the technique of averaging in the spatial domain, most often formulated as a multiple scales method. Whole books have been written about it, for example those by Bensoussan *et al.* (1978) and Sanchez-Palencia (1983). For application to porous media, see, for example, Ene's article in the book edited by Cushman (1990).

Piping

Many dams are built of concrete, and in this case the problems associated with seepage do not arise, owing to the virtual impermeability of concrete. Earth and rockfill dams do exist, however, and are liable to failure by a mechanism called *piping*. The Darcy flow through the porous dam causes channels to form by eroding away fine particles. The resultant channelisation concentrates the flow, increasing the force exerted by the flow on the medium and thus increasing the erosion/collapse rate of the channel wall. We can write Darcy's law as a force balance on the liquid phase,

$$0 = -\phi \boldsymbol{\nabla} p - \frac{\phi \mu}{k} \mathbf{v}_l - \phi \rho_l g \mathbf{k} \tag{2.138}$$

($\mathbf{k}$ being vertically upwards) and $\phi \mu \mathbf{v}_l / k$ is an *interactive drag term*; then the corresponding force balance for the solid phase is

$$0 = -(1-\phi) \boldsymbol{\nabla} p_s + \frac{\phi \mu}{k} \mathbf{v}_l - (1-\phi) \rho_s g \mathbf{k}, \tag{2.139}$$

where p_s is the pressure in the solid. For a granular solid, we can expect grain motion to occur if the interactive force is large enough to overcome friction and cohesion; the typical kind of criterion is that the shear stress τ satisfies

$$\tau \geq c + p_e \tan \phi, \tag{2.140}$$

but in view of the large confining pressure and the necessity of dilatancy for soil deformation, the piping criterion will in practice be satisfied at the toe of the dam (i.e. the front), and piping channels will eat their way back into the dam, in much the same way that river drainage channels eat their way into a hillslope. A simpler criterion at the toe then follows from the necessity that the effective pressure on the grains be positive. A lucid discussion by Bear and Bachmat (1990, p. 153) indicates that the solid pressure is related to the *effective* pressure p_e which controls grain deformation by

$$p_e = (1-\phi)(p_s - p), \tag{2.141}$$

and in this case the piping criterion at the toe is that $p_e < 0$ in the soil there, or $\partial p_e/\partial z > 0$. From (2.138), (2.139) and (2.141), this implies piping if

$$\frac{\mu v}{k} > (\rho_s - \rho_l)(1-\phi)g, \tag{2.142}$$

where v is the vertical component of $\mathbf{v}_l$. This criterion is given by Bear (1972). More generally, piping can be expected to occur if p_e reaches 0 in the soil interior (ignoring cohesion). Sellmeijer and Koenders (1991) develop a model for piping.

Taylor dispersion

Taylor dispersion is named after its investigation by Taylor (1953), who carried out experiments on the dispersal of solute in flow down a tube. The dispersion is enabled by the combination of differential axial advection by the down tube velocity, typically a Poiseuille flow, and the rapid cross stream diffusion which renders the cross-sectional concentration profile radially uniform. The theory of Taylor is somewhat heuristic; it was later elaborated by Aris (1956). For a formal derivation using asymptotic methods, see Fowler (1997, p. 222, exercise 2).

Its application to porous media stems from the conceptual idea that the pore space consists of a network of narrow tubules connected at pore junctions. If the tubes are of radius a and length d_p, the latter corresponding to grain size, then the Darcy flux $|\mathbf{u}| \sim \pi\phi U$, while the pore radius $a \sim d_p\sqrt{\phi}$. This would suggest a Taylor dispersion coefficient of

$$D_T \approx \frac{a^2U^2}{48D} \sim \frac{d_p^2|\mathbf{u}|^2}{48\pi^2 D\phi}, \tag{2.143}$$

as opposed to the measured values which more nearly have $D_T \sim |\mathbf{u}|$. Taylor dispersion in porous media has been studied by Saffman (1959), Brenner (1980) and Rubinstein and Mauri (1986), the latter using the method of homogenisation.

Biofilm growth

Monod kinetics was described by Monod (1949), by way of analogy with enzyme kinetics, where one considers the uptake of nutrients as occurring through a series of fast intermediary reactions; when two nutrients control growth, as in respiration,

it is usual to take the growth rate as proportional to the product of two Monod factors (Bader 1978). A variety of enhancements to this simple model have also been proposed to account for nutrient consumption due to maintenance, inactivation of cells in adverse conditions, and other observed effects (Beeftink *et al.* 1990, Wanner *et al.* 2006).

Bacteria in soils commonly grow as attached biofilms on soil grains, with a thickness of the order of 100 μ. A variety of models to describe biofilm growth have been presented, with an ultimate view of being able to parameterise the uptake rate of contaminant species in soils and other environments (Rittmann and McCarty 1980, Picioreanu *et al.* 1998, Eberl *et al.* 2001, Dockery and Klapper 2001, Cogan and Keener 2004).

Remediation sites

The three sites described in section **??** are under study by the Groundwater Restoration and Protection Group at the University of Sheffield, led by Professor David Lerner. The site at Four Ashes is described by Mayer *et al.* (2001), that at Rexco by Hüttmann *et al.* (2003), and that at St. Alban's by Wealthall *et al.* (2001).

The description of the two species reaction front given by (**??**) is similar to that for a *diffusion flame* (Buckmaster and Ludford 1982) in combustion, and also corrosion in alloys (Hagan *et al.* 1986). It is not conceptually difficult to extend this approach to an arbitrary number of reactions, although it may become awkward when multiple reaction fronts are present (see, for example, Dewynne *et al.* (1993)).

Diagenesis

The first order reaction kinetics (**??**) for the smectite-illite transition was proposed by Eberl and Hower (1976). Information on solubility limits is given by Aagard and Helgeson (1983) and Sass *et al.* (1987). The asymptotic approximation called here the weak solubility limit is called solid density asymptotics by Ortoleva (1994). Details of the use of the weak solubility approximation can be found in Fowler and Yang (2003).

Compaction

Interest in compaction is motivated by its occurrence in sedimentary basins, and also by issues of subsidence due to groundwater or natural gas extraction (see, for example, Baú et al. 2000). The constitutive law used here for effective pressure is that of Smith (1971); it mimics the normal consolidation behaviour of compacting sediments (such as soils), and is further discussed by Audet and Fowler (1992) and Jones (1994).

Athy's law comes from the paper by Athy (1930). Smith (1971) advocates the use of the high exponent $m = 8$ in (2.128). Further details of the asymptotic solution of the compaction profiles are given by Fowler and Yang (1998). Freed and Peacor (1989) show examples of the flattened porosity profiles at depth.

Early work on pressure solution in sedimentary basins was by Angevine and Turcotte (1983) and Birchwood and Turcotte (1984). More recently, Fowler and Yang (1999) derived the viscous compaction law. An extension to viscoelastic compaction has been studied by Yang (2000).

Seals

One process which we have not described is the formation of high pressure seals. In certain circumstances, pore pressures undergo fairly rapid jumps across a 'seal', typically at depths of 3000 m. Such jumps cannot be predicted within the confines of a simple compaction theory, and require a mechanism for pore-blocking. Mineralisation is one such mechanism, as some seals are found to be mineralised with calcite and silica (Hunt 1990). In fact, a generalisation of the clay diagenesis model to allow for calcite precipitation could be used for this purpose. As it stands, (**??**) predicts a source for ϕ, but mineralisation would cause a corresponding sink term. Reduction of ϕ leads to reduction of diffusive transport, and the feedback is self-promoting. Problems of this type have been studied by Ortoleva (1994), for example.

Exercises

2.1 Show that for a porous medium idealised as a cubical network of tubes, the permeability is given (approximately) by $k = d_p^2\phi^2/72\pi$, where d_p is the grain size. How is the result modified if the pore space is taken to consist of planar sheets between identical cubical blocks? (The volume flux per unit width between two parallel plates a distance h apart is $-h^3p'/12\mu$, where p' is the pressure gradient.)

2.2 A sedimentary rock sequence consists of two type of rock with permeabilities k_1 and k_2. Show that in a unit with two horizontal layers of thickness d_1 and d_2, the effective horizontal permeability (parallel to the bedding plane) is

$$k_\parallel = k_1 f_1 + k_2 f_2,$$

where $f_i = d_i/(d_1 + d_2)$, whereas the effective vertical permeability is given by

$$k_\perp^{-1} = f_1 k_1^{-1} + f_2 k_2^{-1}.$$

Show how to generalise this result to a sequence of n layers of thickness $d_1, \ldots, d_n$.

Hence show that the effective permeabilities of a thick stratigraphic sequence containing a distribution of (thin) layers, with the proportion of layers having permeabilities in $(k, k + dk)$ being $f(k)dk$, are given by

$$k_\parallel = \int_0^\infty kf(k)\,dk, \quad k_\perp^{-1} = \int_0^\infty \frac{f(k)\,dk}{k}.$$

2.3 Groundwater flows between an impermeable basement at $z = h_b(x, y, t)$ and a phreatic surface at $z = z_p(x, y, t)$. Write down the equations governing the flow, and by using the Dupuit approximation, show that the saturated depth h satisfies

$$\phi h_t = \frac{k\rho g}{\mu} \boldsymbol{\nabla}.[h \boldsymbol{\nabla} z_p],$$

where $\boldsymbol{\nabla} = (\partial/\partial x, \partial/\partial y)$. Deduce that a suitable time scale for flows in an aquifer of typical depth h_0 and extent l is $t_{gw} = \phi\mu l^2/k\rho g h_0$.

I live a kilometer from the river, on top of a layer of sediments 100 m thick (below which is impermeable basement). What sort of sediments would those need to be if the river responds to rainfall at my house within a day; within a year?

2.4 A two-dimensional earth dam with vertical sides at $x = 0$ and $x = l$ has a reservoir on one side ($x < 0$) where the water depth is h_0, and horizontal dry land on the other side, in $x > l$. The dam is underlain by an impermeable basement at $z = 0$.

Write down the equations describing the saturated groundwater flow, and show that they can be written in the dimensionless form

$$u = -p_x, \quad \varepsilon^2 w = -(p_z + 1),$$

$$p_{zz} + \varepsilon^2 p_{xx} = 0,$$

and define the parameter ε. Write down suitable boundary conditions on the impermeable basement, and on the phreatic surface $z = h(x, t)$.

Assuming $\varepsilon \ll 1$, derive the Dupuit-Forchheimer approximation for h,

$$h_t = (hh_x)_x \quad \text{in} \quad 0 < x < 1.$$

Show that a suitable boundary condition for h at $x = 0$ (the dam end) is

$$h = 1 \quad \text{at} \quad x = 0.$$

Now define the quantity

$$U = \int_0^h p \, dz,$$

and show that the horizontal flux

$$q = \int_0^h u \, dz = -\frac{\partial U}{\partial x}.$$

Hence show that the conditions of hydrostatic pressure at $x = 0$ and constant (atmospheric) pressure at $x = 1$ (the seepage face) imply that

$$\int_0^1 q \, dx = \tfrac{1}{2}.$$

Deduce that, if the Dupuit approximation for the flux is valid all the way to the toe of the dam at $x = 1$, then $h = 0$ at $x = 1$, and show that in the steady state, the (dimensional) discharge at the seepage face is

$$q_D = \frac{k\rho g h_0^2}{2\mu l}.$$

Supposing the above description of the solution away from the toe to be valid, show that a possible boundary layer structure near $x = 1$ can be described by writing

$$x = 1 - \varepsilon^2 X, \quad h = \varepsilon H, \quad z = \varepsilon Z, \quad p = \varepsilon P,$$

and write down the resulting leading order boundary value problem for P.

2.5 I get my water supply from a well in my garden. The well is of depth h_0 (relative to the height of the water table a large distance away) and radius r_0. Show that the Dupuit approximation for the water table height h is

$$\phi \frac{\partial h}{\partial t} = \frac{k\rho g}{\mu} \frac{1}{r} \frac{\partial}{\partial r} \left(rh \frac{\partial h}{\partial r} \right).$$

If my well is supplied from a reservoir at $r = l$, where $h = h_0$, and I withdraw a constant water flux q_0, find a steady solution for h, and deduce that my well will run dry if

$$q_0 > \frac{\pi k\rho g h_0^2}{\mu \ln[l/r_0]}.$$

Use plausible values to estimate the maximum yield (gallons per day) I can use if my well is drilled through sand, silt or clay, respectively.

2.6 A volume V of effluent is released into the ground at a point ($r = 0$) at time t. Use the Dupuit approximation to motivate the model

$$\phi \frac{\partial h}{\partial t} = \frac{k\rho g}{\mu} \frac{1}{r} \frac{\partial}{\partial r} \left(rh \frac{\partial h}{\partial r} \right),$$

$$h = h_0 \ \text{ at } \ t = 0, \ r > 0,$$

$$\int_0^\infty r(h - h_0) dr = V/2\pi, \ t > 0,$$

where h_0 is the initial height of the water table above an impermeable basement. Find suitable similarity solutions in the two cases (i) $h_0 = 0$ (ii) $h_0 > 0$, $h - h_0 \ll h_0$, and comment on the differences you find.

2.7 Fluid flows through a porous medium in the x direction at a linear velocity U. At $t = 0$, a contaminant of concentration c_0 is introduced at $x = 0$. If the longitudinal dispersivity of the medium is D, write down the equation which

determines the concentration c in $x > 0$, together with suitable initial and boundary conditions. Hence show that c is given by

$$\frac{c}{c_0} = \frac{1}{2}\left[\operatorname{erfc}\left\{\frac{x-Ut}{2\sqrt{Dt}}\right\} + \exp\left(\frac{Ux}{D}\right)\operatorname{erfc}\left\{\frac{x+Ut}{2\sqrt{Dt}}\right\}\right],$$

where

$$\operatorname{erfc}\xi = \frac{2}{\sqrt{\pi}}\int_{\xi}^{\infty} e^{-s^2}\,ds.$$

[*Hint: you might try Laplace transforms, or else simply verify the result.*]

Show that for large ξ, $\operatorname{erfc}\,\xi = e^{-\xi^2}\left[\frac{1}{\sqrt{\pi}\xi} + \ldots\right]$, and deduce that if $x = Ut + 2\sqrt{Dt}\,\eta$, with $\eta = O(1)$, then

$$\frac{c}{c_0} \approx \frac{1}{2}\operatorname{erfc}\eta + O\left(\frac{1}{\sqrt{t}}\right).$$

Hence show that at a fixed station $x = X$ far downstream, the measured profile is approximately given by

$$c \approx c_0\left[1 - \frac{1}{2}\operatorname{erfc}\left\{\frac{1}{2}\left(\frac{U^3}{DX}\right)^{1/2}\left(t - \frac{X}{U}\right)\right\}\right].$$

This is called the breakthrough curve, and indicates that dispersion causes breakthrough to occur over a time interval (at large distance) of order $\Delta t_b = (DX/U^3)^{1/2}$. If $D \approx aU$, show that the ratio of Δt_b to $t_b = X/U$ is $\Delta t_b/t_b \sim (a/X)^{1/2}$.

2.8 Rain falls steadily at a rate q (volume per unit area per unit time) on a soil of saturated hydraulic conductivity K_0 ($= k_0\rho_w g/\mu$, where k_0 is the saturated permeability). By plotting the relative permeability k_{rw} and suction characteristic $\sigma\psi/d$ as functions of S (assuming a residual liquid saturation S_0), show that a reasonable form to choose for $k_{rw}(\psi)$ is $k_{rw} = e^{-c\psi}$. If the water table is at depth h, show that, in a steady state, ψ is given as a function of the dimensionless depth $z^* = z/z_c$, where $z_c = \sigma/\rho_w g d$ (σ is the surface tension, d the grain size) by

$$h^* - z^* = \tfrac{1}{2}\psi - \frac{1}{c}\ln\left[\frac{\sinh\{\frac{1}{2}(\ln\frac{1}{q^*} - c\psi)\}}{\sinh\{\frac{1}{2}\ln\frac{1}{q^*}\}}\right],$$

where $h^* = h/z_c$, providing $q^* = q/K_0 < 1$. Deduce that if $h \gg z_c$, then $\psi \approx \frac{1}{c}\ln\frac{1}{q^*}$ near the surface. What happens if $q > K_0$?

2.9 Derive the Richards equation

$$\phi\frac{\partial S}{\partial t} = -\frac{\partial}{\partial z}\left[\frac{k_0}{\mu}k_{rw}(S)\left\{\frac{\partial p_c}{\partial z} + \rho_w g\right\}\right]$$

for one-dimensional infiltration of water into a dry soil, explaining the meaning of the terms, and giving suitable boundary conditions when the surface flux q is prescribed. Show that if the surface flux is large compared with $k_0\rho_w g/\mu$, where k_0 is the saturated permeability, then the Richards equation can be approximated, in suitable non-dimensional form, by a nonlinear diffusion equation of the form

$$\frac{\partial S}{\partial t} = \frac{\partial}{\partial z}\left[D\frac{\partial S}{\partial z}\right].$$

Show that, if $D = S^m$, a similarity solution exists in the form

$$S = t^\alpha F(\eta), \quad \eta = z/t^\beta,$$

where $\alpha = \dfrac{1}{m+2}$, $\beta = \dfrac{m+1}{m+2}$, and F satisfies

$$(F^m F')' = \alpha F - \beta\eta F', \quad F^m F' = -1 \ \text{ at } \ \eta = 0, \quad F \to 0 \ \text{ as } \ \eta \to \infty.$$

Deduce that

$$F^m F' = -(\alpha+\beta)\int_\eta^{\eta_0} F\,d\eta - \beta\eta F,$$

where η_0 (which may be ∞) is where F first reaches zero. Deduce that $F' < 0$, and hence that η_0 must be finite, and is determined by

$$\int_0^{\eta_0} F\,d\eta = \frac{1}{\alpha+\beta}.$$

What happens for $t > F(0)^{-1/\alpha}$?

2.10 Write down the equations describing one-dimensional consolidation of wet sediments in terms of the variables ϕ, v^s, v^l, p, p_e, these being the porosity, solid and liquid (linear) velocities, and the pore and effective pressures. Neglect the effect of gravity.

Saturated sediments of depth h lie on a rigid but permeable (to water) basement, through which a water flux W is removed. Show that

$$v^s = \frac{k}{\mu}\frac{\partial p}{\partial z} - W,$$

and deduce that ϕ satisfies the equation

$$\frac{\partial \phi}{\partial t} = \frac{\partial}{\partial z}\left[(1-\phi)\left\{\frac{k}{\mu}\frac{\partial p}{\partial z} - W\right\}\right].$$

If the sediments are overlain by water, so that $p =$ constant (take $p = 0$) at $z = h$, and if $\phi = \phi_0 + p/K$, where the compressibility K is large (so $\phi \approx \phi_0$), show that a suitable reduction of the model is

$$\frac{\partial p}{\partial t} - W\frac{\partial p}{\partial z} = c\frac{\partial^2 p}{\partial z^2},$$

where $c = K(1-\phi_0)k/\mu$, and $p = 0$ on $z = h$, $p_z = \mu W/k$. Non-dimensionalise the model using the length scale h, time scale h^2/c, and pressure scale $\mu W h/k$. Hence describe the solution if the parameter $\varepsilon = \mu W h/k$ is small, and find the rate of surface subsidence. What has this to do with Venice?

2.11 Write down a model for vertical flow of two immiscible fluids in a porous medium. Deduce that the saturation S of the wetting phase satisfies the equation

$$\phi\frac{\partial S}{\partial t} + \frac{\partial}{\partial z}\left[M_{\text{eff}}\left\{\frac{q}{M_{nw}} + g\Delta\rho\right\}\right] = -\frac{\partial}{\partial z}\left[M_{\text{eff}}\frac{\partial p_c}{\partial z}\right],$$

where z is a coordinate pointing *downwards*,

$$p_c = p_{nw} - p_w, \quad \Delta\rho = \rho_w - \rho_{nw}, \quad M_{\text{eff}}^{-1} = (M_w^{-1} + M_{nw}^{-1}),$$

q is the total downward flux, and the suffixes w and nw refer to the wetting and non-wetting fluid respectively. Define the phase mobilities M_i. Give a criterion on the capillary suction p_c which allows the Buckley-Leverett approximation to be made, and show that for $q = 0$ and $\mu_w \gg \mu_{nw}$, waves typically propagate downwards and form shocks. What happens if $q \neq 0$? Is the Buckley-Leverett approximation realistic — e.g. for air and water in soil? (Assume $p_c \sim 2\gamma/r_p$, where $\gamma = 70$ mN m^{-1}, and r_p is the pore radius: for clay, silt and sand, take $r_p = 1\ \mu$, $10\ \mu$, $100\ \mu$, respectively.)

2.12 A model for snow melt run-off is given by the following equations:

$$u = \frac{k}{\mu}\left[\frac{\partial p_c}{\partial z} + \rho_l g\right],$$

$$k = k_0 S^3,$$

$$\phi\frac{\partial S}{\partial t} + \frac{\partial u}{\partial z} = 0,$$

$$p_c = p_0\left(\frac{1}{S} - S\right).$$

Explain the meaning of the terms in these equations, and describe the assumptions of the model.

The intrinsic permeability k_0 is given by

$$k_0 = 0.077\ d^2 \exp[-7.8\ \rho_s/\rho_l],$$

where ρ_s and ρ_l are snow and water densities, and d is grain size. Take $d = 1$ mm, $\rho_s = 300$ kg m^{-3}, $\rho_l = 10^3$ kg m^{-3}, $p_0 = 1$ kPa, $\phi = 0.4$, $\mu = 1.8 \times 10^{-3}$ Pa s, $g = 10$ m s^{-2}, and derive a non-dimensional model for melting of a one metre thick snow pack at a rate (i.e. u at the top surface $z = 0$) of 10^{-6} m s^{-1}. Determine whether capillary effects are small; describe the nature of the model equation, and find an approximate solution for the melting of an initially dry snowpack. What is the (meltwater flux) run-off curve?

2.13 Consider the following model, which represents the release of a unit quantity of groundwater at $t = 0$ in an aquifer $-\infty < x < \infty$, when the Dupuit approximation is used:

$$h_t = (hh_x)_x,$$

$$h = 0 \text{ at } t = 0,\ x \neq 0,$$

$$\int_{-\infty}^{\infty} h\, dx = 1$$

(i. e., $h = \delta(x)$ at $t = 0$). Show that a similarity solution to this problem exists in the form

$$h = t^{-1/3} g(\xi), \quad \xi = x/t^{1/3},$$

and find the equation and boundary conditions satisfied by g. Show that the water body spreads at a finite rate, and calculate what this is.

Formulate the equivalent problem in three dimensions, and write down the equation satisfied by the similarity form of the solution, assuming cylindrical symmetry. Does this solution have the same properties as the one-dimensional solution?

2.14 The tensor D_{ij} $(i, j = 1, 2, 3)$ has three invariants

$$D_I = D_{ii}, \quad D_{II} = D_{ij}D_{ij}, \quad D_{III} = D_{ij}D_{jk}D_{ki}.$$

(Summation over repeated indices is implied.) Show that the invariants of the tensor

$$D_{ij} = \alpha_\perp u \delta_{ij} + (\alpha_\parallel - \alpha_\perp)\frac{u_i u_j}{u},$$

where $u = |\mathbf{u}|$ and δ_{ij} is the Kronecker delta ($= 1$ if $i = j$, $= 0$ if $i \neq j$), are the same as those of the tensor

$$\mathbf{D} = \begin{pmatrix} \alpha_\parallel u & 0 & 0 \\ 0 & \alpha_\perp u & 0 \\ 0 & 0 & \alpha_\perp u \end{pmatrix}.$$

.

2.15 Suppose that a doubly porous medium consists of a periodic sequence of blocks M with boundaries (fractures) ∂M. The concentration of a chemical reactant c is taken to be a function of the fast space variable $\mathbf{X}$ and the slow space variable $\mathbf{x} = \varepsilon\mathbf{X}$, and we assume that $c = \bar{c}(\mathbf{x}) + \varepsilon^2 c_k(\mathbf{X})$, where the suffix k refers to fractures (f) or matrix block (m).

Let $G(\mathbf{X}, \mathbf{Y})$ be a Green's function satisfying

$$\nabla^2_{\mathbf{Y}} G = \delta(\mathbf{X} - \mathbf{Y}) \text{ in } M, \quad G = 0 \text{ for } \mathbf{Y} \in \partial M,$$

and suppose that

$$\begin{aligned} \nabla^2_{\mathbf{X}}\psi &= 0 \quad \text{in} \quad M, \\ \psi &= \chi \quad \text{on} \quad \partial M. \end{aligned}$$

Show that

$$\psi = \int_{\partial M} \frac{\partial G(\mathbf{X},\mathbf{Y})}{\partial N_Y}\chi(\mathbf{Y})\, dS(\mathbf{Y}),$$

where $\dfrac{\partial G}{\partial N_Y} = \mathbf{n}.\boldsymbol{\nabla}_{\mathbf{Y}} G(\mathbf{X},\mathbf{Y})$, and hence show that

$$\left.\frac{\partial \psi}{\partial N}\right|_{\partial M} = \int_{\partial M} K(\mathbf{X},\mathbf{Y})\,\chi(\mathbf{Y})\, dS(\mathbf{Y}),$$

where

$$K(\mathbf{X},\mathbf{Y}) = \frac{\partial^2 G(\mathbf{X},\mathbf{Y})}{\partial N_X \partial N_Y}.$$

Now suppose that the fluctuating matrix and fracture concentrations of the chemical reactant are given by

$$\nabla^2_{\mathbf{X}} c_m = Pe\left[\frac{\partial \bar{c}}{\partial t} + \mathbf{u}_m.\boldsymbol{\nabla}_{\mathbf{x}}\bar{c}\right] + Pe\,\Lambda\bar{c} - \nabla^2_{\mathbf{x}}\bar{c} \equiv R_m \quad \text{in} \quad M,$$

subject to $c_m = c_f$ on ∂M, and

$$\nabla^2_{\mathbf{X}} c_f - \frac{(1-\phi_f)}{\phi_f}\left.\frac{\partial c_m}{\partial N}\right|_{\partial M} = Pe\left[\frac{\partial \bar{c}}{\partial t} + \mathbf{u}_f.\boldsymbol{\nabla}_{\mathbf{x}}\bar{c}\right] + Pe\,\Lambda\bar{c} - \nabla^2_{\mathbf{x}}\bar{c} \equiv R_f \quad \text{on} \quad \partial M,$$

subject to conditions of periodicity with zero mean.

Show that, if we define c^* to be the solution of

$$\nabla^2_{\mathbf{X}} c^* = 1 \quad \text{in} \quad M,$$

with

$$c^* = 0 \quad \text{on} \quad \partial M,$$

then

$$c_m = \int_{\partial M} \frac{\partial G(\mathbf{X},\mathbf{Y})}{\partial N_Y} c_f(\mathbf{Y})\, dS(\mathbf{Y}) + R_m c^*,$$

and deduce that, for $\mathbf{X} \in M$,

$$\phi_f \nabla^2_{\mathbf{X}} c_f - (1-\phi_f)\int_{\partial M} K(\mathbf{X},\mathbf{Y}) c_f(\mathbf{Y})\, dS(\mathbf{Y}) = \phi_f R_f + (1-\phi_f) R_m \left.\frac{\partial c^*}{\partial N}\right|_{\partial M}.$$

By integrating this equation over ∂M, show that the condition of periodicity of c_f implies that the equation to determine $\bar{c}$ is

$$Pe\left[\frac{\partial \bar{c}}{\partial t} + \mathbf{u}.\boldsymbol{\nabla}_{\mathbf{x}}\bar{c}\right] = \nabla^2_{\mathbf{x}}\bar{c} - Pe\,\Lambda\bar{c},$$

where $\mathbf{u} = \phi_f \mathbf{u}_f + (1-\phi_f)\mathbf{u}_m$.

2.16 The reaction rates in the reactions

$$\begin{aligned} \mathrm{S}^S &\xrightarrow{r_1} \mathrm{X}^L + n\mathrm{H_2O}, \\ \mathrm{KFs} &\xrightarrow{r_2} \mathrm{K}^{+L} + \mathrm{AlO_2}^{-L} + s\,\mathrm{SiO_2}^L, \\ \mathrm{K}^{+L} + \mathrm{AlO_2}^{-L} + f\mathrm{X}^L &\xrightarrow{r_3} f\mathrm{I}^S + \mathrm{SiO_2}^L, \\ \mathrm{SiO_2}^L &\underset{r_4^-}{\overset{r_4^+}{\rightleftharpoons}} \mathrm{Qz}, \end{aligned}$$

are related by

$$\begin{aligned} r_1 &\approx f r_3, \\ r_2 &\approx r_3, \\ r_4^+ - r_4^- &\approx (s+1) r_3. \end{aligned}$$

The reaction rate r_3 is given by

$$r_3 = \phi_I R_3 \left[1 - \frac{f r_3}{\phi_S R_1}\right] \left[1 - \frac{r_3}{\phi_F R_2 \left\{\theta_L - 1 - \dfrac{(s+1) r_3}{\phi_Q R_4^+}\right\}}\right],$$

where ϕ_i are porosities, R_k are rate factors (such that $r_k \propto R_k$), and the stoichiometric constants f and s, and the constant θ_L, may be taken as $O(1)$ (and $\theta_L > 1$). Show that r_3 can be written explicitly in the form

$$\frac{2}{r_3} = \left\{ \frac{1}{(\theta_L - 1)\gamma_F} + \frac{s+1}{(\theta_L - 1)\gamma_Q} + \frac{f}{\gamma_S} + \frac{1}{\gamma_I} \right\}$$

$$+ \left[\left\{ \frac{1}{(\theta_L - 1)\gamma_F} + \frac{s+1}{(\theta_L - 1)\gamma_Q} + \frac{f}{\gamma_S} + \frac{1}{\gamma_I} \right\}^2 + \frac{4(s+1)}{(\theta_L - 1)\gamma_Q} \left(\frac{f}{\gamma_S} + \frac{1}{\gamma_I} \right) \right]^{1/2},$$

where the coefficients γ_Y represent the porosity weighted rate factors, i. e.,

$$\gamma_I = \phi_I R_3, \quad \gamma_S = \phi_S R_1, \quad \gamma_Q = \phi_Q R_4^+, \quad \gamma_F = \phi_F R_2.$$

Deduce that the slowest reaction of the four (as measured by γ_Y) controls the overall rate, and give explicit approximations for r_3 for each of the consequent four possibilities.

Chapter 3

Convection

Convection is the fluid motion induced by buoyancy; buoyancy is the property of a fluid whereby its density depends on external properties. The most common form of convection is *thermal convection*, which occurs due to the dependence of density on temperature: warm fluid is light, and therefore rises. Everyday examples of this are the circulation induced by a convector heater, or the motion which can be seen in a saucepan of oil when it is heated. (In the latter case, one can see convection rolls in the fluid, regular but time-dependent.) Another common form of convection is *compositional convection*, which is induced by density changes dependent on composition. An example of this occurs during the formation of sea ice in the polar regions. As salty sea water freezes, it rejects the salt (the ice is almost pure water substance), and the resulting salty water is denser than the sea water from which it forms, and thus induces a convective motion below the ice. Below, we discuss three geophysical examples from convection, but convection is everywhere: it drives the oceanic circulation, it drives the atmospheric circulation, it causes thunderstorms, it occurs in glass manufacture, in a settling pint of Guinness, in back boilers, in solar panels. And, it has formed the thematic core of the subject of geophysical fluid dynamics for almost a century.

3.1 Mantle convection

Most people have heard of continental drift, the process whereby the Earth's continents drift apart relative to each other. The Atlantic Ocean is widening at the rate of several centimetres a year, the crashing of India into Asia over the last 50 My (fifty million years) has caused the continuing uplift of the Himalayas, Scotland used to be joined to Newfoundland. The continents ride, like rafts of debris, on the tectonic plates of the Earth, which separate at mid-ocean ridges and converge at subduction zones. The theory of plate tectonics, which originated with the work of Wegener and Holmes in the early part of the twentieth century, and which was finally accepted by geophysicists in the 'plate tectonics revolution of the 1960's, describes the surface of the Earth as being split up into some thirteen major tectonic plates: see figure

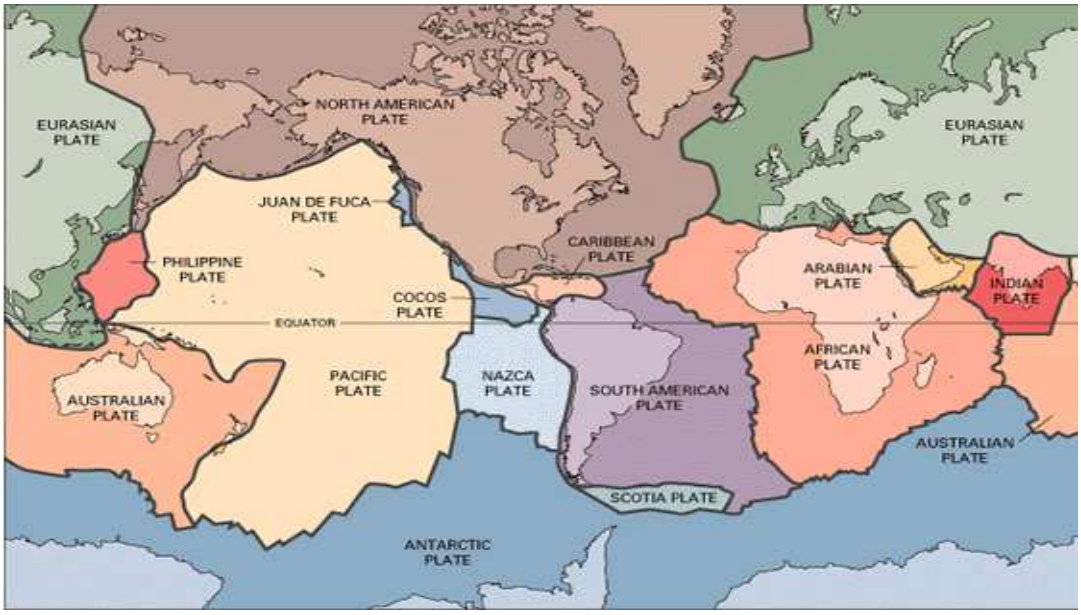

Figure 3.1: The tectonic plates of the Earth.

3.1. These plates move relative to each other across the surface, and this motion is the surface manifestation of a convective motion in the Earth's *mantle*, which is the part of the Earth from the surface to a depth of about 3,000 kilometres, and which consists of an assemblage of polycrystalline silicate rocks. Upwelling occurs at mid-ocean ridges, for example the mid-Atlantic ridge which passes through Iceland, and the East Pacific Rise off the coast of South America, which passes through the Galapagos Islands. The plates sink into the mantle at subduction zones, which adjoin continental boundaries, and which are associated with the presence of oceanic trenches.

The plates are so called because they are conceived of as moving quasi-rigidly. They are in fact the cold upper thermal boundary layers of the convective motion, in-

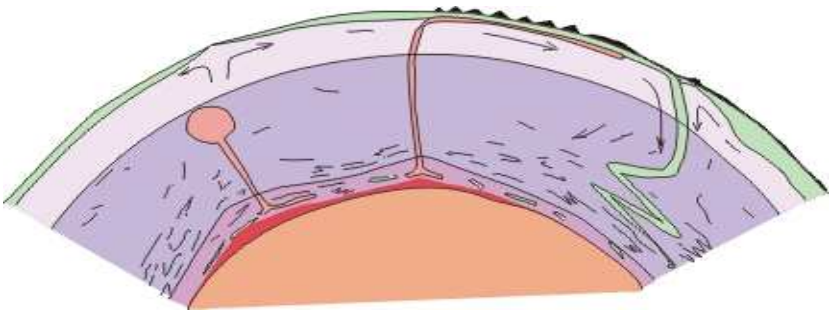

Figure 3.2: A cartoon of mantle convection. We see plumes, mid-ocean ridges, subducting slabs.

dicated schematically in figure 3.2, and are plate-like because the strong temperature dependence of mantle viscosity renders these relatively cold rocks extremely viscous. One may wonder how the mantle moves at all, consisting as it does of mostly solid polycrystalline rocks. In fact, solids will deform just as fluids do when subjected to stress. The deformation is enabled by the migration of dislocations within the crystalline lattice of the solid grains of the rock. The effective viscosity of the Earth's mantle is a whopping 10^{21} Pa s; this is about eight orders of magnituse greater than the viscosity of ice, and twenty-four orders greater than the viscosity of water.

The reason that the mantle convects is that the Earth is cooling. The primordial heat of formation has gradually been lost over the Earth's history, but the central core of the planet is still very hot; some six thousand degrees Celsius at the centre of the planet. This heat from the core is instrumental in heating the mantle from below, and driving the convective flow. Radioactive heating also contributes to an extent which is not certain, but which is thought to be significant.

3.2 The Earth's core

Part of the heat which drives mantle convection is derived from cooling the Earth's core. The core is the part of the Earth which lies between its centre and the mantle. Like the mantle, it is also some three thousand kilometres deep, and consists of a molten outer core of iron, alloyed with some lighter element, usually thought to be sulphur or oxygen, in a concentration of some 10%. The inner core is solid (pure) iron, of radius 1,000 km. It is generally thought that the core was initially molten throughout, and that the inner core has gradually solidified from the outer core over the course of geological time. It is the consequent release of latent heat which, at least partly, powers mantle convection.

One may wonder how the outer core can be liquid, and the inner core solid, if the inner core is hotter (as it must be). The reason for this is that the solidification temperature (actually the liquidus temperature, see below) depends on pressure, through the Clapeyron effect. This is the effect whereby a pressure cooker works: the boiling temperature increases with pressure, and similarly, the solidification temperature of the outer core iron alloy increases with pressure, and thus also depth. Thus, the inner core can be below the solidification temperature because of the greater pressure there.

The convection in the outer core is partly due to the dependence of density on temperature, but the primary dependence is, as often the case when composition varies, due to the dependence of density on the concentration of sulphur (or oxygen). In order to understand how the solidification of the inner core leads to convection, we need to understand the general thermodynamic way in which melting and solidification occur in multi-component materials. This is illustrated in figure 3.3, which indicates how the solidification temperatures vary with composition in a two-component melt. At a given temperature, there are two curves which describe the concentrations of the solid and liquid, when these are in thermodynamic equilibrium with each other. These two curves are called the solidus and liquidus, respectively. Often there are two sets of

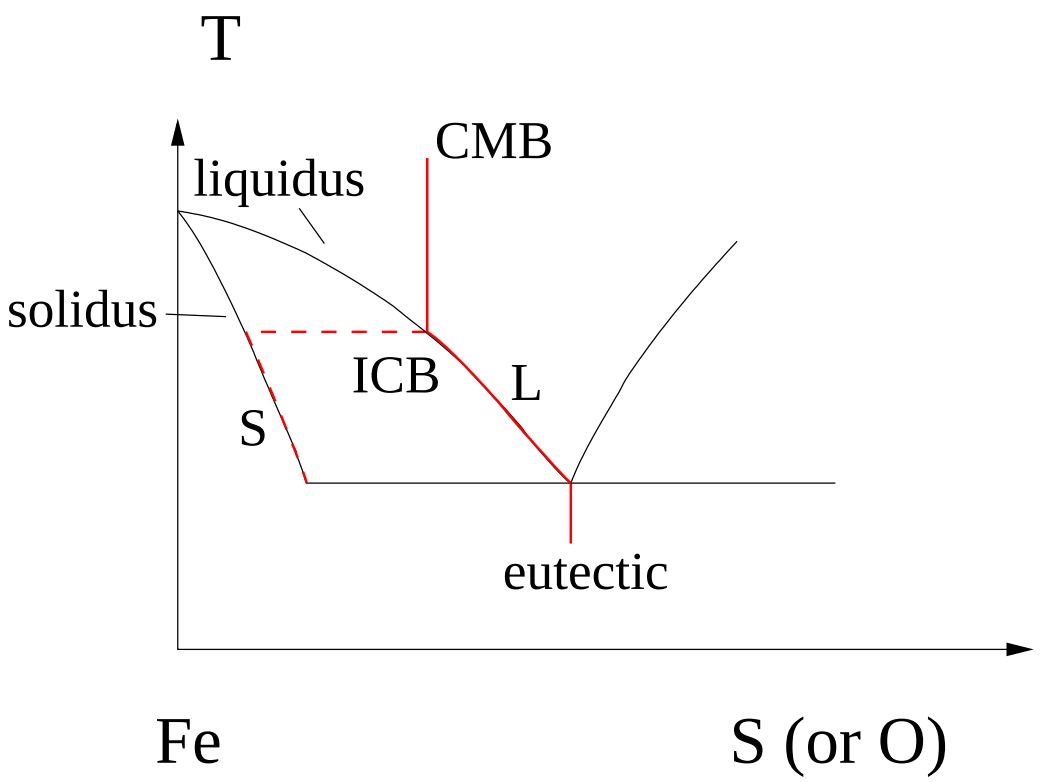

Figure 3.3: Typical phase diagram for a two-component alloy with a eutectic point. When the liquid reaches the liquidus (L), the resulting solid has the concentration of the solidus (S). When the liquid reaches the eutectic point, two solids, iron-rich and sulphur-rich respectively, will be formed.

solidus and liquidus curves, and they meet at a point called the eutectic point. The way in which a liquid alloy solidifies is then indicated by the red line in figure 3.3. In the outer core, the composition is relatively constant, but the temperature decreases (relative to the liquidus) from the core-mantle boundary (CMB) to the inner core boundary (ICB), where solidification occurs. (The phase diagram is indicated as if at constant pressure; in reality, the curves will also vary with pressure.)

At this temperature, the solid which crystallises has the solidus concentration, which is richer in iron than the liquid, and so as the temperature cools during freezing, the liquid concentration of sulphur or oxygen increases because of its rejection at the freezing interface. It is this source of buoyancy which provides the driving force for compositional convection.

Actually, it is typically the case that when alloys solidify, they do not form a solid with a clear interface. Rather, such a situation is typically *morphologically unstable*, and a dendritic mush consisting of a solid–liquid mixture is formed, as shown in figure 3.4. The convection caused by the release of light fluid now occurs throughout the mush, and leads to the formation of narrow 'chimneys', from which plumes emerge.

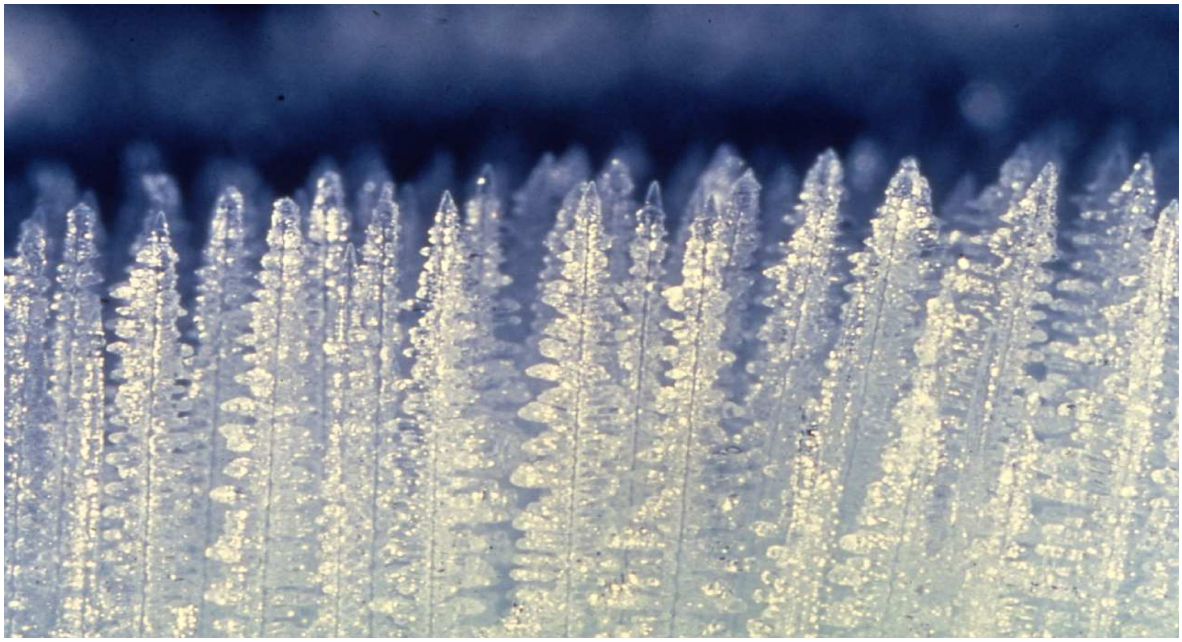

Figure 3.4: A dendritic mush in the solidification of ammonium chloride in the laboratory. Convection occurs within the mush, leading to the formation of 'chimneys' which act as sources of plumes in the residual melt. Photo courtesy of Grae Worster.

In the Earth's core, it is this convection which forms the magnetic field. Convection in an electrically conducting fluid causes a magnetic field to grow, providing the magnetic diffusivity is sufficiently small, through the action of the Lorentz force. The study of such instabilities is a central part of the subject of magnetohydrodynamics.

3.3 Magma chambers

Our final example of convection arises in the formation and cooling of magma chambers. When mantle rock upwells, either at mid-ocean ridges, or in isolated thermal plumes such as that below Hawaii, the slight excess temperature causes the rock to partially melt. It is thought that the melt fraction can then ascend through the residual porous matrix, forming rivulets and channels which allow the escape of the magma through the lithosphere to the crust.[1] As the magma ascends into the crust, it can typically encounter unconformities, where the rock types alter, and where the density may be less than that of the magma. In that case, the magma will stop rising, but will spread laterally, simultaneously uplifting the overlying strata. Thus forms the *laccolith*, a magmatic intrusion, and over the course of time such intrusions, or

[1]The *lithosphere* is the cold surface boundary layer of the convecting mantle, of depth some 100 km in the oceanic mantle, somewhat greater beneath continents; the crust is a relatively thin layer of rocks near the surface, formed through partial melting of the mantle and the resulting volcanism.

Figure 3.5: Graded layering in the Skaergaard intrusion. Photograph courtesy of Kurt Hollocher.

magma chambers, will solidify, forming huge cauldrons of rock which may later be exposed at the Earth's surface.

Convection undoubtedly occurs in such chambers, which may be tens of kilometres in extent. The hot magma is continuously chilled at the roof and sides of the chamber, and this leads to convective currents continually draining towards the floor of the chamber. There they will accumulate, leading to a cold, crystal-rich layer lying stagnant below the convecting upper portion. This is essentially the filling box mechanism which is discussed further below.

Magmas are multi-component alloys, and their convective solidification can lead to various exotic phenomena. The phase diagram of the type in figure 3.3 causes chemical differentiation on the large scale (in metal alloy castings this is called macrosegregation). For example, in an olivine–plagioclase magma, the heavy olivine will crystallise out first, and the crystals may settle to the base of the chamber. The residual liquid is then plagioclase-rich and lighter. So the end result would be a chamber having two distinct layers. Successive injections of magma may then lead to a sequence of such layers, as is seen in the Scottish island of Rum, and this has been suggested as an explanation for these particular layers.

Other magma chambers show layering at a much finer scale, and the origin of these layers is a mystery. An example is shown in figure 3.5. The layers are reminiscent of double-diffusive layering, which we discuss in section 3.11, but efforts to build a theory round this idea, or indeed any other, have so far not met with success.

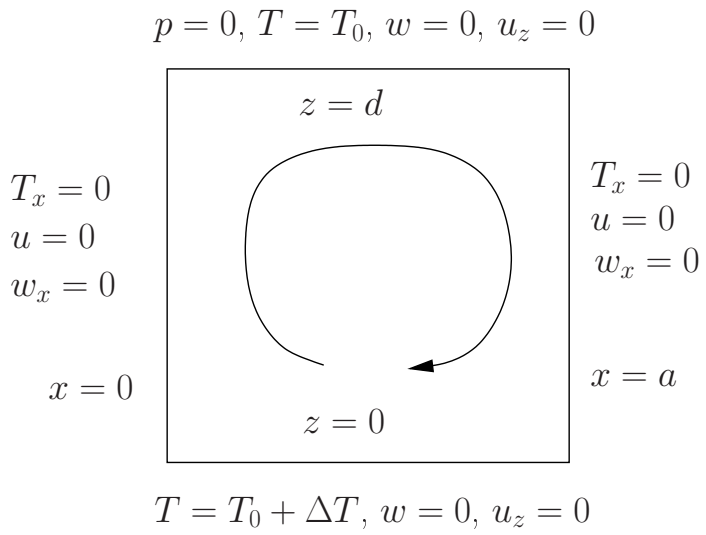

Figure 3.6: Geometry of a convection cell.

3.4 Rayleigh–Bénard convection

The simplest model of convection is the classical Rayleigh-Bénard model in which a layer of fluid is heated from below, by application of a prescribed temperature difference across the layer. Depending on the nature of the boundaries, one may have a no slip condition or a no shear stress condition applied at the bounding surfaces. For the case of mantle convection, one conceives of both the oceans (or atmosphere) and the underlying fluid outer core as exerting no stress on the extremely viscous mantle, so that no stress conditions are appropriate, and in fact it turns out that this is the simplest case to consider. The geometry of the flow we consider is shown in figure 3.6. It is convenient to assume lateral boundaries, although in a wide layer, these simply represent the convection cell walls, and can be an arbitrary distance apart.

The equations describing the flow are the Navier-Stokes equations, allied with the energy equation and an equation of state, and can be written in the form

$$\begin{aligned} \rho_t + \boldsymbol{\nabla}.(\rho\mathbf{u}) &= 0, \\ \rho[\mathbf{u}_t + (\mathbf{u}.\boldsymbol{\nabla})\mathbf{u}] &= -\boldsymbol{\nabla}p - \rho g\mathbf{k} + \mu\nabla^2\mathbf{u}, \\ \rho c_p[T_t + \mathbf{u}.\boldsymbol{\nabla}T] &= k\nabla^2 T, \\ \rho &= \rho_0[1 - \alpha(T - T_0)]; \end{aligned} \tag{3.1}$$

in these equations, ρ is the density, $\mathbf{u}$ is the velocity, p is the pressure, g is the acceleration due to gravity, $\mathbf{k}$ is the unit upwards vector, μ is viscosity, c_p is the specific heat, T is temperature, k is thermal conductivity, ρ_0 is the density at the reference temperature T_0 at the surface of the fluid layer, and α is the thermal expansion coefficient. The boundary conditions for the flow are indicated in figure 3.6, and correspond to

prescribed temperature at top and bottom, no flow through the boundaries, and no shear stress at the boundaries. The lateral boundaries represent stress free 'walls', but as mentioned above, these simply indicate the boundaries of the convection cells (across which there is no heat transport, hence the no flux condition for temperature).

To proceed, we non-dimensionalise the variables as follows. We use the convective time scale, and a thermally related velocity scale, and use the depth of the box d as the length scale:

$$\begin{aligned} \mathbf{u} \sim \frac{\kappa}{d}, \quad \kappa = \frac{k}{\rho_0 c_p}, \qquad t \sim \frac{d^2}{\kappa}, \quad \mathbf{x} \sim d, \\ p - [p_0 + \rho_0 g(d-z)] \sim \frac{\mu\kappa}{d^2}, \qquad T - T_0 \sim \Delta T. \end{aligned} \tag{3.2}$$

Here p_0 is the (prescribed) pressure at the surface, which we take as constant. We would also scale $\rho \sim \rho_0$, but in the scaled equations below, the density has been algebraically eliminated. The scaled equations take the form

$$\begin{aligned} -BT_t + \boldsymbol{\nabla}.[(1-BT)\mathbf{u}] &= 0, \\ \frac{1}{Pr}[1-BT][\mathbf{u}_t + (\mathbf{u}.\boldsymbol{\nabla})\mathbf{u}] &= -\boldsymbol{\nabla}p + Ra\,T\mathbf{k} + \nabla^2\mathbf{u}, \\ (1-BT)(T_t + \mathbf{u}.\boldsymbol{\nabla}T) &= \nabla^2 T, \end{aligned} \tag{3.3}$$

and the dimensionless parameters are defined as

$$B = \alpha\Delta T, \quad Pr = \frac{\mu}{\rho_0\kappa}, \quad Ra = \frac{\alpha\rho_0\Delta T g d^3}{\mu\kappa}; \tag{3.4}$$

the parameters Ra and Pr are known as the Rayleigh and Prandtl numbers, respectively. The Prandtl number is a property of the fluid; for air it is 0.7, and for water it is 7. The Rayleigh number is a measure of the strength of the heating. As we shall see, convective motion occurs if the Rayleigh number is large enough, and it becomes vigorous if the Rayleigh number is large. The parameter B might be termed a Boussinesq number, although this is not common usage.

Suppose we think of values typical for a layer of water in a saucepan. We take $d = 0.1$ m, $\mu = 2 \times 10^{-3}$ Pa s, $\Delta T = 100$ K, $\alpha = 3 \times 10^{-5}$ K^{-1}, $\rho_0 = 10^3$ kg m^{-3}, $\kappa = 0.3 \times 10^{-6}$ m^2 s^{-1}, $g = 9.8$ m s^{-2}. Then we have $Pr \approx 7$, $B \approx 3 \times 10^{-3}$, and $Ra \approx 5 \times 10^7$. In this case, we have that $B \ll 1$ and $Ra \gg 1$. This is typically the case. We now make the Boussinesq approximation, which says that $B \ll 1$, and we ignore the terms in B in (3.3). In words, we assume that the density is constant, except in the buoyancy term. The mathematical reason for this exception is that, although $Ra \propto B$ (and so $Ra \to 0$ as $B \to 0$), the actual numerical sizes of the two parameters are typically very different. The adoption of the Boussinesq approximation leads to what are called the Boussinesq equations of thermal convection:

$$\begin{aligned} \boldsymbol{\nabla}.\mathbf{u} &= 0, \\ \frac{1}{Pr}[\mathbf{u}_t + (\mathbf{u}.\boldsymbol{\nabla})\mathbf{u}] &= -\boldsymbol{\nabla}p + \nabla^2\mathbf{u} + Ra\,T\hat{\mathbf{k}}, \\ T_t + \mathbf{u}.\boldsymbol{\nabla}T &= \nabla^2 T, \end{aligned} \tag{3.5}$$

with associated boundary conditions for free slip:

$$\begin{aligned} T = 1, \quad \mathbf{u}.\mathbf{n} = \tau_{nt} = 0 \quad \text{on} \quad z = 0, \\ T = 0, \quad \mathbf{u}.\mathbf{n} = \tau_{nt} = 0 \quad \text{on} \quad z = 1, \end{aligned} \tag{3.6}$$

where τ_{nt} represents the shear stress.

3.4.1 Linear stability

It is convenient to study the problem of the onset of convection in two dimensions (x, z). In this case we can define a stream function ψ which satisfies

$$u = -\psi_z, \quad w = \psi_x. \tag{3.7}$$

(The sign is opposite to the usual convention; for $\psi > 0$ this describes a clockwise circulation.) We eliminate the pressure by taking the curl of the momentum equation $(3.5)_2$, which leads, after some algebra (see also question 3.2), to the pair of equations for ψ and T:

$$\begin{aligned} \frac{1}{Pr}\left[\nabla^2\psi_t + \psi_x\nabla^2\psi_z - \psi_z\nabla^2\psi_x\right] &= Ra\, T_x + \nabla^4\psi, \\ T_t + \psi_x T_z - \psi_z T_x &= \nabla^2 T, \end{aligned} \tag{3.8}$$

with the associated boundary conditions

$$\begin{aligned} \psi = \nabla^2\psi = 0 \quad \text{at} \quad z = 0, 1, \\ T = 0 \quad \text{at} \quad z = 1, \\ T = 1 \quad \text{at} \quad z = 0. \end{aligned} \tag{3.9}$$

In the absence of motion, $\mathbf{u} = \mathbf{0}$, the steady state temperature profile is linear,

$$T = 1 - z, \tag{3.10}$$

and the lithostatic pressure is modified by the addition of

$$p = -\frac{Ra}{2}(1 - z)^2. \tag{3.11}$$

(Even if Ra is large, this represents a small correction to the lithostatic pressure, of relative $O(B)$.) The stream function is just

$$\psi = 0. \tag{3.12}$$

We define the temperature perturbation θ by

$$T = 1 - z + \theta. \tag{3.13}$$

This yields

$$\begin{aligned} \frac{1}{Pr}\left[\nabla^2\psi_t + \psi_x\nabla^2\psi_z - \psi_z\nabla^2\psi_x\right] &= \nabla^4\psi + Ra\,\theta_x, \\ \theta_t - \psi_x + \psi_x\theta_z - \psi_z\theta_x &= \nabla^2\theta, \end{aligned} \tag{3.14}$$

and the boundary conditions are

$$\psi_{zz} = \psi = \theta = 0 \quad \text{on} \quad z = 0, 1. \tag{3.15}$$

In the Earth's mantle, the Prandtl number is large, and we will now simplify the algebra by putting $Pr = \infty$. This assumption does not in fact affect the result which is obtained for the critical Rayleigh number at the onset of convection. The linear stability of the basic state is determined by neglecting the nonlinear advective terms in the heat equation. We then seek normal modes of wave number k in the form

$$\begin{aligned} \psi &= f(z)e^{\sigma t + ikx}, \\ \theta &= g(z)e^{\sigma t + ikx}, \end{aligned} \tag{3.16}$$

whence f and g satisfy (putting $Pr = \infty$)

$$\begin{aligned} (D^2 - k^2)^2 f + ikRa\,g &= 0, \\ \sigma g - ikf &= (D^2 - k^2)g, \end{aligned} \tag{3.17}$$

where $D = d/dz$, and

$$f = f'' = g = 0 \quad \text{on} \quad z = 0, 1. \tag{3.18}$$

By inspection, solutions are

$$f = \sin m\pi z, \quad g = b\sin m\pi z, \tag{3.19}$$

$(n = 1, 2, ...)$ providing

$$\sigma = \frac{k^2 Ra}{(m^2\pi^2 + k^2)^2} - (m^2\pi^2 + k^2), \tag{3.20}$$

which determines the growth rate for the m-th mode of wave number k.

Since σ is real, instability is characterised by a positive value of σ. We can see that σ decreases as m increases; therefore the value $m = 1$ gives the most unstable value of σ. Also, σ is negative for $k \to 0$ or $k \to \infty$, and has a single maximum. Since σ increases with Ra, we see that $\sigma > 0$ (for $m = 1$) if $Ra > Ra_{ck}$, where

$$Ra_{ck} = \frac{(\pi^2 + k^2)^3}{k^2}. \tag{3.21}$$

In turn, this value of the Rayleigh number depends on the selected wave number k. Since an arbitrary disturbance will excite all wave numbers, it is the minimum

value of Ra_{ck} which determines the absolute threshold for stability. The minimum is obtained when

$$k = \frac{\pi}{\sqrt{2}}, \tag{3.22}$$

and the resulting critical value of the Rayleigh number is

$$Ra_c = \frac{27\pi^4}{4} \approx 657.5; \tag{3.23}$$

That is, the steady state is linearly unstable if $Ra > Ra_c$.

For other boundary conditions, the solutions are still exponentials, but the coefficients, and hence also the growth rate, must be found numerically. The resultant critical value of the Rayleigh number is higher for no slip boundary conditions, for example, (it is about 1707), and in general, thermal convection is initiated at values of $Ra \gtrsim O(10^3)$.

3.5 Double-diffusive convection

Double-diffusive convection refers to the motion which is generated by buoyancy, when the density depends on two diffusible substances or quantities. The simplest examples occur when salt solutions are heated; then the two diffusing quantities are heat and salt. Double-diffusive processes occur in sea water and in lakes, for example. Other simple examples occur in multi-component fluids containing more than one dissolved species; convection in magma chambers is one such.

The guiding principle behind double-diffusive convection is still that light fluid rises, and convection occurs in the normal way (the direct mode) when the steady state is statically unstable (i. e., when the density increases with height), but confounding factors arise when, as normally the case, the two substances diffuse at different rates. Particularly when we are concerned with temperature and salt, the ratio of thermal to solutal diffusivity is large, and in this case different modes of convection occur near the statically neutral buoyancy state: the cells can take the form of long thin 'fingers', or the onset of convection can be oscillatory. In practice, fingers are seen, but oscillations are not.

A further variant on Rayleigh-Bénard convection arises in the form of convective layering. This is a long-lived transient form of convection, in which separately convecting layers form, and is associated partly with the high diffusivity ratio, and partly with the usual occurrence of no flux boundary conditions for diffusing chemical species.

We pose a model for double-diffusive convection based on a density which is related linearly to temperature T and salt composition c in the form

$$\rho = \rho_0[1 - \alpha(T - T_0) + \beta c], \tag{3.24}$$

where we take α and β to be positive constants; thus the presence of salt makes the fluid heavier. The equation that then needs to be added to (3.1) is that for convective

diffusion of salt:

$$c_t + \mathbf{u}.\boldsymbol{\nabla} c = D\nabla^2 c, \tag{3.25}$$

where D is the solutal diffusion coefficient, assuming a dilute solution. We adopt the same scaling of the variables as before, with the extra choice

$$c \sim c_0, \tag{3.26}$$

where c_0 is a relevant salinity scale (in our stability analysis, it will be the prescribed salinity at the base of the fluid layer). The Boussinesq form of the scaled equations, based on the assumptions that $\alpha\Delta T \ll 1$ and $\beta c_0 \ll 1$, are then

$$\begin{aligned} \boldsymbol{\nabla}.\mathbf{u} &= 0, \\ \frac{1}{Pr}[\mathbf{u}_t + (\mathbf{u}.\boldsymbol{\nabla})\mathbf{u}] &= -\boldsymbol{\nabla} p + \nabla^2\mathbf{u} + Ra\, T\hat{\mathbf{k}} - Rs\, c\hat{\mathbf{k}}, \\ T_t + \mathbf{u}.\boldsymbol{\nabla} T &= \nabla^2 T, \\ c_t + \mathbf{u}.\boldsymbol{\nabla} c &= \frac{1}{Le}\nabla^2 c. \end{aligned} \tag{3.27}$$

The Rayleigh number Ra and the Prandtl number Pr are defined as before, and the solutal Rayleigh number Rs and the Lewis number Le are defined by

$$Rs = \frac{\beta\rho_0 c_0 g d^3}{\mu\kappa}, \quad Le = \frac{\kappa}{D}. \tag{3.28}$$

Note that in the absence of temperature gradients, the quantity $-Rs\, Le$ would be the effective Rayleigh number determining convection.

3.5.1 Linear stability

Now we study the linear stability of a steady state maintained by prescribed temperature and salinity differences ΔT and c_0 across a stress-free fluid layer. In dimensionless terms, we pose the boundary conditions

$$\begin{aligned} \psi = \nabla^2\psi = 0 \quad &\text{at} \quad z = 0, 1, \\ T = c = 0 \quad &\text{at} \quad z = 1, \\ T = c = 1 \quad &\text{at} \quad z = 0, \end{aligned} \tag{3.29}$$

where as before, we restrict attention to two dimensions, and adopt a stream function ψ. The steady state is

$$c = 1 - z, \quad T = 1 - z, \quad \psi = 0, \tag{3.30}$$

and we perturb it by writing

$$c = 1 - z + C, \quad T = 1 - z + \theta, \tag{3.31}$$

and then linearising the equations on the basis that $C, \theta, \psi \ll 1$. This leads to

$$\begin{aligned} \frac{1}{Pr}\nabla^2\psi_t &\approx Ra\,\theta_x - Rs\,C_x + \nabla^4\psi, \\ \theta_t - \psi_x &\approx \nabla^2\theta, \\ C_t - \psi_x &\approx \frac{1}{Le}\nabla^2 C, \end{aligned} \tag{3.32}$$

with

$$C = \psi = \psi_{zz} = \theta = 0 \quad \text{on} \quad z = 0, 1. \tag{3.33}$$

By inspection, solutions satisfying the temperature and salinity equations are

$$\begin{aligned} \psi &= \exp(ikx + \sigma t)\sin m\pi z, \\ \theta &= \frac{ik}{\sigma + K^2}\exp(ikx + \sigma t)\sin m\pi z, \\ C &= \frac{ik}{\sigma + \dfrac{K^2}{Le}}\exp(ikx + \sigma t)\sin m\pi z, \end{aligned} \tag{3.34}$$

where we have written

$$K^2 = k^2 + m^2\pi^2. \tag{3.35}$$

Substituting these into the momentum equation leads to the dispersion relation determining σ in terms of k:

$$(\sigma + K^2 Pr)(\sigma + K^2)\left(\sigma + \frac{K^2}{Le}\right) + k^2 Pr\left[\frac{(Rs - Ra)\sigma}{K^2} + Rs - \frac{Ra}{Le}\right] = 0. \tag{3.36}$$

This is a cubic in σ, which can be written in the form

$$\sigma^3 + a\sigma^2 + b\sigma + c = 0, \tag{3.37}$$

where

$$\begin{aligned} a &= K^2\left(Pr + 1 + \frac{1}{Le}\right), \\ b &= K^4\left(Pr + \frac{1}{Le} + \frac{Pr}{Le}\right) + \frac{k^2}{K^2}Pr(Rs - Ra), \\ c &= \frac{K^6}{Le}Pr + k^2 Pr\left(Rs - \frac{Ra}{Le}\right). \end{aligned} \tag{3.38}$$

Instability occurs if any one of the three roots of (3.37) has positive real part. Since Le and Pr are properties of the fluid, we take them as fixed, and study the effect of varying Ra and Rs on the stability boundaries where $\mathrm{Re}\,\sigma = 0$. Firstly, if $Ra < 0$ and $Rs > 0$, then a, b and c are all positive. We can then show (see question 3.3) that $\mathrm{Re}\,\sigma < 0$ for all three roots providing $ab > c$, and this is certainly the case if

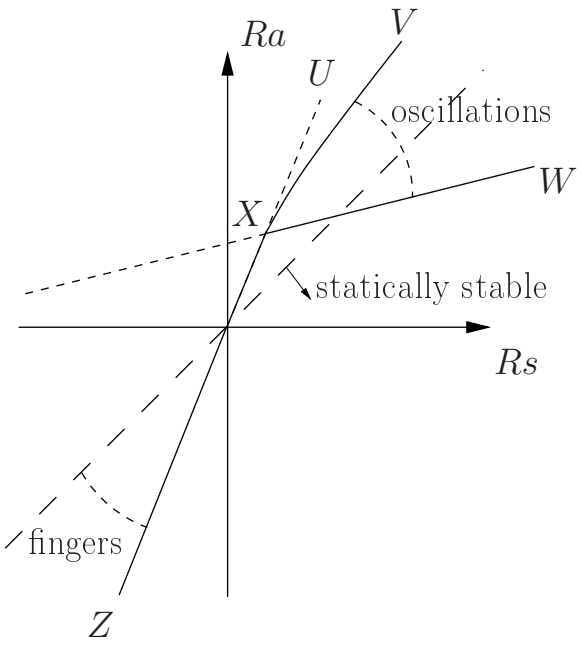

Figure 3.7: Stability diagram for double-diffusive convection.

$Le > 1$, which is always true for heat and salt diffusion. Thus when both temperature and salinity fields are stabilising, the state of no motion is linearly stable.

To find regions of instability in the (Rs, Ra) plane, it thus suffices to locate the curves where $\mathrm{Re}\,\sigma = 0$. There are two possibilities. The first is referred to as exchange of stability, or the direct mode, and occurs when $\sigma = 0$. From (3.37), this is when $c = 0$, or $Rs = \dfrac{Ra}{Le} - \dfrac{K^6}{k^2 Le}$. This is a single curve (for each k), and since we know that $\mathrm{Re}\,\sigma < 0$ in $Ra < 0$ and $Rs > 0$, this immediately tells us that a direct instability occurs if

$$Ra - Le\,Rs > R_c = \min_k \frac{K^6}{k^2} = \frac{27\pi^4}{4}. \tag{3.39}$$

This direct transition is the counterpart of the onset of Rayleigh-Bénard convection, and shows that $Ra - Le\,Rs$ is the effective Rayleigh number. This is consistent with the remark just after (3.28).

The other possibility is that there is a Hopf bifurcation, i. e., a pair of complex conjugate values of σ cross the imaginary axis at $\pm i\Omega$, say. The condition for this is $ab = c$, which is again a single curve, and one can show (see question 3.4) that oscillatory instability occurs for

$$Ra > \frac{\left(Pr + \dfrac{1}{Le}\right) Rs}{1 + Pr} + \frac{\left(1 + \dfrac{1}{Le}\right)\left(Pr + \dfrac{1}{Le}\right)}{Pr} R_c. \tag{3.40}$$

Direct instability occurs along the line XZ in figure 3.7, while oscillatory instability occurs at the line XW. Between XW and the continuation XU of XZ, there

are two roots with positive real part and one with negative real part. As Ra increases above XW, it is possible that the two complex roots coalesce on the real axis, so that the oscillatory instability is converted to a direct mode. One can show (see question 3.5) that the criterion for this is that $b < 0$ and

$$c = \tfrac{1}{9}\left[ab + \frac{(a^2 - 6b)}{3}\left\{-a + (a^2 - 3b)^{1/2}\right\}\right]. \tag{3.41}$$

For large Rs, this becomes (for $k^2 = \dfrac{\pi^2}{2}$)

$$Ra \approx Rs + \frac{3R_c^{1/3} Rs^{2/3}}{2^{2/3} Pr^{1/3}}, \tag{3.42}$$

and is shown as the line XW in figure 3.7. Thus the onset of convection is oscillatory only between the lines XW and XV, and beyond (above) XV it is direct. In practice, oscillations are rarely seen.

Fingers

If we return to the cubic in the form (3.36), and consider the behaviour of the roots in the third quadrant as $Ra, Rs \to -\infty$, it is easy to see that one root is

$$\sigma \approx \frac{K^2\left[\dfrac{Ra}{Le} - Rs\right]}{Rs - Ra}, \tag{3.43}$$

while the other two are oscillatorily stable (see question 3.6). Thus this growth rate is positive when $Le\, Rs < Ra < Rs$ and grows unboundedly with the wave number k (since $K^2 = k^2 + \pi^2$ when $m = 1$). This is an indication of ill-posedness, and in fact we anticipate that σ will become negative at large k. To see when this occurs, inspection of (3.36) shows that the neglected terms in the approximation (3.43) become important when $k \sim |Ra|^{1/4}$, where σ is maximum (of $O|Ra|^{1/4}$), and then $\sigma \sim -k^2$ for larger k. Thus in the 'finger' régime sector indicated in figure 3.7, the most rapidly growing wavelengths are short, and the resulting waveforms are tall and thin. This is what is seen in practice, and the narrow cells are known as fingers.

3.5.2 Layered convection

The linear stability analysis we have given above is only partially relevant to double diffusive convection. It is helpful in the understanding of the finger régime, but the oscillatory mode of convection is not particularly relevant. The other principal phenomenon which double diffusive systems exhibit is that of layering. This is a transient, but long-term, phenomenon associated often with the heating of a stable salinity gradient, and arises because in normal circumstances, more appropriate

boundary conditions for salt concentration are to suppose that there is no flux at the bounding surfaces.

In pure thermal convection, the heating of an initially stably thermally stratified fluid will lead to the formation of a layer of convecting fluid below the stable region. This (single) convecting layer will grow in thickness until it fills the entire layer. This is essentially the 'filling box' which is described in section 3.10. Suppose now we have a stable salinity gradient which is heated from below. Again a convecting layer forms, which mixes the temperature and concentration fields to be uniform within the layer. At the top of the convecting layer, there will be a step down ΔT in temperature, and a step down Δc in salinity. It is found experimentally that $\alpha\Delta T = \beta\Delta c$, that is, the *boundary layer*[2] is neutrally stable. However, the disparity in diffusivities (typically $Le \gg 1$) means that there is a thicker thermal conductive layer ahead of the interface. In effect, the stable salinity gradient above the convecting layer is heated by the layer itself, and a second, and then a third, layer forms. In this way, the entire fluid depth can fill up with a sequence of long-lived, separately convecting layers. The layers will eventually merge and form a single convecting layer over a time scale controlled by the very slow transport of salinity between the convecting layers. Such layers are very suggestive of some of the fossilised layering seen in magma chambers, as for example in figure 3.5, but the association may be a dangerous one.

3.6 High Rayleigh number convection

We have seen that convection occurs if the Rayleigh number is larger than $O(10^3)$ in general, depending on the precise boundary conditions which apply. In the Earth's mantle, suitable values of the constituent parameters are $\alpha = 3 \times 10^{-5}$ K^{-1}, $\Delta T =$ 3000 K, $\rho_0 = 3 \times 10^3$ kg m^{-3}, $g = 10$ m s^{-2}, $d = 3000$ km, $\eta_0 = 10^{21}$ Pa s, $\kappa_0 = 10^{-6}$ m^2 s^{-1}, and for these values, the Rayleigh number is slightly less than 10^8. Thus the Rayleigh number is much larger than the critical value, and as a consequence we can expect the convection to be vigorous (if velocities of centimetres per year can be said to be vigorous).

There are various intuitive ways in which we can get a sense of the likely behaviour of the convective solutions of the Boussinesq equations when $Ra \gg 1$. Since Ra multiplies the buoyancy term, any $O(1)$ lateral temperature gradient will cause enormous velocities. One might thus expect the flow to organise itself so that either horizontal temperature gradients are small, or they are confined to thin regions, or both. Since $O(1)$ temperature variations are enforced by the boundary conditions, the latter is more plausible, and thus we have the idea of the *thermal plume*, a localised upwelling of hot fluid which will be instantly familiar to glider pilots and seabirds.

A mathematically intuitive way of inferring the same behaviour follows from the expectation that increasing Ra drives increasing velocities; then large Ra should imply large velocity, and the conduction term in the heat equation $\mathbf{u}.\boldsymbol{\nabla}T = \nabla^2 T$ is correspondingly small. Since the conduction term represents the highest derivative

[2]For discussion of boundary layers, see section 3.6.1.

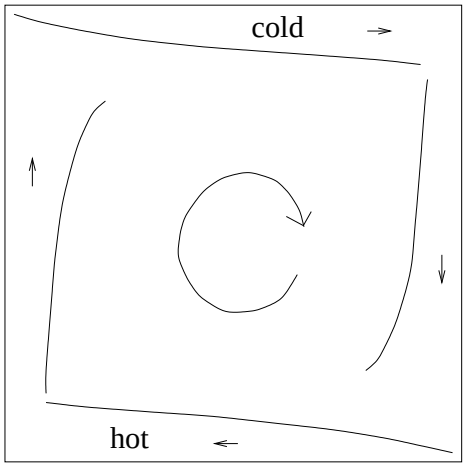

Figure 3.8: Schematic representation of boundary layer convection

in the equation, its neglect would imply a reduction of order, and correspondingly we would expect *thermal boundary layers* to exist at the boundaries of the convecting cell. This is in fact what we will find: a hot thermal boundary layer adjoins the lower boundary, and a cold one adjoins the upper boundary, and a rapid circulation in the interior of the cell detaches these as upwelling and downwelling plumes. The general structure of the resulting flow is shown in figure 3.8. We analyse this structure in the following sections.

3.6.1 Boundary layer theory

The Boussinesq equations describing thermal convection are written in the following dimensionless form:

$$\begin{aligned} \boldsymbol{\nabla}.\mathbf{u} &= 0, \\ \frac{1}{Pr}\frac{d\mathbf{u}}{dt} &= -\boldsymbol{\nabla}p + \nabla^2\mathbf{u} + Ra\,T\mathbf{j}, \\ \frac{dT}{dt} &= \nabla^2 T, \end{aligned} \tag{3.44}$$

where $\mathbf{u}$ is velocity, p is pressure, T is temperature, and the Rayleigh and Prandtl numbers are defined in (3.4).

By considering only two-dimensional motion in the (x, z) plane, we define the stream function ψ by

$$u = -\psi_z, \quad v = \psi_x; \tag{3.45}$$

the vorticity is then $(0, \omega, 0)$, where $\omega = -\nabla^2\psi$. Taking the curl of the momentum equation, we derive the set

$$\omega = -\nabla^2\psi,$$
$$\frac{dT}{dt} = T_t + \psi_x T_y - \psi_y T_x = \nabla^2 T,$$
$$\frac{1}{Pr}\frac{d\omega}{dt} = -Ra\, T_x + \nabla^2\omega, \tag{3.46}$$

which are supplemented by the boundary conditions

$$\psi, \omega = 0 \quad \text{on} \quad x = 0,\ a,\ z = 0, 1,$$
$$T = \tfrac{1}{2} \quad \text{on} \quad z = 0, \quad T = -\tfrac{1}{2} \quad \text{on} \quad z = 1,$$
$$T_x = 0 \quad \text{on} \quad x = 0,\ a; \tag{3.47}$$

here a is the aspect ratio, and we have chosen free slip (no stress) conditions at the cell boundaries.

Rescaling

The idea is that when $Ra \gg 1$, thermal boundary layers of thickness $\delta \ll 1$ will form at the edges of the flow, and both ψ and ω will be $\gg 1$ in the flow. To scale the equations properly, we rescale the variables as

$$\psi,\ \omega \sim 1/\delta^2, \tag{3.48}$$

and define

$$\delta = Ra^{-1/3}. \tag{3.49}$$

Rescaled, the equations are thus, in the steady state,

$$\omega = -\nabla^2\psi,$$
$$\psi_x T_z - \psi_z T_x = \delta^2 \nabla^2 T,$$
$$\nabla^2\omega = \frac{1}{\delta} T_x + \frac{1}{Pr\,\delta^2}\frac{d\omega}{dt}. \tag{3.50}$$

In order that the inertia terms be unimportant, we require $Pr\,\delta^2 \gg 1$, i.e. $Pr \gg Ra^{2/3}$. This assumption is easily vindicated in the earth's mantle, but is difficult to achieve in the laboratory.

As in any singular perturbation procedure, we now examine the flow region by region, introducing special rescalings in regions where boundary conditions cannot be satisfied.

Core flow

The temperature equation is linear in T, and implies $T = T_0(\psi) + O(\delta^2)$. For a flow with closed streamlines, the Prandtl-Batchelor theorem then implies $T_0 =$ constant (this follows from the exact integral $\oint_C \frac{\partial T}{\partial n} ds = 0$, where the integral is around a streamline, whence $T_0'(\psi) \oint_C \frac{\partial \psi}{\partial n} ds = 0$); it then follows that T is constant to all (algebraic) orders of δ, and is in fact zero by the symmetry of the flow. Thus

$$\begin{aligned} T &= 0, \\ \nabla^4 \psi &= 0, \end{aligned} \tag{3.51}$$

and clearly the core flow cannot have $\psi = \omega = 0$ at the boundaries, for non-zero ψ. In fact, ω jumps at the side-walls where the plume buoyancy generates a non-zero vorticity. We examine the plumes next.

Plumes

Near $x = 0$, for example, we rescale the variables as

$$x \sim \delta, \quad \psi \sim \delta, \tag{3.52}$$

and denote rescaled variables by capital letters. At leading order, we then have

$$\Psi_{XX} \sim 0, \tag{3.53}$$

whence $\Psi \sim v_p(y) X$, and to match to the core flow, we define $v_p = \psi_x \,|_{x=0}$ as the core velocity at $x = 0$. Also

$$\begin{aligned} \Psi_X T_y - \Psi_y T_X &\sim T_{XX}, \\ \omega_{XX} &\sim T_X, \end{aligned} \tag{3.54}$$

the latter of which integrates to give

$$\omega = \int_0^X T \, dX, \quad \omega_p = \int_0^\infty T \, dX, \tag{3.55}$$

where matching requires ω_p to be the core vorticity at $x = 0$. Integration of $(3.54)_1$ gives

$$\int_0^\infty T \, d\Psi = C, \tag{3.56}$$

where C is constant, and it follows that the core flow must satisfy the boundary condition $\omega \psi_x = C$ on $x = 0$. In summary, the effective boundary conditions for the core flow are

$$\psi = 0 \quad \text{on} \quad x = 0, a;\ z = 0, 1;$$
$$\psi_{zz} = 0 \quad \text{on} \quad z = 0, 1; \tag{3.57}$$
$$\psi_x \psi_{xx} = -C \quad \text{on} \quad x = 0, \quad \psi_x \psi_{xx} = C \quad \text{on} \quad x = a,$$

and the solution can be found as $\psi = C^{1/2}\hat{\psi}$, where $\hat{\psi}$ is determined numerically. It thus remains to determine C. This requires consideration of the thermal boundary layers.

Thermal boundary layers

Near the base, for example, we rescale the variables

$$z \sim \delta, \quad \psi \sim \delta, \quad \omega \sim \delta, \tag{3.58}$$

to find the leading order rescaled equations as

$$\Psi_{ZZ} \sim 0, \tag{3.59}$$

whence $\Psi \sim u_b(x)Z$, and $-u_b$ is the core value of the basal velocity. Then $\Omega_{ZZ} \sim T_x$ determines Ω (with $\Omega = 0$ on $Z = 0$, and $\Omega \sim \omega_b(x)Z$ as $Z \to \infty$, where ω_b is the core value of the basal vorticity), and T satisfies

$$\Psi_x T_Z - \Psi_Z T_X \sim T_{ZZ}. \tag{3.60}$$

In Von Mises coordinates x, Ψ, the equation is

$$-T_x \sim \frac{\partial}{\partial \Psi}\left[\Psi_Z \frac{\partial T}{\partial \Psi}\right], \tag{3.61}$$

and putting $\xi = \displaystyle\int_x^a u_b(x)dx$ (so ξ marches from right to left in the direction of flow), this is just the diffusion equation

$$T_\xi = T_{\Psi\Psi}, \tag{3.62}$$

with

$$T = \tfrac{1}{2} \quad \text{on} \quad \Psi = 0, \quad T \to 0 \quad \text{as} \quad \Psi \to \infty. \tag{3.63}$$

A quantity of interest is the Nusselt number, defined as

$$Nu = -\int_0^1 \frac{\partial T}{\partial z}(x, 0)dx, \tag{3.64}$$

and from the above, this can be written as

$$Nu \sim \left[\int_0^\infty T|_{z=0}\, d\Psi\right]_{x=a}^{x=0} R^{1/3}. \tag{3.65}$$

Notice that the plume temperature equation can also be written as (3.62), where ξ is extended as $\int^z v_p(z)\,dz$, etc.

Corner flow

The core flow has a singularity in each corner, where (if r is distance from the corner), then $\psi \sim r^{3/2}$, $\omega \sim r^{-1/2}$, and (for the corner at $x = 0$, $z = 0$, for example) $x, z \sim r$. There must be a region where this singularity is alleviated by the incorporation of the buoyancy term. This requires $\omega/r^2 \sim 1/\delta r$, whence $r \sim \delta^{2/3}$. Rescaling the variables as indicated ($x, z \sim \delta^{2/3}$, $\psi \sim \delta$, $\omega \sim \delta^{-1/3}$) then gives the temperature equation as

$$\Psi_X T_Z - \Psi_Z T_X \sim \delta \nabla^2 T, \tag{3.66}$$

which shows that (since the ψ scale, δ, is the same as that of the boundary layers adjoining the corner) the boundary layer temperature field is carried through the corner region. The corner flow has $T \sim T(\Psi)$, so that

$$\nabla^4 \Psi + T'(\Psi)\Psi_X = 0, \tag{3.67}$$

with appropriate matching conditions.

Solution strategy

The temperature equation (3.62) must now be solved in the four regions corresponding to the boundary layer at $z = 0$, plume at $x = 0$, boundary layer at $z = 1$, plume at $x = a$, with T being continuous at each corner, and

$$\begin{aligned}
&T \to 0 \quad \text{as} \quad \Psi \to \infty, \\
&T = \tfrac{1}{2} \quad \text{on} \quad \Psi = 0 \qquad [z = 0,\ \text{base}], \\
&\frac{\partial T}{\partial \Psi} = 0 \quad \text{on} \quad \Psi = 0 \quad [x = 0,\ \text{left}], \\
&T = -\tfrac{1}{2} \quad \text{on} \quad \Psi = 0 \quad [z = 1,\ \text{top}], \\
&\frac{\partial T}{\partial \Psi} = 0 \quad \text{on} \quad \Psi = 0 \quad [x = a,\ \text{right}];
\end{aligned} \tag{3.68}$$

in addition, T is periodic in ξ. Beginning from $x = a$, $z = 0$, denote the values of ξ at the corners as ξ_A ($x = 0$, $z = 0$), ξ_B ($x = 0$, $z = 1$), ξ_C ($x = a$, $z = 1$). From the definition of ξ, we have $\xi_k = C^{1/2}\hat{\xi}_k$, where $\hat{\xi}_k$ are independent of C. Putting

$$\xi = C^{1/2}\hat{\xi}, \quad \Psi = C^{1/4}\hat{\Psi}, \tag{3.69}$$

then the problem for $T(\hat{\xi}, \hat{\Psi})$ is independent of C. If we can solve this numerically, then $\int T\,d\Psi = C^{1/4} \int T\,d\hat{\Psi}$, thus

$$Nu \sim C^{1/4} \left[\int_0^\infty T \, d\hat{\Psi} \right]_0^{\hat{\xi}_A} R^{1/3}, \tag{3.70}$$

and lastly, C is determined from

$$C = \left[\int_0^\infty T \, d\hat{\Psi} \right]^{4/3}, \tag{3.71}$$

where the integral is evaluated at $\hat{\xi}_A$. Since also $-C^{3/4} = \int_0^\infty T \, d\hat{\Psi}$ at $\xi = 0$, (3.70) can be written as

$$Nu \sim 2CRa^{1/3}. \tag{3.72}$$

No-slip boundary conditions

For no slip boundary conditions, the necessary preliminary rescaling is $\psi \sim 1/\delta^3$, $\omega \sim 1/\delta^3$, where $\delta = Ra^{-1/5}$. Thus the Nusselt number $Nu \sim Ra^{1/5}$. There is no longer parity between the thermal boundary layers and plumes, as the former are slowed down by the no slip conditions. The rescaled equations are

$$\begin{aligned} \omega &= -\nabla^2 \psi, \\ \psi_x T_z - \psi_z T_x &= \delta^3 \nabla^2 T, \\ \nabla^2 \omega &= \frac{1}{\delta^2} T_x. \end{aligned} \tag{3.73}$$

The core flow is as before; the thermal boundary layers have $\psi \sim \delta^2$, $\omega \sim 1$, $z \sim \delta$, so that vorticity balances buoyancy (an omission in Roberts' 1979 paper, precluding his similarity solution), and all three equations are necessary to solve for T; it is still the case that $\int T \, d\psi$ is conserved at corners, but now in the plume, $x \sim \delta^{3/2}$, $\psi \sim \delta^{3/2}$, and $T \sim \delta^{1/2}$. The initial plume profile is effectively a delta function, and the plume temperature is just the resultant similarity solution. The remainder of the structure must be computed numerically, something which has not been done.

3.7 Parameterised convection

The boundary layer theory described in section **??** applies to steady state solutions at high Rayleigh number, but in fact real convection becomes time-varying at such parameter values. The behaviour becomes first oscillatory, and then becomes increasingly irregular, so that at very high Rayleigh numbers, the cellular structure of convection in a fluid layer breaks down. The upwelling and downwelling plumes of the boundary layer theory still exist, but their detachment is sporadic and irregular. In these circumstances, the theoretical description of convection become, paradoxically, easier. Just as for turbulent shear flows at high Reynolds numbers, one uses

Figure 3.9: On the left, a sub-oceanic black smoker issuing from a vent at the ocean floor; on the right, an industrial chimney. Images from `http://oceanexplorer.noaa.gov`.

empirically-based measures of the fluxes at boundaries to describe the flow. Turbulence mixes the fluid, so that, as in the boundary layer theory, the interior of a convecting cell is taken to be isothermal. In this section, we describe two particular examples of turbulent convection to illustrate these ideas. The first is that of the turbulent convective plume, and the second is a classic example of the use of turbulent convective theory in the decription of the 'filling box'.

3.8 Plumes

A plume is an isolated convective upwelling. Examples are the rise of smoke from an industrial chimney, the formation of cumulus clouds over oceans, 'black smokers' at mid-ocean rise vents, and explosive volcanic eruptions. In these examples, a source of buoyancy at (essentially) a point drives a convective flow in the fluid above. As suggested in figure 3.9, the plume forms a turbulent, approximately conical region, with a fairly sharp (but time-varying) boundary. The turbulence causes rapid convective mixing, and allows us to conceptualise the plume as a relatively homogeneous cloud of density $\rho = \rho_0 - \Delta\rho$ rising through an ambient medium of density ρ_0. If ρ_0 depends on height z, then the medium is called a stratified medium, and it is stably stratified if $\rho_0'(z) < 0$.

Suppose that a cylindrical plume of radius $r = b(z)$ rises at speed w through a medium of density ρ_0. We take the density deficit in the plume to be $\Delta\rho$, and define the reduced gravity to be

$$g' = \frac{g\Delta\rho}{\rho_0}. \tag{3.74}$$

The (constant) buoyancy flux is then defined to be

$$B = 2\pi \int_0^b rwg' \, dr. \tag{3.75}$$

The boundary of the plume is taken to be such that $g' = w = 0$ at $r = b$.

Self-similarity

If we take ρ_0 to be constant, and suppose that the dynamics of the plume depend only on the reduced gravity and on position, or equivalently on the buoyancy flux and position, then purely dimensional reasoning leads to the conclusion that we must have

$$\begin{aligned} w &= c_1 \left(\frac{r}{b}\right) B^{1/3} z^{-1/3}, \\ g' &= c_2 \left(\frac{r}{b}\right) B^{2/3} z^{-5/3}, \\ b &= \beta z. \end{aligned} \tag{3.76}$$

The functions c_1 and c_2 are found experimentally to be well fitted by Gaussians. Note that the volume flux

$$Q = 2\pi \int_0^b rw \, dr \tag{3.77}$$

is then given by

$$Q \sim B^{1/3} z^{5/3}, \tag{3.78}$$

so that the plume volume flux increases with height. As a consequence, for this solution to be appropriate, the plume must entrain ambient fluid.

Entrainment

In fact this is found to be the case. The turbulent eddies of the plume incorporate the ambient fluid, and dramatically increase the plume volume flux. If the entrainment velocity at the edge of the plume is u_e, then we have that

$$(ru)|_b = -bu_e. \tag{3.79}$$

The entrainment velocity needs to be constituted, and a common assumption is to suppose that

$$u_e = \alpha w, \tag{3.80}$$

where the value of α is found experimentally to be approximately 0.1. It is no longer quite obvious that with this assumption, the similarity solution (3.76) will still be appropriate, but in fact it is (but depends on the fact that (3.80) is linear).

Plumes in a stratified environment

If, as for example in the atmosphere, the ambient density decreases with height, then a similarity solution for Q is no longer feasible. To derive a model for such a plume, we consider a turbulent, cylindrical plume to be described by the inviscid, conductionless (or diffusionless) Boussinesq equations:

$$\begin{aligned} uu_r + wu_z &= -\frac{1}{\rho_0}p_r, \\ uw_r + ww_z &= -\frac{1}{\rho_0}p_z - \frac{\rho}{\rho_0}g, \\ u\rho_r + w\rho_z &= 0, \\ \frac{1}{r}(ru)_r + w_z &= 0, \end{aligned} \tag{3.81}$$

where (r, z) are cylindrical coordinates, (u, w) the corresponding velocity components, p the pressure, ρ the density, ρ_0 the reference density, and g is the acceleration due to gravity. The Boussinesq approximation is based on the assumption that $\Delta\rho \ll \rho_0$.

Dissipative processes are ignored in writing (3.81) because in a turbulent flow, the molecular diffusivites are unimportant. There are turbulent diffusivities arising from velocity fluctuations (for example Reynolds stresses), but these are also relatively small, and are feasibly ignored in a first approximation. The density evolution equation $(3.81)_3$ merits some comment. It indicates the assumption that density depends on a variable such as temperature or composition (or both) which is described by an advection-diffusion equation, so that when the diffusion term is ignored, we obtain the purely advective equation indicated, assuming that density depends linearly on the advected variables (as is typically the case in Boussinesq fluids).

By integrating the equations across the plume, we can derive three equations for the buoyancy flux B, the volume flux Q, and the momentum flux

$$M = 2\pi \int_0^b rw^2\, dr, \tag{3.82}$$

and these are (see question 3.8)

$$\begin{aligned} \frac{dB}{dz} &= -N^2 Q, \\ \frac{dQ}{dz} &= 2\pi\alpha b w, \\ \frac{dM}{dz} &= 2\pi \int_0^b rg'\, dr, \end{aligned} \tag{3.83}$$

where N is the Brunt–Väisälä frequency,

$$N = \left(-\frac{g\rho_0'(z)}{\rho_0}\right)^{1/2}. \tag{3.84}$$

Note that when $N = 0$, i. e., the ambient fluid is not stratified, then the buoyancy flux is constant, as we assumed.

To progress, we must make closure assumptions about the average fluxes B, Q and M in terms of the plume (average) velocity w and radius b. The simplest assumptions are based on the notion that the profiles of density and velocity are uniform across the plume (as opposed to the more realistic Gaussians), and this leads to the formulae

$$B = \pi b^2 w g', \quad Q = \pi b^2 w, \quad M = \pi b^2 w^2; \tag{3.85}$$

in addition we take

$$2\pi \int_0^b r g' \, dr = \pi b^2 g'. \tag{3.86}$$

Eliminating w and b finally yields the equations (see question 3.9)

$$\begin{aligned} \frac{dB}{dz} &= -N^2 Q, \\ \frac{dM}{dz} &= \frac{BQ}{M}, \\ \frac{dQ}{dz} &= 2\pi^{1/2} \alpha M^{1/2}. \end{aligned} \tag{3.87}$$

We can see from this that the buoyancy flux continually decreases with height, while the volume flux increases. When $B = 0$, the plume reaches its level of neutral buoyancy, but continues to rise because of its momentum. With $B < 0$, M decreases, and will not rise any further when M reaches 0. According to the equations, the volume flux is still positive, but in fact the plume spreads out laterally, and the one-dimensional description becomes irrelevant. Thus a plume in a stratified medium will level out at a height z_s which can be determined from (3.87) in the form (see question 3.9)

$$z_s = c B_0^{1/4} N^{-3/4}, \tag{3.88}$$

where B_0 is the buoyancy flux at $z = 0$, and N is assumed constant.

3.9 Turbulent convection

As the Rayleigh number increases in Rayleigh–Bénard convection, the convective rolls which can be seen at the onset of convection bifurcate to three-dimensional planforms, typically either square cells or hexagons. In a layer of large horizontal extent, convective rolls tend to be weakly chaotic, because the alignment in different parts of the layer is different, and thus defects or dislocations are formed in the cellular structure, and these migrate slowly, sometimes permanently. Three-dimensional cells tend to be more stable, because they are essentially confined, but at higher Rayleigh number, an oscillatory instability sets in. The thermal boundary layers which migrate across the base of the cells and detach at the cell boundaries start to prematurely thicken and then thin again before detachment, causing an oscillation which is a

manifestation of budding plume development. Eventually, these budding plumes do begin to detach before reaching the cell walls, and at this point the convection becomes temporally and spatially disordered. Thermal boundary layers thicken and plumes detach irregularly, and a defined cellular structure disappears, being replaced by a host of upwelling and downwelling thermal plumes. In fact, a large scale circulation does come into existence, but this is on a much larger scale than the typical plume spacing.

A very famous but simple model of turbulent thermal convection was put forward by Lou Howard in 1964, at the International Congress of Mechanics in Munich. In his model, a quiescent thermal boundary layer grows into an isothermal core until it reaches a critical thickness, when it suddenly forms a plume and detaches, mixing the fluid and returning to isothermal conditions. The average heat flux is then determined by that during the quiescent, conductive phase. The conductive temperature in the growing boundary layer is given by the solution of

$$T_t = \kappa T_{zz}, \tag{3.89}$$

with

$$\begin{aligned} T &= \frac{\Delta}{2} \quad \text{on} \quad z = 0, \\ T \to 0 \quad &\text{as} \quad z \to \infty; \end{aligned} \tag{3.90}$$

here we imagine a convecting fluid layer of depth d, across which the prescribed temperature difference is ΔT (and thus half across the boundary layers on each surface). Starting from an isothermal state $T = 0$ (boundary layer of thickness zero), the solution is

$$T = \frac{\Delta}{2}\text{erfc}\left(\frac{z}{2\sqrt{\kappa t}}\right), \tag{3.91}$$

and thus the average heat flux from the surface $z = 0$ is

$$F = \frac{1}{t_c}\int_0^{t_c}\left(-k\frac{\partial T}{\partial z}\right)\bigg|_{z=0} dt, \tag{3.92}$$

where t_c is the time of detachment of the boundary layer. Using (3.91), we then find

$$F = \frac{k\Delta T}{2\sqrt{\kappa t_c}} = \frac{k\Delta T}{2d_c}, \tag{3.93}$$

where $d_c = \sqrt{\kappa t_c}$ is the thickness of the thermal boundary layer at detachment.

Howard hypothesised that detachment would occur when a locally defined Rayleigh number, using the boundary layer thickness as the depth scale, became critical, of order

$$Ra_c \sim 10^3; \tag{3.94}$$

thus we define the critical thickness d_c via the effective critical Rayleigh number condition

$$\frac{\alpha\rho_0 g d_c^3 \Delta T}{2\mu\kappa} = Ra_c, \tag{3.95}$$

where the factor 2 allows for the temperature drop of $\Delta T/2$ across the boundary layer. In terms of the Rayleigh number of the fluid layer

$$Ra = \frac{\alpha \rho_0 g d^3 \Delta T}{\mu \kappa}, \tag{3.96}$$

we thus have the dimensionless heat flux, called the Nusselt number Nu, given by

$$Nu = \frac{F}{(k\Delta T/d)} = \frac{d}{d_c} = cRa^{1/3}, \tag{3.97}$$

where

$$c = (2Ra)^{-1/3} \approx 0.08. \tag{3.98}$$

Thus the heat flux can be parameterised as

$$F = c \left(\frac{\alpha g c_p}{\mu} \right)^{1/3} (\rho_0 k)^{2/3} \Delta T^{4/3}, \tag{3.99}$$

which is the famous four-thirds law for turbulent convection. it is reasonably consistent with experimental results.

3.10 The filling box

3.11 Double-diffusive layering

3.12 Notes and references

The theory of continental drift was famously published by Alfred Wegener, a German meteorologist, in 1915. An English translation of his book was published later, see Wegener (1924). His ideas were scorned by the geophysical establishment, and in particular, in Britain, by the colossal figure of Harold Jeffreys. The blind ignorance with which he and other fellow geologists refuted Wegener's ideas should serve (but have not) as a lesson for scientists against the perils of treating science as religion, and hypothesis as dogma. A notable supporter of the thesis of continental drift was Holmes (1978), who understood that mantle convection was the driving mechanism. A more modern treatment of geodynamics is the classic book by Turcotte and Schubert (1982), while Davies (1999) gives a readable but technically undemanding account.

The layered magma chamber known as the Skaergaard intrusion was the subject of a massive memoir by Wager and Brown (1968), who gave painstaking descriptions of the series of layered rocks. They made some attempts at a theoretical description, as did McBirney and Noyes (1979), based on analogous processes in chemical reaction-diffusion theory. Neither of these, nor any subsequent attempts at a theoretical model, have been altogether successful.

Baines and Gill (1969), Turner (1979)

Balmforth *et al.* (2001)

The basic description of boundary layer at high Rayleigh number and infinite Prandtl number was first done successfully by Turcotte and Oxburgh (1967). A more complete theory is due to Roberts (1979), although even this is not quite watertight.. The necessary numerical results to compute C in (3.57) are given by Roberts (1979) and Jimenez and Zufiria (1987). The results are slightly different, with the latter paper considering Roberts' numerical results to be wrong. For $a = O(1)$, then $2C \approx 0.1$.

Jimenez and Zufiria (1987) claim that the equivalent problem to (3.67) for the case of no-slip boundary conditions has no solution, but do not adduce details. Their inference is that the boundary layer approximation fails: this seems a hazardous conclusion.

Linden (2000, Morton *et al.* (1956).

The model of turbulent thermal convection described in section 3.9 is due to Howard (1966). Baines and Turner (1969).

Exercises

3.1

3.2 A two-dimensional, incompressible fluid flow has velocity $\mathbf{u} = (u, 0, w)$, and depends only on the coordinates x and z. Show that there is a stream function ψ satisfying $u = -\psi_z$, $w = \psi_x$, and that the vorticity

$$\boldsymbol{\omega} = \boldsymbol{\nabla} \times \mathbf{u} = -\nabla^2 \psi \mathbf{j},$$

and thus that

$$\mathbf{u} \times \boldsymbol{\omega} = (\psi_x \nabla^2 \psi, 0, \psi_z \nabla^2 \psi),$$

and hence

$$\boldsymbol{\nabla} \times (\mathbf{u} \times \boldsymbol{\omega}) = (\psi_x \nabla^2 \psi_z - \psi_z \nabla^2 \psi_x)\mathbf{j}.$$

Use the vector identity $(\mathbf{u}.\boldsymbol{\nabla})\mathbf{u} = \boldsymbol{\nabla}(\frac{1}{2}u^2) - \mathbf{u} \times \boldsymbol{\omega}$ to show that

$$\boldsymbol{\nabla} \times \frac{d\mathbf{u}}{dt} = \left[-\nabla^2 \psi_t - \psi_x \nabla^2 \psi_z + \psi_z \nabla^2 \psi_x\right] \mathbf{j}.$$

Show also that

$$\boldsymbol{\nabla} \times \theta \mathbf{k} = -\theta_x \mathbf{j},$$

and use the Cartesian identity

$$\nabla^2 \equiv \text{grad}\,\text{div} - \text{curl}\,\text{curl}$$

to show that

$$\boldsymbol{\nabla} \times \nabla^2 \mathbf{u} = -\nabla^4 \psi\, \mathbf{j},$$

and hence deduce that the momentum equation for Rayleigh–Bénard convection can be written in the form

$$\frac{1}{Pr}\left[\nabla^2\psi_t + \psi_x\nabla^2\psi_z - \psi_z\nabla^2\psi_x\right] = Ra\,\theta_x + \nabla^4\psi.$$

3.3 Suppose that σ satisfies

$$p(\sigma) \equiv \sigma^3 + a\sigma^2 + b\sigma + c = 0,$$

and that a, b and c are positive. Suppose, firstly, that the roots are all real. Show in this case that they are all negative.

Now suppose that one root (α) is real and the other two are complex conjugates $\beta \pm i\gamma$. Show that $\alpha < 0$. Show also that $\beta < 0$ if $a > \alpha$. Show that $a > \alpha$ if $p(-a) < 0$, and hence show that $\beta < 0$ if $c < ab$.

If

$$\begin{aligned} a &= K^2\left(Pr + 1 + \frac{1}{Le}\right), \\ b &= K^4\left(Pr + \frac{1}{Le} + \frac{Pr}{Le}\right) + \frac{k^2}{K^2}Pr(Rs - Ra), \\ c &= \frac{K^6}{Le}Pr + k^2 Pr\left(Rs - \frac{Ra}{Le}\right), \end{aligned}$$

show that $a, b, c > 0$ if $Ra < 0$, $Rs > 0$, and show that if $Le > 1$, then $c < ab$.

What does this tell you about the stability of a layer of fluid which is both thermally and salinely stably stratified?

3.4 Suppose that σ satisfies

$$p(\sigma) \equiv \sigma^3 + a\sigma^2 + b\sigma + c = 0,$$

and that all the roots have negative real part if $c < ab$. Show that the condition that there be two purely imaginary roots $\pm i\Omega$ is that $c = ab$, and deduce that there are two (complex) roots with positive real part if $c > ab$. With

$$\begin{aligned} a &= K^2\left(Pr + 1 + \frac{1}{Le}\right), \\ b &= K^4\left(Pr + \frac{1}{Le} + \frac{Pr}{Le}\right) + \frac{k^2}{K^2}Pr(Rs - Ra), \\ c &= \frac{K^6}{Le}Pr + k^2 Pr\left(Rs - \frac{Ra}{Le}\right), \end{aligned}$$

show that this condition reduces to

$$Ra > \frac{\left(Pr + \frac{1}{Le}\right)Rs}{1 + Pr} + \frac{\left(1 + \frac{1}{Le}\right)\left(Pr + \frac{1}{Le}\right)}{Pr}\frac{K^6}{k^2}.$$

Assuming $K^2 = k^2 + m^2\pi^2$, where m is an integer, show that the minimum value of Ra where this condition is satisfied is when $m = 1$, and give the corresponding critical value $Ra_{\rm osc}$.

3.5 On the line XV in figure 3.7, the cubic

$$p(\sigma) = \sigma^3 + a\sigma^2 + b\sigma + c$$

has two positive real roots β and one negative real root α. Show that the condition for this to be the case is that

$$a = \alpha - 2\beta, \quad b = \beta^2 - 2\alpha\beta, \quad c = \alpha\beta^2,$$

and deduce that

$$a\beta^2 + 2b\beta + 3c = 0. \quad (1)$$

Show also that at the double root β,

$$3\beta^2 + 2a\beta + b = 0. \quad (2)$$

Deduce from (1) and (2) that

$$\beta = \frac{9c - ab}{a^2 - 6b},$$

and hence, using (2), that

$$\beta = \tfrac{1}{3}\left[-a + \{a^2 - 3b\}^{1/2}\right]. \quad (3)$$

Explain why the positive root is taken in (3), and why we can assume $b < 0$.

Use the definitions

$$\begin{aligned} a &= K^2\left(Pr + 1 + \frac{1}{Le}\right), \\ b &= K^4\left(Pr + \frac{1}{Le} + \frac{Pr}{Le}\right) + \frac{k^2}{K^2}Pr(Rs - Ra), \\ c &= \frac{K^6}{Le}Pr + k^2 Pr\left(Rs - \frac{Ra}{Le}\right), \end{aligned}$$

to show that if $Ra \sim Rs \gg 1$, $Ra - Rs \gg 1$ and $Le \gg 1$, then XV is approximately given by

$$Ra \approx Rs + \frac{3K^2 Rs^{2/3}}{(4k^2 Pr)^{2/3}}.$$

3.6 The growth rate σ for finger instabilities is given by

$$(\sigma + K^2 Pr)(\sigma + K^2)\left(\sigma + \frac{K^2}{Le}\right) + k^2 Pr\left[\frac{(Rs - Ra)\sigma}{K^2} + Rs - \frac{Ra}{Le}\right] = 0,$$

and $Ra, Rs < 0$ with $-Ra, -Rs \gg 1$; K is defined by $K^2 = k^2 + \pi^2$.

Define $Rs = Ra\, r$, and consider the behaviour of the roots when $Ra \to -\infty$ with r fixed. Show that when k is $O(1)$, one root is given by

$$\sigma = \frac{\left(r - \dfrac{1}{Le}\right) K^2}{1 - r} + O\left(\frac{1}{|Ra|}\right), \quad (*)$$

and that this is positive if

$$\frac{1}{Le} < r < 1.$$

Show that the other two roots are of $O\left(|Ra|^{1/2}\right)$, and by putting

$$\sigma = |Ra|^{1/2}\Sigma_0 + \Sigma_1 + \ldots,$$

show that they are given by

$$\sigma = \pm i\frac{k}{K}\{Pr(Ra - Rs)\}^{1/2} - \tfrac{1}{2}K^2\left(Pr + \frac{1 - \dfrac{1}{Le}}{1 - r}\right) + O\left(\frac{1}{|Ra|^{1/2}}\right),$$

and thus represent stable modes.

Show further that when k is large, an appropriate scaling when $(*)$ breaks down is given by

$$k = |Ra|^{1/4}\alpha, \quad \sigma = |Ra|^{1/4}\Sigma,$$

and write down the equation satisfied by Σ in this case. Show also that when α is large, the three roots are all negative, with $\Sigma \sim -\alpha^2 S$, and $S = Pr$, 1, or $\dfrac{1}{Le}$.

Deduce that the maximal growth rate for finger instability occurs when $k \sim |Ra|^{1/4}$.

3.7 The scaled Boussinesq equations for two-dimensional thermal convection at infinite Prandtl number and large Rayleigh number R in $0 < x < a$, $0 < z < 1$, can be written in the form

$$\omega = -\nabla^2\psi,$$

$$\nabla^2\omega = \frac{1}{\delta}T_x,$$

$$\psi_x T_z - \psi_z T_x = \delta^2\nabla^2 T,$$

where $\delta = R^{-1/3}$. Explain what is meant by the Boussinesq approximation, and explain what the equations represent. Explain why suitable boundary conditions for these equations which represent convection in a box with stress free boundaries, as appropriate to convection in the Earth's mantle, are given by

$$\psi = 0, \quad \omega = 0, \quad \text{on} \quad x = 0,\ a, \quad z = 0,\ 1,$$

$$T = \tfrac{1}{2} \quad \text{on} \quad z = 0, \quad T = -\tfrac{1}{2} \quad \text{on} \quad z = 1, \quad T_x = 0 \quad \text{on} \quad x = 0,\ a.$$

Show that, if $\delta \ll 1$, there is an interior 'core' in which $T \approx 0$, $\nabla^4 \psi = 0$.

By writing $1 - z = \delta Z$, $\psi = \delta \Psi$ and $\omega = \delta \Omega$, show that $\Psi \approx u_s(x) Z$, and deduce that the temperature in the thermal boundary layer at the surface is described by the approximate equation

$$u_s T_x - Z u_s' T_Z \approx T_{ZZ},$$

with

$$T = -\tfrac{1}{2} \quad \text{on} \quad Z = 0, \quad T \to 0 \quad \text{as} \quad Z \to \infty.$$

If u_s is constant, find a similarity solution, and show that the scaled surface heat flux $q = \partial T / \partial Z|_{Z=0}$ is given by

$$q = \frac{1}{2}\sqrt{\frac{u_s}{\pi x}}.$$

3.8 An isolated turbulent cylindrical plume in a stratified medium of density $\rho_0(z)$ is described by the inviscid Boussinesq equations

$$u u_r + w u_z = -\frac{1}{\rho_0} p_r,$$

$$u w_r + w w_z = -\frac{1}{\rho_0} p_z - \frac{\rho}{\rho_0} g,$$

$$u \rho_r + w \rho_z = 0,$$

$$\frac{1}{r}(ru)_r + w_z = 0,$$

where (r, z) are cylindrical coordinates, (u, w) the corresponding velocity components, p the pressure, ρ the density, ρ_0 the reference density, and g is the acceleration due to gravity. If $\rho = \rho_0 - \Delta\rho$, explain what is meant by the Boussinesq approximation.

Suppose the edge of the plume is at radius $r = b$, such that $w = 0$ there. Suppose also that the plume entrains ambient fluid, such that

$$(ru)|_b = -b\alpha\bar{w},$$

where $\bar{w}$ denotes the cross-sectional average value of w. Deduce that the plume volume flux

$$Q = 2\pi \int_0^b rw\, dr$$

satisfies

$$\frac{dQ}{dz} = 2\pi\alpha b\bar{w}.$$

The momentum flux is defined by

$$M = 2\pi \int_0^b rw^2\, dr.$$

Show that, assuming that

$$\frac{\partial p}{\partial z} = -\rho_0 g$$

throughout the plume, that

$$\frac{dM}{dz} = 2\pi \int_0^b rg'\, dr,$$

where

$$g' = \frac{g\Delta\rho}{\rho_0}.$$

Why would the hydrostatic approximation be appropriate?

The buoyancy flux is defined by

$$B = 2\pi \int_0^b rwg'\, dr;$$

assuming $g' = 0$ at $r = b$, show that

$$\frac{dB}{dz} = -N^2 Q,$$

where the Brunt–Väisälä frequency N is defined by

$$N = \left(-\frac{g\rho_0'(z)}{\rho_0}\right)^{1/2}.$$

3.9 The buoyancy flux B, momentum flux M, and mass flux Q of a turbulent plume in a stratified atmosphere satisfy the equations

$$\frac{dB}{dz} = -N^2 Q,$$

$$\frac{dM}{dz} = 2\pi \int_0^b rg'\, dr,$$

$$\frac{dQ}{dz} = 2\pi\alpha b w,$$

where w is the plume velocity, b is its radius, g' is the reduced gravity, N is the Brunt–Väisälä frequency, $\alpha \approx 0.1$ is an entrainment coefficient, and r and z are radial and axial coordinates. Assuming that

$$2\pi \int_0^b rA\,dr = \pi b^2 A$$

for any plume quantity, assumed to be approximated by a top hat profile, show that

$$\frac{dB}{dz} = -N^2 Q,$$
$$\frac{dM}{dz} = \frac{BQ}{M},$$
$$\frac{dQ}{dz} = 2\pi^{1/2}\alpha M^{1/2}.$$

Now suppose that $B = B_0$, $M = Q = 0$ at $z = 0$. By non-dimensionalising the equations appropriately, show that the level of neutral buoyancy where $B = 0$ is given by

$$z_s = \frac{\zeta_s}{(2\alpha\pi^{1/2})^{1/2}} \frac{B_0^{1/4}}{N^{3/4}},$$

where ζ_s is a numerical constant (it is approximately measured to be 1.5). Write down the equations and boundary conditions necessary to determine ζ_s, and by integrating them, show that

$$\zeta_s = \int_0^1 \frac{db}{\left[2\int_b^1 (1-\beta^2)^{1/4}\,d\beta\right]^{1/2}}.$$

If, instead, $w = w_0$ and $b = b_0$ at $z = 0$, show that the same model to determine z_s is valid provided w_0 and b_0 are small enough, and specifically if

$$w_0 \ll \frac{g'}{N}, \quad b_0^2 w_0 \ll \frac{g'^3}{N^5}.$$

Show that if the first inequality is satisfied, then the second is also provided

$$b_0 \lesssim \frac{g'}{N^2}.$$

If the scale height of the medium is h (i. e., $\rho_0'/\rho \sim 1/h$), show that these two inequalities take the form

$$w_0 \ll \frac{\Delta\rho}{\rho_0}\sqrt{gh}, \quad b_0 \lesssim \frac{\Delta\rho}{\rho_0} h.$$

Chapter 4

Atmospheric and oceanic circulation

If we had to define what environmental fluid dynamics was about, we might be tempted to limit ourselves to physical oceanography and numerical weather prediction. The wind and the sea are the most obvious examples of fluids in motion around us, and the nightly weather forecast is a commonplace in our perception of our surroundings. Certainly, groundwater levels and river flood forecasting are other environmental fluid flows of concern, but they are more often associated with stochastic behaviour and uncertainty, whereas we all know that ocean currents and weather systems are described, however inexactly, by partial differential equations. The general idea (which may or may not be correct) is that we know, at least in principle, the governing equations. The difficulty with weather prediction is then that the solutions are chaotic.[1]

Oceanography and atmospheric sciences, together tagged with the epithet of geophysical fluid dynamics (GFD), are huge and related subjects which each can and do have whole books devoted to them. In this chapter we describe briefly some of the principal phenomena of GFD with a view to making sense of how the Earth's oceans and winds operate.

The atmosphere is a layer of thin fluid draped around the Earth. The Earth has a radius of some 6,370 kilometres, but the bulk of the atmosphere lies in a film only 10 kilometres deep. This layer is called the *troposphere*. The atmosphere extends above this, into the stratosphere and then the mesosphere, but the fluid density is very small in these upper layers (though not inconsequential), and we will simplify the discussion by conceiving of atmospheric fluid motion as being (largely) confined to the troposphere.

Atmospheric winds (and thus weather) are driven by heating from the sun. The sun heats the Earth non-uniformly, because of the curvature of the Earth's surface, but the outgoing long wave radiation is much more uniform. Consequently, there

[1]This paradigm, that we know the model but can't solve it well enough, is one which is a matter of current concern in weather forecasting circles.

is an energy imbalance between the equator and the poles. The equator is differentially heated, and the poles are differentially cooled. It is important to realise that the primary climatic energy balance (which determines the mean temperature of the Earth) is between net incoming short wave radiation and outgoing long wave radiation; the Earth's weather systems and general circulation arise as a consequence of spatial variation in this balance, and as such are a perturbation to the basic energy balance. Weather is a *detail*.

The oceans are similar. The fluid is water and not air, but the oceans also lie in a thin layer on the Earth. For various reasons, their motion is more complicated and less well understood. For a start, their motion is baulked by continents. The great oceans lie in basins, and their global circulation is dictated to some extent by the topography of these basins. The atmosphere may have to flow over mountains, but it can do so: oceans have to flow round continents.

In addition, the oceans are driven not only by the same differential heating which drives the atmosphere, but also by the atmospheric winds themselves; this is the *wind driven circulation*. It is not even clear whether this is the primary driving force. A final complication is that the density of ocean water depends on salinity as well as temperature, so that oceanic convection is double-diffusive in nature. (One might say in compensation that cloud formation in the atmosphere means that atmospheric convection is multi-phase convection, but this is not conceived of as being fundamental to the nature of atmospheric motion.)

The basic nature of the atmospheric general circulation is thus that it is a convecting fluid. Hot air rises, and so the equatorial air will rise at the expense of the cold polar air. In the simplest situation, the Earth's differential heating would drive a convection cell with warm air rising in the tropics and sinking at the poles; this circulation is called the *Hadley circulation*.

In reality, the hemispheric circulation consists of three cells rather than one. The tropical cell (terminating at about 30° latitude) is still called the Hadley cell, then there is a mid-latitude cell and a polar cell. This basic circulation is strongly distorted by the rotation of the Earth, which as we shall see is rapid, so that the north/south Hadley type circulations are flung to the east (at mid-latitudes): hence the prevailing westerly winds of common European experience.[2]

This eastwards wind is called the *zonal wind*. And it is unstable: a phenomenon called *baroclinic instability* causes the uniform zonal wind to form north to south waves, and these meandering waves form the weather systems which can be seen on television weather forecast charts. At a smaller scale, such instabilities lead to weather fronts, essentially like shocks, and in the tropics these lead to cyclones and hurricanes. In order to begin to understand how this all works, we need a mathematical model, and this is essentially a model of shallow water theory (or shallow air theory) on a rapidly rotating sphere.

[2]A westerly wind is one coming from the west. It will be less confusing to call such a wind eastwards, and *vice versa* for easterlies, i. e., westwards winds.

4.1 Basic equations

The basic equations describing atmospheric (or indeed, oceanic) motion are those of mass, momentum and energy in a rotating frame, and can be written in the form

$$\begin{aligned} \frac{d\rho}{dt} + \rho \boldsymbol{\nabla}.\,\mathbf{u} &= 0, \\ \rho\left\{\frac{d\mathbf{u}}{dt} + 2\boldsymbol{\Omega}\times\mathbf{u}\right\} &= -\boldsymbol{\nabla}p - \rho\boldsymbol{\nabla}\Phi + \mathbf{F}, \\ \rho c_p \frac{dT}{dt} - \beta T\frac{dp}{dt} &= \boldsymbol{\nabla}.\,\mathbf{q} + Q. \end{aligned} \tag{4.1}$$

In these equations, ρ is the density, $\mathbf{u}$ is the velocity, p is the pressure, T is the temperature. d/dt is the material derivative following a fluid element, i. e., $d/dt = \partial/\partial t + \mathbf{u}.\boldsymbol{\nabla}$. $\boldsymbol{\Omega}$ is the angular velocity of the Earth, and the equations have been written with respect to a set of coordinates fixed in the (rotating) Earth.[3]

$\mathbf{q}$ is the heat flux, and is principally due to turbulent thermal conduction (molecular thermal conductivity is negligible) and radiative transport. The latter is awkward to quantify, but in the limit of an opaque, 'grey' atmosphere, it can be approximated by a Fourier heat conduction law, with an effective 'radiative' thermal conductivity

$$k_R = \frac{16\sigma T^3}{3\kappa\rho}; \tag{4.2}$$

here σ is the Stefan–Boltzmann constant, and κ is the radiative absorption coefficient (independent of wavelength in a grey atmosphere). If the eddy thermal conductivity is k_T, then we suppose

$$\mathbf{q} = -\bar{k}\boldsymbol{\nabla}T, \tag{4.3}$$

where

$$\bar{k} = k_R + k_T. \tag{4.4}$$

Φ is called the geopotential; it is the gravitational potential corrected for the effect of centrifugal force, and is defined by

$$\Phi = \Phi_g - \tfrac{1}{2}\left|\boldsymbol{\Omega}\times\mathbf{r}\right|^2, \tag{4.5}$$

where Φ_g is the gravitational potential. The surface $\Phi = 0$ is called sea level; the surface of the oceans would be this geopotential surface in the absence of motion. We take z to be the coordinate normal to $\Phi = 0$; essentially it is in the radial direction, and to a good approximation we can take $\Phi = gz$, where g is called the gravitational acceleration (although in fact it includes a small component due to centrifugal force).

[3]The effect of the rotating coordinate system is that time derivatives of vectors $\mathbf{a}$ are transformed as $\left.\frac{d\mathbf{a}}{dt}\right|_{\text{fix}} = \left.\frac{d\mathbf{a}}{dt}\right|_{\text{rot}} + \boldsymbol{\Omega}\times\mathbf{a}$, because in differentiating $\mathbf{a} = a_i\mathbf{e}_i$, both the components a_i and the unit vectors $\mathbf{e}_i$ change with time, and $\dot{\mathbf{e}}_i = \boldsymbol{\Omega}\times\mathbf{e}_i$.

(4.1) must be supplemented by an equation of state. In the atmosphere, we take the perfect gas law

$$p = \frac{\rho RT}{M_a}, \tag{4.6}$$

where R is the gas constant and M_a is the molecular weight of dry air.

4.1.1 Eddy viscosity

The force $\mathbf{F}$ represents the effects of friction. Molecular viscosity is insignificant in the atmosphere and oceans, but the flows are turbulent, and the result of this is that momentum transport by small scale eddying motion is often modelled by a a diffusive frictional term of the form $\rho\varepsilon_T\nabla^2\mathbf{u}$, where ε_T is an 'eddy' (kinematic) viscosity. More generally, it varies with distance from rough boundaries.[4] A complication in the atmosphere is that the vertical motion is much smaller than the horizontal, and this leads to the idea that different eddy viscosities are appropriate for horizontal and vertical momentum transport. We denote these coefficients as ε_H and ε_V, and take them as constants. To be precise, we then represent the frictional terms in the form

$$\mathbf{F} = \rho\varepsilon_H\nabla_H^2\mathbf{u} + \rho\varepsilon_V\frac{\partial^2\mathbf{u}}{\partial z^2}, \tag{4.7}$$

where $\nabla_H^2 = \dfrac{\partial^2}{\partial x^2} + \dfrac{\partial^2}{\partial y^2}$, and x, y are 'horizontal' coordinates, z is the vertical coordinate. In the following chapter, we discuss a more precise definition of the relation of these local cartesian coordinates to the appropriate spherical coordinates of the system. Since the friction terms will only be important in boundary layers where the sphericity is unimportant, we need not concern ourselves with such niceties in defining $\mathbf{F}$. Frictional effects are generally relatively small. In the atmosphere, they are confined to a 'boundary layer' adjoining the surface, having a typical depth of 1000 metres, and bulk motion above this layer is effectively inviscid.

4.2 Geostrophic flow

Two simplifications render the equations (4.1) more tractable. The first follows from the fact that the atmosphere is relatively *shallow*: the length scale of major weather systems (the synoptic scale) is of the order of thousands of kilometres, whereas the depth scale of the troposphere is about ten kilometres. Thus the aspect ratio of the flow is large.

A typical consequence of this large aspect ratio is that of lubrication theory: vertical velocities are small, and the pressure field is almost hydrostatic.

The second simplification arises from the observation that on the Earth, one form of the Rossby number

$$Ro = \frac{U}{2\Omega l} \tag{4.8}$$

[4] And then, we write $\mathbf{F} = \boldsymbol{\nabla}.\boldsymbol{\tau}_T$, $\boldsymbol{\tau}_T = \frac{1}{2}\varepsilon_T(\boldsymbol{\nabla}\mathbf{t}u + \boldsymbol{\nabla}\mathbf{t}u^T)$.

is small; here U is a typical atmospheric wind speed, Ω is the angular speed of rotation of the Earth, and l is the synoptic length scale. A typical value is $Ro \approx 0.1$. Ignoring inertia and friction, the momentum equation in the horizontal is then

$$2\rho\boldsymbol{\Omega} \times \mathbf{u} \approx -\boldsymbol{\nabla}_H p. \tag{4.9}$$

A consequence of this is that $\mathbf{u}.\boldsymbol{\nabla}_H p \approx 0$, which implies that horizontal winds are approximately along isobars. This odd behaviour (we normally expect fluid flow to be down pressure gradients) is entirely due to the rapid rotation of the planet.

Tis kind of flow is called *geostrophic flow*. In it the velocity is determined by the pressure field. The question then arises, how to determine the pressure field. The answer to this question lies in formalising the geostrophic approximation by means of an expansion in powers of the Rossby number, and this leads to an equation called the *quasi-geostrophic* equation. Its derivation is sketched in the following chapter.

4.3 The quasi-geostrophic potential vorticity equation

We now write the equations in terms of spherical polar coordinates. We take r to be the radius measured from the Earth's centre, λ to be the angle of latitude, and ϕ to be the angle of longitude. In terms of the more usual definition of spherical polar coordinates (r, θ, ϕ), r and ϕ are the same, and $\lambda = \frac{\pi}{2} - \theta$. We denote velocity components in ϕ, λ, r directions as u, v, w (because we are setting up ϕ, λ, r, i. e., east, north, upwards, as future x, y, z cartesian variables), and we denote the vector velocity $\mathbf{u} = (u, v, w)$.[5] Then the material derivative takes the form

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u}.\boldsymbol{\nabla}, \tag{4.10}$$

and conservation of mass can be written in the form

$$\frac{d\rho}{dt} + \rho\boldsymbol{\nabla}.\,\mathbf{u} = 0, \tag{4.11}$$

where the definitions of the vector derivatives are

$$\begin{aligned} \boldsymbol{\nabla}.\,\mathbf{u} &= \frac{1}{r\cos\lambda}\frac{\partial u}{\partial \phi} + \frac{1}{r\cos\lambda}\frac{\partial (v\cos\lambda)}{\partial \lambda} + \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 w\right), \\ \boldsymbol{\nabla} &= \left(\frac{1}{r\cos\lambda}\frac{\partial}{\partial \phi}, \frac{1}{r}\frac{\partial}{\partial \lambda}, \frac{\partial}{\partial r}\right). \end{aligned} \tag{4.12}$$

[5]It should be pointed out that the Earth deviates noticeably from being a sphere; it is more nearly an oblate spheroid, whose radius varies by some 20 km between pole and equator. This is of some conceptual importance, since gravity is the most important force, and the use of a purely spherical coordinate system would yield large 'horizontal' forces in the momentum equations. The correct procedure is to define the level 'horizontal' surfaces to be geopotential surfaces, so that there are no horizontal gravitational forces. But the geometric deviation from sphericity is so small that in effect we regain the form of the equations in spherical polars, as presented here.

The momentum equations have the form

$$\begin{aligned}
&\frac{du}{dt}+\frac{uw}{r}-\frac{uv}{r}\tan\lambda-2\Omega v\sin\lambda+2\Omega w\cos\lambda=-\frac{1}{\rho r\cos\lambda}\frac{\partial p}{\partial\phi}+\frac{F_\phi}{\rho},\\
&\frac{dv}{dt}+\frac{vw}{r}+\frac{u^2}{r}\tan\lambda+2\Omega u\sin\lambda=-\frac{1}{\rho r}\frac{\partial p}{\partial\lambda}+\frac{F_\lambda}{\rho},\\
&\frac{dw}{dt}-\frac{(u^2+v^2)}{r}-2\Omega u\cos\lambda=-\frac{1}{\rho}\frac{\partial p}{\partial r}-g+\frac{F_r}{\rho}.
\end{aligned}\tag{4.13}$$

The energy equation is

$$\rho c_p\frac{dT}{dt}-\frac{dp}{dt}=\boldsymbol{\nabla}.(\bar{k}\boldsymbol{\nabla}T)+Q_a+\rho LC,\tag{4.14}$$

where we assume that $\bar{k}$ represents a combined effective radiative and sensible thermal conductivity.

These are awkward equations, but they can be simplified by scaling and approximation. One of the features of the Earth's weather systems is that they have a horizontal length scale which, though large, is not global in extent. The description of such systems is facilitated by using a local, near cartesian coordinate system.

However, there is a difficulty in doing this. It is necessary to choose a particular latitude on which to put the cartesian origin, and this then limits the applicability of the resulting approximate model to phenomena appropriate to this latitude. Luckily, as we have seen, there is a natural division of the global circulation into three bands (one in each hemisphere): tropical, mid-latitude and polar. We associate these three latitudes with values of λ near zero, of $O(1)$, and near $\pm\frac{\pi}{2}$. Particularly, the polar régime is an awkward one, because of the degeneracy of the equations near $\lambda=\frac{\pi}{2}$. We will concentrate our discussion on mid-latitude phenomena, and take $\lambda=\lambda_0$ to define the x-z plane.

Specifically, we define east, north and vertical coordinates x, y and z by the relations

$$x=\phi r\cos\lambda_0,\quad y=(\lambda-\lambda_0)r,\quad z=r-r_0,\tag{4.15}$$

where r_0 is the radius at sea level. We then have

$$\frac{1}{r\cos\lambda}\frac{\partial}{\partial\phi}=\mu\frac{\partial}{\partial x},\quad \frac{1}{r}\frac{\partial}{\partial\lambda}=\frac{\partial}{\partial y},\quad \frac{\partial}{\partial r}=\frac{\partial}{\partial z}+\frac{1}{r}\left(x\frac{\partial}{\partial x}+y\frac{\partial}{\partial y}\right),\tag{4.16}$$

where

$$\mu=\frac{\cos\lambda_0}{\cos\lambda},\tag{4.17}$$

so that

$$\begin{aligned}
\boldsymbol{\nabla}&=\left(\mu\frac{\partial}{\partial x},\frac{\partial}{\partial y},\frac{\partial}{\partial z}\right)+\frac{\mathbf{k}}{r}\left(x\frac{\partial}{\partial x}+y\frac{\partial}{\partial y}\right),\\
\boldsymbol{\nabla}.\mathbf{u}&=\mu\frac{\partial u}{\partial x}+\mu\frac{\partial(v/\mu)}{\partial y}+\frac{\partial w}{\partial z}+\frac{1}{r}\left\{x\frac{\partial w}{\partial x}+y\frac{\partial w}{\partial y}+2w\right\}.
\end{aligned}\tag{4.18}$$

The mass and energy equations are still (4.11) and (4.14), and the momentum equations are then

$$\begin{aligned}
&\frac{du}{dt} - 2\Omega v \sin\lambda + 2\Omega w \cos\lambda + \frac{1}{r}\left[uw - uv\tan\lambda\right] = -\frac{\mu}{\rho}\frac{\partial p}{\partial x} + f_x,\\
&\frac{dv}{dt} + 2\Omega u \sin\lambda + \frac{1}{r}\left[vw + u^2\tan\lambda\right] = -\frac{1}{\rho}\frac{\partial p}{\partial y} + f_y,\\
&\frac{dw}{dt} - 2\Omega u\cos\lambda - \frac{(u^2+v^2)}{r} = -\frac{1}{\rho}\left[\frac{\partial p}{\partial z} + \frac{1}{r}\left(x\frac{\partial p}{\partial x} + y\frac{\partial p}{\partial y}\right)\right] - g + f_z,
\end{aligned} \tag{4.19}$$

where

$$f_x = \frac{F_\phi}{\rho}, \quad f_y = \frac{F_\lambda}{\rho}, \quad f_z = \frac{F_r}{\rho}. \tag{4.20}$$

Following (4.7), we take the vector $\mathbf{f} = (f_x, f_y, f_z)$ as

$$\mathbf{f} = \mathcal{F}\mathbf{u}, \tag{4.21}$$

where

$$\mathcal{F} = \varepsilon_H\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + \varepsilon_V\frac{\partial^2}{\partial z^2}. \tag{4.22}$$

4.4 Non-dimensionalisation

There are three obvious length scales of immediate relevance. These are the depth h of the troposphere, the radius r_0 of the Earth, and the length scale l of horizontal atmospheric motions. We have $h = 10$ km, $r_0 = 6370$ km, and the largest (*synoptic*) scales of mid-latitude weather systems are observed to be $l = 1000$ km. These lengths combine to form two dimensionless parameters,

$$\delta = \frac{h}{l}, \quad \Sigma = \frac{l}{r_0}. \tag{4.23}$$

both of which are small: $\delta \approx 0.01$, $\Sigma \approx 0.16$. The ideas of lubrication theory, using the fact that $\delta \ll 1$, suggest that in the vertical momentum equation, $\dfrac{\partial p}{\partial z} \approx -\rho g$, i. e., the pressure is approximately hydrostatic, as in our basic state. Lubrication theory also suggests that if U is a suitable horizontal velocity scale, then the appropriate vertical velocity scale is hU/l, in order that the material derivative retains vertical acceleration.

Sphericity in the equations is manifested by the terms in $1/r$ and the trigonometric terms in λ. The terms in $1/r$ are generally small, of order Σ or less, and serve as a regular perturbation to the cartesian derivative terms, except near the poles, where $\tan\lambda \to \infty$ and a different discussion is necessary.

We scale the variables as follows:

$$\begin{aligned}
&x, y \sim l, \quad z \sim h, \quad u, v \sim U, \quad w \sim \delta U,\\
&t \sim \frac{l}{U}, \quad \rho \sim \rho_0, \quad p \sim p_0, \quad T \sim T_0,
\end{aligned} \tag{4.24}$$

where we choose

$$p_0 = \frac{\rho_0 R T_0}{M_a} = \rho_0 g h \tag{4.25}$$

(which actually defines h as the (dry) atmospheric scale height, cf. question 4.2). The length scales l and r_0 are those we have described, the horizontal wind speed U is typically about 20 m s^{-1}, and the density and temperature scales ρ_0 and T_0 are their values at sea level. (These are determined by the mass of the atmosphere and the effective radiative temperature.) For the moment we assume they are constant. This is a reasonable approximation for ρ_0 but less so for temperature.

We then have dimensionless expressions

$$\begin{aligned} \boldsymbol{\nabla}.\mathbf{u} &= \mu\frac{\partial u}{\partial x} + \frac{1}{\cos\lambda}\frac{\partial (v\cos\lambda)}{\partial y} + \frac{\partial w}{\partial z} + \delta\Sigma\left\{x\frac{\partial w}{\partial x} + y\frac{\partial w}{\partial y} + 2w\right\}, \\ \frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u}.\boldsymbol{\nabla} &= \frac{\partial}{\partial t} + \mu u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z} + \delta\Sigma w\left(x\frac{\partial}{\partial x} + y\frac{\partial}{\partial y}\right), \end{aligned} \tag{4.26}$$

and the momentum equations take the dimensionless form

$$\begin{aligned} &Ro\frac{du}{dt} - v\left(\sin\lambda + \tfrac{1}{2}Ro\Sigma v\tan\lambda\right) + \delta w\left(\cos\lambda + \tfrac{1}{2}Ro\Sigma u\right) = -\frac{Ro}{F^2}\frac{\mu}{\rho}\frac{\partial p}{\partial x} + f_x^*, \\ &Ro\frac{dv}{dt} + u\left(\sin\lambda + \tfrac{1}{2}Ro\Sigma u\tan\lambda\right) + \tfrac{1}{2}\delta Ro\Sigma vw = -\frac{Ro}{F^2}\frac{1}{\rho}\frac{\partial p}{\partial y} + f_y^*, \\ &\delta\left[\delta Ro\frac{dw}{dt} - u\cos\lambda - Ro\Sigma\left(u^2 + v^2\right)\right] \\ &\qquad = -\frac{Ro}{F^2}\left[\frac{1}{\rho}\left\{\frac{\partial p}{\partial z} + \delta\Sigma\left(x\frac{\partial p}{\partial x} + y\frac{\partial p}{\partial y}\right)\right\} + 1\right] + \delta f_z^*, \end{aligned} \tag{4.27}$$

in which

$$\lambda = \lambda_0 + \Sigma y, \tag{4.28}$$

$$f_k^* = \frac{f_k}{2\Omega U} \quad \text{for} \quad k = x,\, y,\, z, \tag{4.29}$$

and the extra parameters are a form of the Rossby number,

$$Ro = \frac{U}{2\Omega l}, \tag{4.30}$$

and the Froude number

$$F = \frac{U}{\sqrt{gh}}. \tag{4.31}$$

For $U = 20$ m s^{-1}, $\Omega = 0.7\times10^{-4}$ s^{-1}, $l = 10^3$ km, $g = 10$ m s^{-2}, $h = 10$ km, we have $Ro \approx 0.14$, $F \approx 0.06$, and thus $F^2/Ro \approx 0.03$. Evidently the pressure is essentially hydrostatic, as we expect for a shallow flow.

It is convenient also to define a dimensionless measure of the (vertical) friction term. This is the Ekman number, defined by

$$E = \frac{\varepsilon_V}{fh^2}, \tag{4.32}$$

where ε_V is the vertical eddy diffusivity defined in (4.22), and f is the Coriolis parameter, given by

$$f = 2\Omega \sin \lambda_0. \tag{4.33}$$

The energy equation is commonly written in terms of the *potential temperature*, defined as

$$\theta = T \left(\frac{p_0}{p} \right)^{R/M_a c_p}; \tag{4.34}$$

the use of this variable is that

$$\rho c_p T \frac{d\theta}{\theta} = \rho c_p \, dT - dp, \tag{4.35}$$

so that θ is constant for the dry adiabatic basic state of question 4.2.[6] If we scale θ as well as T with T_0, then the dimensionless definition of θ is

$$\theta = \frac{T}{p^\alpha}, \tag{4.36}$$

in which

$$\alpha = \frac{R}{M_a c_p}. \tag{4.37}$$

The equation of state is simply

$$\rho = \frac{p}{T}. \tag{4.38}$$

The dimensionless energy equation takes the form

$$\frac{p}{\theta} \frac{d\theta}{dt} = \frac{1}{Pe} \left[\frac{\partial}{\partial z} \left(k^* \frac{\partial T}{\partial z} \right) + O(\delta^2) \right] + Q_a^* + C^*; \tag{4.39}$$

the reduced Péclet number, internal heating rate and condensation rate are given by

$$Pe = \frac{Uh^2}{\kappa_0 l}, \quad Q_a^* = \frac{Q_a l}{\rho_0 c_p T_0 U}, \quad C^* = \frac{LCl\rho}{c_p T_0 U}, \tag{4.40}$$

we have written

$$\bar{k} = k_0 k^* \quad (k^* = O(1)), \tag{4.41}$$

and

$$\kappa_0 = \frac{k_0}{\rho_0 c_p}. \tag{4.42}$$

[6] Thus $s = c_p \ln \theta$, where s is entropy.

4.5 Parameter estimates

We have already estimated typical values $\delta \approx 0.01$, $Ro \approx 0.14$, $F \approx 0.06$, $\Sigma \approx 0.16$, and we need further to estimate values of f_k^*, Pe, Q_a^* and C^*. We estimate the internal radiative heating $Q_a h \sim 0.2\, q_i \sim 68$ W m^{-2}; using values $\rho \approx 1$ kg m^{-3}, $h \approx 10$ km, $c_p \approx 10^3$ J kg^{-1} K^{-1}, $l = 10^3$ km, $T_0 = 288$ K, $U = 20$ m s^{-1}, we obtain $Q_a^* \sim 1.2 \times 10^{-3}$. Internal radiative heating is therefore very small.

In order to estimate Pe and f_k^*, we need estimates of eddy viscosities. A typical estimate in the horizontal is $\varepsilon_H \sim 0.1\, Uh \sim 10^4$ m^2 s^{-1}, and a typical estimate in the vertical is $\varepsilon_V \sim 0.1 \delta U h \sim 10^2$ m^2 s^{-1}. Therefore $\varepsilon_H \nabla_H^2 \sim 10^{-8}$ s^{-1}, $\varepsilon_V \partial^2/\partial z^2 \sim 10^{-6}$ s^{-1}, so that the vertical diffusivity is dominant. Then $f_{x,y}^* \sim \varepsilon_V / 2\Omega h^2 \sim 10^{-2}$, while $f_z^* \sim \delta f_{x,y}^*$.

We already estimated $Pe \sim 20$ in question 4.1, based on a radiative effective thermal conductivity of $k_R \approx 10^5$ W m^{-1} K^{-1}. A corresponding estimate for the (vertical) eddy thermal conductivity is $k_T \approx \rho c_p \varepsilon_V \approx 10^5$ W m^{-1} K^{-1}, comparable to the radiative value. This suggests that $\bar{k} \sim k_0 \approx 2 \times 10^5$ W m^{-1} K^{-1} is a reasonable estimate, which would then suggest that $Pe \sim 10$.

In order to estimate the dimensionless condensation rate C^*, we suppose that $C \approx -dm/dt$, assuming saturation. We use the expression, obtained by solving the Clausius–Clapeyron equation,

$$p_{SV} = p_{SV}^0 \exp\left[a\left(1 - \frac{T_0}{T}\right)\right], \tag{4.43}$$

where

$$a = \frac{M_v L}{R T_0}. \tag{4.44}$$

Appropriate values are $a \approx 18.8$ and $p_{SV}^0 \approx 1,688$ Pa.[7] From these we find

$$m = \frac{M_v p_{SV}^0}{M_a p} \exp\left[a\left(1 - \frac{T_0}{T}\right)\right], \tag{4.45}$$

and in terms of the dimensionless temperature and pressure,

$$m = \nu M(T, p), \tag{4.46}$$

where

$$M(T, p) = \frac{1}{p} \exp\left[a\left(1 - \frac{1}{T}\right)\right], \tag{4.47}$$

and

$$\nu = \frac{M_v p_{SV}^0}{M_a p_0}. \tag{4.48}$$

[7]This is different from the triple point value of 600 Pa because we use 288 K as the reference temperature, not 273 K.

Approximately, $\nu \approx 0.01$. In dimensionless terms, we thus have

$$C^* = \nu St \left(-\rho \frac{dM}{dt}\right), \tag{4.49}$$

where the Stefan number is

$$St = \frac{L}{c_p T_0}. \tag{4.50}$$

The value of St is 8.7, so that $\nu St \approx 0.087$. Because a is large and M is $O(1)$, $dM/dt \sim aM$, and thus $C^* \sim O(1)$ (the value of $\nu St\, a$ is ≈ 1.6).

4.5.1 Basic reference state

Using the definitions of M in (4.47), and of ρ and T in (4.38) and (4.36), we can write the energy equation (4.39) in the form

$$\frac{p}{\theta}\left[1 + \frac{\nu St\, aM}{T^2}\right]\frac{d\theta}{dt} = -\frac{\nu St\, M(\alpha a - T)}{T^2}\frac{dp}{dt} + \frac{1}{Pe}\left[\frac{\partial}{\partial z}\left(k^* \frac{\partial T}{\partial z}\right)\right], \tag{4.51}$$

in which we neglect Q_a^* and $O(\delta^2 Pe)$.

If we further neglect the conductive term of $O(1/Pe)$, then to leading order, (4.27) and (4.51) can be written as

$$\frac{\partial p}{\partial z} = -\rho, \quad \frac{p}{\theta}\left[1 + \frac{\nu St\, aM}{T^2}\right]\frac{d\theta}{dt} = -\frac{\nu St\, M(\alpha a - T)}{T^2}\frac{dp}{dt}, \tag{4.52}$$

representing a wet adiabatic hydrostatically balanced atmosphere. This tells us that in such an atmosphere, θ is a well-defined function of p, and hence (because of hydrostatic balance) also of z. We define this basic wet potential temperature function as $\theta_w(p)$, and the corresponding pressure and density profiles as p_w and ρ_w. Thus θ_w and p_w are determined by solving the simultaneous differential equations (noting that $\rho = p^{1-\alpha}/\theta$ and $T = \theta p^\alpha$)

$$\begin{aligned} \frac{dp_w}{dz} &= -\frac{p_w^{1-\alpha}}{\theta_w}, \\ \frac{d\theta_w}{dz} &= \frac{\nu St\,(a\alpha - \theta_w p_w^\alpha)M}{[\theta_w^2 p_w^{2\alpha} + \nu St\, aM]p_w^\alpha}, \end{aligned} \tag{4.53}$$

with $p_w = \theta_w = 1$ at $z = 0$. We have $\alpha \approx 0.29$, $a \approx 18.8$, and thus $\alpha a \approx 5.45$, and so $d\theta/dz > 0$. Also $\nu \approx 0.01$, $St \approx 8.7$, so that $\nu St \approx 0.087$, and the potential temperature gradient appears on this basis to be small, of $O(Ro)$. Figure 4.1 shows a numerical solution of (4.53), which shows that pressure decreases approximately exponentially (with scale height of about 10 km) and θ_w increases approximately linearly, in this model. The numerical solution indicates that the potential temperature gradient is indeed small, of order 0.1. We associate this with the fact that $\nu St \approx 0.087$ is small. Below, we define a parameter ε (the Rossby number) which is of the same order as the wet potential temperature gradient; then $\theta_w = 1 + O(\varepsilon)$ defines a wet adiabat, whereas a dry reference state in which the moisture term is absent is simply $\theta = 1$. Reality is somewhere between the two, though nearer the wet state.

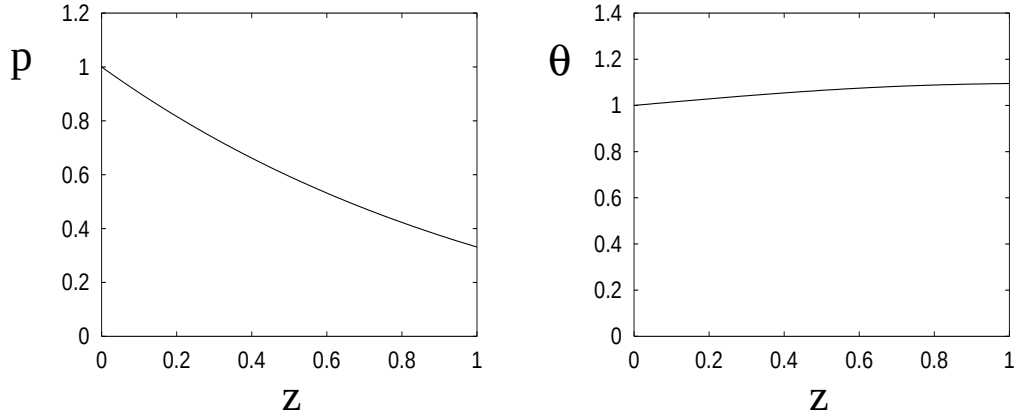

Figure 4.1: Solution of (4.53). The pressure is excellently approximated by $p \approx e^{-1.08z}$, and the potential temperature is excellently approximated by $\theta \approx 1 + 0.15z - 0.05z^2$.

4.5.2 A reduced model

In order to approximate the model, we note that

$$\begin{aligned} \delta \sim Ro\,\Sigma \sim \frac{F^2}{Ro} \sim f^*_{x,y} \sim 10^{-2}, \\ \frac{1}{Pe} \sim \nu St \sim Ro \sim \Sigma \sim 10^{-1}, \end{aligned} \tag{4.54}$$

and the other parameters Q^*_a and f^*_z are much smaller. These suggest that we should think of $1/Pe$, νSt, Ro and Σ as small, but an order of magnitude larger than δ and F^2/Ro. In fact, $\partial T/\partial z \sim \alpha$, and $\alpha/Pe \approx 0.04$; therefore we shall consider the conductive term in (4.51) to be of $O(\delta)$. In fact, to be specific, we now define the length scale l and velocity scale U by requiring that

$$\frac{F^2 \sin\lambda_0}{Ro} = \frac{\alpha}{Pe} = \varepsilon^2, \tag{4.55}$$

where it is conventional to define the *Rossby number* as

$$\varepsilon = \frac{Ro}{\sin\lambda_0} = \frac{U}{fl}, \tag{4.56}$$

in which the Coriolis parameter f is defined as

$$f = 2\Omega \sin\lambda_0. \tag{4.57}$$

This leads to definitions

$$U = \left(\frac{\alpha\kappa_0 g}{fh}\right)^{1/2}, \quad l = U\left(\frac{h^2}{\alpha\kappa_0 f^2}\right)^{1/3}, \tag{4.58}$$

and calculation of these using values suggested previously leads to $U \approx 26$ m s^{-1}, $l \approx 1290$ km.

Next, we adopt the formal asymptotic limits

$$\begin{aligned} &\nu St \sim Ro \sim \Sigma \sim \varepsilon, \\ &\delta \sim Ro\Sigma \sim f^*_{x,y} \sim \varepsilon^2. \end{aligned} \tag{4.59}$$

Expanding the equations in powers of ε, the vertical momentum equation is

$$\frac{\partial p}{\partial z} \approx -\rho + O(\varepsilon^3), \tag{4.60}$$

where

$$\rho = \frac{p}{T}, \quad \theta = \frac{T}{p^\alpha}. \tag{4.61}$$

Also,

$$\begin{aligned} \boldsymbol{\nabla}.\,\mathbf{u} &\approx \mu\frac{\partial u}{\partial x} + \frac{1}{\mu}\frac{\partial (v/\mu)}{\partial y} + \frac{\partial w}{\partial z} + O(\varepsilon^3), \\ \frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{u}.\boldsymbol{\nabla} &\approx \frac{\partial}{\partial t} + \mu u\frac{\partial}{\partial x} + v\frac{\partial}{\partial y} + w\frac{\partial}{\partial z} + O(\varepsilon^3), \end{aligned} \tag{4.62}$$

the horizontal momentum equations are approximately

$$\begin{aligned} \varepsilon \sin\lambda_0 \frac{du}{dt} - v\sin\lambda &= -\frac{\sin\lambda_0}{\varepsilon^2}\frac{\mu}{\rho}\frac{\partial p}{\partial x} + O(\varepsilon^2), \\ \varepsilon \sin\lambda_0 \frac{dv}{dt} + u\sin\lambda &= -\frac{\sin\lambda_0}{\varepsilon^2}\frac{1}{\rho}\frac{\partial p}{\partial y} + O(\varepsilon^2), \end{aligned} \tag{4.63}$$

and the energy equation is approximately

$$\frac{p}{\theta}\left[1 + \frac{\nu St\, aM}{T^2}\right]\frac{d\theta}{dt} = -\frac{\varepsilon s M(\alpha a - T)}{T^2}\frac{dp}{dt} + \varepsilon^2\left[\frac{\partial}{\partial z}\left(\frac{k^*}{\alpha}\frac{\partial T}{\partial z}\right)\right], \tag{4.64}$$

where we have written

$$\nu St = \varepsilon s \tag{4.65}$$

to delineate the smallness of νSt (but noting that $\nu St\, a \approx 1.64$ is $O(1)$). Together with the conservation of mass equation

$$\frac{d\rho}{dt} + \rho\boldsymbol{\nabla}.\,\mathbf{u} = 0, \tag{4.66}$$

this completes the basic approximate model, valid locally everywhere except near the poles (where μ and $\tan\lambda \to \infty$). There are seven equations in (4.60), (4.61), (4.63), (4.64) and (4.66) for the seven variables θ, ρ, T, p, u, v and w. The frictional terms $f^*_{x,y}$ can be neglected in the main flow, but they are important in the planetary boundary layer.

4.5.3 Geostrophic balance

Geostrophic flow is described by the leading order approximation which considers both curvature and inertial effects to be small, that is, $\varepsilon \ll 1$. At leading order, the pressure is hydrostatic, and (4.63) indicates that the correction is of $O(\varepsilon^2)$. This is consistent with (4.64), which indicates that $\theta = \bar{\theta}(z) + O(\varepsilon^2)$. We do not yet assume that $\bar{\theta}$ is equal to the reference state θ_w defined in (4.53); this would have to be deduced. But we do anticipate that $\bar{\theta}'(z) = O(\varepsilon)$ (since also $\theta_w' = O(\varepsilon)$.) We put

$$p = \bar{p}(z) + \varepsilon^2 P, \tag{4.67}$$

where $\bar{p}$ is the hydrostatic pressure corresponding to $\bar{\theta}$, and we denote the corresponding density and temperature as $\bar{\rho}$ and $\bar{T}$. Then, since $\mu \approx 1$, $\lambda \approx \lambda_0$ and $\rho \approx \bar{\rho}$, the momentum equations become

$$\begin{aligned} \bar{\rho} v &\approx \frac{\partial P}{\partial x}, \\ \bar{\rho} u &\approx -\frac{\partial P}{\partial y}, \end{aligned} \tag{4.68}$$

and mass conservation reduces to

$$\frac{\partial(\bar{\rho}u)}{\partial x} + \frac{\partial(\bar{\rho}v)}{\partial y} + \frac{\partial(\bar{\rho}w)}{\partial z} \approx 0. \tag{4.69}$$

Together (4.68) and (4.69) imply

$$\frac{\partial(\bar{\rho}w)}{\partial z} = 0, \tag{4.70}$$

and thus w is determined by its value on the surface, where it is prescribed by the no flow through boundary condition. In the absence of topography, we have $w = 0$ at $z = 0$, so that $w = 0$ everywhere. The flow is then purely two-dimensional, and the horizontal velocity vector $\mathbf{u}_H = (u, v)$ is given by

$$\bar{\rho}\mathbf{u}_H = \mathbf{k} \times \boldsymbol{\nabla}_H P, \tag{4.71}$$

where $\boldsymbol{\nabla}_H = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$. (4.71) defines the *geostrophic wind*, and shows that $\mathbf{u} \,.\boldsymbol{\nabla} p = 0$, i. e., wind velocities are along isobars. In the northern hemisphere, the wind moves anti-clockwise about regions of low pressure (depressions, or cyclones). The closer the isobars, the higher the wind speed.

The no flow through condition at the ground is modified by the small friction terms in the momentum equation, which causes the existence of a boundary layer (the planetary boundary layer) to occur. The effect of rotation in this boundary layer is to cause an effective vertical velocity at the top of the layer, proportional to the Ekman number E defined by (4.32), and this velocity is often called Ekman pumping (indeed, the viscous boundary layer in a rapidly rotating fluid is called an Ekman layer). The Ekman layer modifies the effective no–flow–through condition to be

$$w = \sqrt{\frac{E}{2}} \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) \quad \text{at} \quad z = 0. \tag{4.72}$$

4.6 The quasi-geostrophic approximation

We now return to the problem of finding the pressure for the geostrophic approximation in which (4.71) applies. To do this, we need to carry the approximation to next order in ε, and this will allow us to deduce the *quasi-geostrophic potential vorticity equation*. The equation of mass conservation (4.66) can be written in the form

$$\frac{\partial \rho}{\partial t} + \mu \frac{\partial (\rho u)}{\partial x} + \mu \frac{\partial (\rho v/\mu)}{\partial y} + \frac{\partial (\rho w)}{\partial z} = O(\varepsilon^3). \tag{4.73}$$

Since $w = 0$ at leading order, we put

$$w = \varepsilon W. \tag{4.74}$$

We also define the perturbed potential temperature Θ by

$$\theta = \bar{\theta}(z) + \varepsilon^2 \Theta; \tag{4.75}$$

evidently $\bar{\theta}(z)$ is the time and space-horizontal average of θ correct to $O(\varepsilon^2)$, and we can in fact define it to be the exact such average of θ, without loss of generality. More generally, we might take $\bar{\theta} = \bar{\theta}(z,t)$, but the energy equation then simply implies that $\bar{\theta}_t = 0$. We might have expected $\bar{\theta}$ to be equal to the wet adiabatic potential temperature θ_w, defined in (4.53), but as we shall see, there is a subtle distinction, and it is necessary to delineate the difference in the equations. Because the hydrostatic correction in (4.60) is $O(\varepsilon^3)$, expansion of that equation to $O(\varepsilon^2)$ yields the hydrostatic approximation for the perturbation pressure P, defined in (4.67):

$$\Theta = \bar{\theta}^2 \frac{\partial}{\partial z} \left[\frac{P}{\bar{p}^{1-\alpha}} \right]. \tag{4.76}$$

The geostrophic wind approximation (4.68) suggests that we write

$$P = \bar{\rho} \psi, \tag{4.77}$$

where ψ is the geostrophic stream function, thus

$$u = -\frac{\partial \psi}{\partial y}, \quad v = \frac{\partial \psi}{\partial x}. \tag{4.78}$$

Bearing in mind that $\bar{\rho} = \bar{p}^{1-\alpha}/\bar{\theta}$, it follows that

$$\Theta = \bar{\theta}^2 \frac{\partial}{\partial z} \left[\frac{\psi}{\bar{\theta}} \right] = \frac{\partial \psi}{\partial z} + O(\varepsilon), \tag{4.79}$$

on the assumption that $\bar{\theta}'(z) = O(\varepsilon)$. This relation, together with the geostrophic wind approximation, gives us the *thermal wind equations*:

$$\frac{\partial u}{\partial z} = -\frac{\partial \Theta}{\partial y}, \quad \frac{\partial v}{\partial z} = \frac{\partial \Theta}{\partial x}. \tag{4.80}$$

Next we form an equation for the (vertical) vorticity

$$\zeta = \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} = \nabla^2 \psi \tag{4.81}$$

by cross differentiating (4.63) (with some care) to eliminate the pressure derivatives. Using the conservation of mass equation, together with (4.74) and the fact that $\rho = \bar{\rho}(z) + O(\varepsilon^2)$, we derive the vorticity equation

$$\frac{D\zeta}{Dt} + \beta \frac{\partial \psi}{\partial x} = \frac{1}{\bar{\rho}} \frac{\partial (\bar{\rho} W)}{\partial z}, \tag{4.82}$$

where D/Dt denotes the horizontal material derivative, and the term in β arises from the variation of $\sin \lambda$ with latitude; β is defined by

$$\beta = \frac{\Sigma \cot \lambda_0}{\varepsilon}, \tag{4.83}$$

and the horizontal material derivative is defined by

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + u \frac{\partial}{\partial x} + v \frac{\partial}{\partial y} = \frac{\partial}{\partial t} - \frac{\partial \psi}{\partial y} \frac{\partial}{\partial x} + \frac{\partial \psi}{\partial x} \frac{\partial}{\partial y}. \tag{4.84}$$

Next, we consider the energy equation (4.64). Expanding in powers of ε, this can be written in the form, correct to terms of $O(\varepsilon^2)$,

$$\varepsilon W \frac{d\bar{\theta}}{dz} + \varepsilon^2 \frac{D\Theta}{Dt} = \varepsilon W \frac{d\theta_w}{dz} + \varepsilon^2 H, \tag{4.85}$$

where

$$H = \frac{\dfrac{\partial}{\partial z}\left(\dfrac{k^*}{\alpha} \dfrac{\partial \bar{T}}{\partial z} \right)}{\dfrac{\bar{p}}{\bar{\theta}} \left[1 + \dfrac{\nu St\, a M(\bar{T}, \bar{p})}{\bar{T}^2} \right]} \tag{4.86}$$

is the heating term.

Now we can see the nature of the assumption about the average potential temperature. Bearing in mind that $d\theta_w/dz = O(\varepsilon)$, we see that the *ansatz* that $d\bar{\theta}/dz = O(\varepsilon)$ is indeed correct. However, it is generally not the case that $\bar{\theta} = \theta_w$. The question then arises how to determine it.

Let us denote the stratification function $S(z)$ by

$$S(z) = \frac{1}{\varepsilon} \left[\frac{d\bar{\theta}}{dz} - \frac{d\theta_w}{dz} \right], \tag{4.87}$$

and note that by observation (and assumption) it is positive and $O(1)$. It is related to the Brunt-Väisälä frequency N, which is the frequency of small vertical oscillations in the atmosphere; in fact $S \propto N^2$. Positive S (and thus real N) indicates a stably

stratified atmosphere. If S were to become negative, the atmosphere would become unstably stratified and it would overturn. The energy equation is thus

$$\frac{D\Theta}{Dt} = H - WS. \tag{4.88}$$

In summary, we have the vorticity ζ and potential temperature Θ defined in terms of the stream function ψ by (4.81) and (4.79). Two separate equations for ζ and Θ are then (4.82) and (4.88), from which W and $S(z)$ must also be determined, the latter by averaging the equations.

By an application of Green's theorem in the plane, we have that

$$\iint_A \frac{D\Gamma}{Dt}\, dS = \frac{\partial}{\partial t} \iint_A \Gamma\, dS - \oint_{\partial A} \Gamma\, d\psi, \tag{4.89}$$

where A is any horizontal area at fixed z. In particular, if A is a closed region on the boundaries of which ψ is constant in space, i. e., there is no flow through ∂A, then the boundary integral is zero.[8] Let an overbar denote a space horizontal average over A. Putting $\Gamma = \Theta$, it follows that

$$\frac{\partial \bar{\Theta}}{\partial t} = H - \overline{W}S, \tag{4.90}$$

where $\overline{W}(z)$ is the horizontal average of W. Applying the same procedure to (4.82), we have

$$\frac{\partial \bar{\zeta}}{\partial t} = \frac{1}{\bar{\rho}} \frac{\partial}{\partial z} [\bar{\rho}\overline{W}]. \tag{4.91}$$

According to the Ekman pumping boundary condition (4.72), the value of $\overline{W}$ at $z = 0$ is

$$\overline{W}_0 = E^* \bar{\zeta}_0, \tag{4.92}$$

where $\bar{\zeta}_0$ is the space averaged vorticity at the surface, and

$$E^* = \sqrt{\frac{E}{2\varepsilon^2}}. \tag{4.93}$$

Integrating (4.91), we have (using $\bar{\rho} = 1$ at $z = 0$)

$$\bar{\rho}\overline{W} = \int_0^z \bar{\rho}\bar{\zeta}_t\, dz + E^* \bar{\zeta}_0, \tag{4.94}$$

and it follows from this that the stratification parameter is defined by the relation

$$\frac{\bar{\rho}}{S} = \frac{\displaystyle\int_0^z \bar{\rho}\bar{\zeta}_t\, dz + E^* \bar{\zeta}_0}{H - \bar{\Theta}_t}. \tag{4.95}$$

[8] We have in mind that A is the region of zonal mid-latitude flow, bounded to the north by the polar front, and to the south by the tropical front. We can allow A to be a periodic strip on the sphere also.

We can go further if we assume that the solutions are stationary (not necessarily steady), i. e., a well-defined time average exists.[9] The time averages of the time derivative terms are zero, and thus it simply follows (since H, S and $\bar{\rho}$ are functions only of z) that

$$H = \widehat{W}S, \quad \bar{\rho}\widehat{W} = \widehat{W}_0, \tag{4.96}$$

where $\widehat{W}$ is the time average of $\overline{W}$, and the constant $\widehat{W}_0$ is the value of the surface boundary value of $\widehat{W}$ at $z = 0$. The Ekman pumping boundary condition (4.92) implies that

$$\widehat{W}_0 = E^* \bar{\zeta}_0, \tag{4.97}$$

where $\bar{\zeta}_0$ is the space averaged vorticity at the surface.

The two equations in (4.96) define S and $\widehat{W}$, and in particular we find that

$$\frac{\bar{\rho}}{S} = \frac{E^* \widehat{\zeta}_0}{H}. \tag{4.98}$$

This equation thus defines the stratification function $S(z)$ for a stationary (but not necessarily steady) atmosphere.[10] Evidently, the wet adiabatic profile ($S = 0$) is obtained (in stationary conditions) only if the heating rate H is zero.

We can now use the identity

$$\frac{\partial}{\partial z}\left[K(z)\frac{D\Theta}{Dt}\right] = \frac{D}{Dt}\left[\frac{\partial}{\partial z}\left(K(z)\frac{\partial \psi}{\partial z}\right)\right] \tag{4.99}$$

to show, using (4.88), that

$$\frac{1}{\bar{\rho}}\frac{\partial}{\partial z}[\bar{\rho}W] = \frac{1}{\bar{\rho}}\frac{\partial}{\partial z}\left[\frac{\bar{\rho}H}{S}\right] - \frac{D}{Dt}\left[\frac{1}{\bar{\rho}}\frac{\partial}{\partial z}\left(\frac{\bar{\rho}}{S}\frac{\partial \psi}{\partial z}\right)\right], \tag{4.100}$$

and therefore (4.82) can be written

$$\frac{D}{Dt}\left[\nabla^2\psi + \beta y + \frac{1}{\bar{\rho}}\frac{\partial}{\partial z}\left(\frac{\bar{\rho}}{S}\frac{\partial \psi}{\partial z}\right)\right] = \frac{1}{\bar{\rho}}\frac{\partial}{\partial z}\left[\frac{\bar{\rho}H}{S}\right]. \tag{4.101}$$

This is one form of the *quasi-geostrophic potential vorticity equation*. It is a single equation for the geostrophic stream function ψ, providing the stratification S is known. In most treatments of its solutions, the stratification parameter S is assumed known (from measurements), and then the equation (4.101) can be considered on its own. However, in reality S must be determined from (4.98), which indicates that the stratification is determined in terms of the solution of the equation (4.101) itself, which is thus an integro-differential equation for the stream function ψ. This equation is only well-posed if $S > 0$, i. e., if the time mean surface value of the vorticity

[9] This is what we would generally expect. Unbounded drift of ψ would indicate breakdown of the perturbation expansion because of the presence of secular terms.

[10] This derivation is somewhat similar to that of Pedlosky (1979); however, he did not provide an explicit recipe for $S(z)$. See also question 4.3.

$\hat{\zeta}_0$ is positive (assuming also that $H > 0$), but there seems to be no obvious reason why this should necessarily be the case. If the stratification became negative, the atmosphere would become inherently unstable. It would start to 'boil', with severe convective storms shrouding the planet. The mild, quasi-geostrophic description of the weather would no longer be apt, and climate change would be a sudden, dramatic reality overnight. The day after tomorrow, indeed. The possibility of such a catastrophic scenario appears to have entirely bypassed climate scientists.

4.7 Poincaré, Kelvin and Rossby waves

The geostrophic wind given by equation (4.71) is an approximate solution to the governing equations which is *quasi-static*, in the sense that the acceleration terms in the momentum equation are ignored; implicitly, any more rapid transients have died out. In this chapter, we consider various classes of wave motion which arise in the model on this shorter transient time scale.

4.8 Gravity waves

Atmospheric motions are dominated by various kinds of waves. Two particular sorts of waves which are familiar in fluid mechanics are *sound waves* and *gravity waves*. Sound waves are associated with compressibility; they travel at a speed (the speed of sound) which depends on density but is independent of wave number: they are *monochromatic*. At sea level this speed is about 330 m s^{-1}: much faster than typical wind speeds; as a consequence, we might expect sound waves to be high frequency phenomena which are not relevant to common atmospheric motions. If we denote the sound wave speed as c_s, then the dispersion relation relating frequency ω to wave speed and wave number k is just $\omega = kc_s$. When this is written in dimensionless units, as in chapter 4, we have

$$c_s^2 = gh\bar{c}^2, \quad \bar{c} = \left(\frac{d\bar{p}}{d\bar{\rho}}\right)^{1/2}, \tag{4.102}$$

and the corresponding dimensionless dispersion relation is just

$$\frac{\omega}{k} = \frac{\bar{c}}{F}, \tag{4.103}$$

where F is the Froude number defined by (4.31). Note from (4.55) that $F = \varepsilon^{3/2}$.

Gravity waves are familiar as the waves which propagate on the surface of the sea. The ingredients of the theory which describes them are mass conservation (where horizontal divergence is accommodated by vertical contraction and expansion), acceleration, gravity, pressure gradient, and a vertical stratification which, in the simplest form of the theory, is manifested by the interface between dense underlying fluid (e. g.,

Figure 4.2: Periodic gravity waves in Lapland, Northern Finland, October 2004.

water) and a lighter overlying fluid (e. g., air). Gravity waves can be seen propagating at the interface between two incompressible liquids such as oil and water, and gravity waves will similarly propagate in a continuously stratified fluid contained in a vertically confined channel; in this case the waves are less easily visualised, and they are often called internal waves, or internal gravity waves.

In the sense that the atmosphere consists of a dense troposphere beneath a light stratosphere, we can expect gravity waves to propagate as undulations in the tropopause altitude. More generally, gravity waves will propagate as internal waves in the stratified atmosphere. Gravity waves can be seen commonly in the atmosphere, because the vertical undulations of the air causes periodic cloud formation as air rises (and thus cools). Figure 4.2 shows a particular striking example from Lapland of low lying periodic gravity waves.

For the simple case of an incompressible fluid of depth h, the dispersion relation between frequency and wavenumber is $\omega^2 = gk \tanh kh$. In the case of a shallow fluid (such as the atmosphere), the long wave limit $kh \ll 1$ may be appropriate, and then the wave speed is constant, and $\omega \approx k\sqrt{gh}$. This applies to waves of wavelength larger than 10 km (the waves in figure 4.2 are of smaller wavelength). In dimensionless terms, the dispersion relation becomes

$$\frac{\omega}{k} = \frac{1}{F}. \tag{4.104}$$

Comparing (4.104) with (4.103), we see that long gravity waves in the atmosphere are essentially the same as sound waves. In an incompressible fluid, density is manifested

as fluid column depth, and the pressure is proportional to this, so that the dimensionless 'sound' speed is equal to one. For internal waves, the height of the column need not change, but the common factor is that the height of geopotential surfaces propagates in both types of wave.

We can recover gravity waves from the scaled atmospheric model by focussing on long waves of wavenumber $k \sim O(\sqrt{\varepsilon})$, and time scales of $O(\varepsilon)$ (i. e., frequencies $\omega \sim O(1/\varepsilon)$). (Note that then $\omega/k \sim 1/\varepsilon^{3/2} = 1/F$, consistent with (4.103) and (4.104).) We write

$$t = \varepsilon\tau, \quad (x, y) = (X, Y)/\sqrt{\varepsilon}, \quad P = \frac{\Pi}{\sqrt{\varepsilon}}, \tag{4.105}$$

and retain leading order terms in equations (4.63) and (4.66), assuming that $w \sim \varepsilon$. Note that $\rho = p^{1-\alpha}/\theta$, and that $\partial\theta/\partial t \approx 0$, so that

$$\frac{1}{\rho}\frac{\partial\rho}{\partial t} \approx \left(\frac{1-\alpha}{p}\right)\frac{\partial p}{\partial t}. \tag{4.106}$$

At leading order, mass conservation takes the form

$$\left(\frac{1-\alpha}{\bar{p}}\right)\frac{\partial\Pi}{\partial\tau} + \frac{\partial u}{\partial X} + \frac{\partial v}{\partial Y} = 0; \tag{4.107}$$

compressibility and stratification are manifested by the first term in this equation.

At leading order, the momentum equations take the form

$$\begin{aligned} \frac{\partial u}{\partial\tau} - v &\approx -\frac{1}{\bar{\rho}}\frac{\partial\Pi}{\partial X}, \\ \frac{\partial v}{\partial\tau} + u &\approx -\frac{1}{\bar{\rho}}\frac{\partial\Pi}{\partial Y}. \end{aligned} \tag{4.108}$$

We can write these equations in terms of the horizontal divergence $\Delta = u_X + v_Y$, the vorticity $\zeta = v_X - u_Y$, and the pressure perturbation Π. We obtain

$$\begin{aligned} \frac{\partial\Delta}{\partial\tau} - \zeta &= -\frac{1}{\bar{\rho}}\nabla^2\Pi, \\ \frac{\partial\zeta}{\partial\tau} + \Delta &= 0, \\ \frac{\partial\Pi}{\partial\tau} + \bar{\rho}\bar{c}^2\Delta &= 0, \end{aligned} \tag{4.109}$$

where

$$\bar{c} = \left[\frac{\bar{p}}{(1-\alpha)\bar{\rho}}\right]^{1/2} \tag{4.110}$$

is the dimensionless isentropic sound speed.

These are linear equations, and solutions exist of the form

$$\begin{pmatrix} \Delta \\ \zeta \\ \Pi \end{pmatrix} = \mathbf{w} \exp\{i(kX + lY + \omega\tau\}, \tag{4.111}$$

provided

$$\begin{pmatrix} 0 & 1 & \dfrac{(k^2+l^2)}{\bar{\rho}} \\ -1 & 0 & 0 \\ -\bar{\rho}\bar{c}^2 & 0 & 0 \end{pmatrix} \mathbf{w} = i\omega\mathbf{w}. \tag{4.112}$$

Solutions to this exist provided either $\omega = 0$, or

$$\omega^2 = 1 + (k^2 + l^2)\bar{c}^2, \tag{4.113}$$

and this latter equation is the dispersion relation for gravity waves in a stratified atmosphere. These waves are called *Poincaré waves*.

Another kind of wave can be found by seeking solutions in which $v = 0$. Such waves are particularly relevant to propagation in a confined zonal channel (for example in the ocean), where the condition $v = 0$ at the north and south boundaries forces $v = 0$ everywhere. This requires $\partial\Delta/\partial Y = -\partial\zeta/\partial X$, and substitution into (4.112) then shows that we must have $l = -ik/\omega$, and thus solutions are exponential in y, and

$$\omega = k\bar{c}; \tag{4.114}$$

these waves are called *Kelvin waves*. They are *edge waves*, because they decay exponentially away from one or other boundary. Together with the geostrophic mode $\omega = 0$, Poincaré and Kelvin waves form the complete spectrum of waves for the flow. The mode $\omega = 0$ is associated with low frequency waves which emerge in the higher order quasi-geostrophic approximation (which is derived in the next section); these slow waves are called *Rossby waves*, or *planetary waves*.

The constant term in (4.113) arises from rotation and the Coriolis force. In the high frequency limit, we see that $\omega \approx k\bar{c}$ (for unidirectional waves), and this is consistent with the long wave limit of gravity wave theory, and the acoustic wave speed given in (4.103). Gravity waves are essentially long wavelength sound waves, and Poincaré waves are their modification by the effects of rotation. The critical length scale $l/\sqrt{\varepsilon}$ above which rotation becomes important is known as the Rossby radius of deformation. Using (4.58), it is found to be equal to $\sqrt{gh}/f$. For atmospheric motion, it is of order 3000 km, so that rotation is unimportant for smaller scale gravity waves.

4.9 Rossby waves

We now seek a wave motion corresponding to the zero frequency geostrophic gravity wave mode satisfying (4.112) with $\omega = 0$. This is the Rossby wave, and it is most

simply examined by studying (4.101) in the absence of heating, and assuming that the stratification parameter S is prescribed. (Such simplifications are in fact commonly made in studying the properties of (4.101).) We define a vertical eigenfunction $\Psi(z)$ satisfying the ordinary differential equation

$$\frac{1}{\bar{\rho}}\left[\frac{\bar{\rho}}{S}\Psi'\right]' = -m^2\Psi, \tag{4.115}$$

where for suitable homogeneous boundary conditions on Ψ, m^2 will be positive. With $H = 0$, $\psi = 0$ is a solution of (4.101), and small amplitude solutions of the equation will satisfy the linearised equation

$$\frac{\partial}{\partial t}\left[\nabla^2\psi + \frac{1}{\bar{\rho}}\frac{\partial}{\partial z}\left(\frac{\bar{\rho}}{S}\frac{\partial\psi}{\partial z}\right)\right] + \beta\frac{\partial\psi}{\partial x} = 0. \tag{4.116}$$

This has solutions

$$\psi = \Psi(z)\exp[i(kx + ly + \omega t)], \tag{4.117}$$

providing

$$\omega = \frac{k\beta}{k^2 + l^2 + m^2}. \tag{4.118}$$

These are Rossby waves. The wave speed $-\omega/k$ is negative, so that the waves move westwards. The sphericity of the Earth (i. e., $\beta > 0$) is essential in causing the waves to move.

4.10 Baroclinic instability

Gravity waves are the sound of the atmosphere. Like a bell which reverberates when struck, gravity waves are excited externally. For example, when the atmosphere flows over mountains, the waves are visualised by the periodic rows of clouds which form in the lee. However, they do not play a prominent part in large scale weather flows, because they are damped fairly rapidly by friction, and they are generated by external effects such as topographic forcing, not by internal dynamics.

Rossby waves, on the other hand, do play an important part in the day to day weather, and this is because they are continually generated by an instability in the underlying basic zonal flow. This instability is called *baroclinic instability*, and it is responsible for the basic wave-like nature of the circulation in mid-latitudes.

We consider the stability of a basic state which is taken to be a purely zonal flow. Because the quasi-geostrophic model is essentially inviscid (and conductionless), there is no unique such state. In the absence of the heating term H on the right hand side of (4.101), *any* zonal stream function $\psi(y, z)$ satisfies the QG equation (4.101). However, we would expect that over sufficiently long time scales, the potential temperature Θ of a zonal flow would become equal to the underlying surface temperature $\Theta_0(y)$, which ultimately is what drives the flow. A local expansion on the mid-latitude length scale

of the global $O(\varepsilon)$ variation in θ suggests the prescription of $\Theta_0 = -y$ at $z = 0$. The choice $\Theta = -y$ implies the zonal flow

$$\psi = k - yz; \tag{4.119}$$

generally, $k = k(z)$ but we will take it as constant, $k = 1$. We will use (4.119) as the basic state whose stability we wish to study.

4.11 The Eady model

The simplest model in which baroclinic instability is manifested is the Eady model. In this model, the tropopause is considered to be a rigid lid, so that we impose

$$W = 0 \quad \text{at} \quad z = 1. \tag{4.120}$$

Basal friction is ignored, corresponding to $E \to 0$ in (4.72), so that

$$W = 0 \quad \text{at} \quad z = 0. \tag{4.121}$$

The Earth's sphericity is ignored by putting $\beta = 0$ in (4.101), the heating term $H = 0$ (consistent with the basic state (4.119), and both the density $\bar{\rho}$ and the stratification S are taken as constant. The equation to be solved is thus the QG equation in the form

$$\frac{D}{Dt}\left[\nabla^2\psi + \frac{1}{S}\frac{\partial^2\psi}{\partial z^2}\right] = 0, \tag{4.122}$$

with boundary conditions which derive from (4.88):

$$\frac{D}{Dt}\left(\frac{\partial\psi}{\partial z}\right) = 0 \quad \text{at} \quad z = 0, 1, \tag{4.123}$$

together with the no flow conditions $\partial\psi/\partial x = 0$ on $y = \pm 1$. In addition, (4.82) implies that

$$\frac{D}{Dt}\int_0^1 \zeta \, dz = 0. \tag{4.124}$$

This is automatically satisfied when ψ satisfies (4.122) and (4.123).

We write

$$\psi = 1 - yz + \Psi, \tag{4.125}$$

and linearise for small Ψ to find

$$\left(\frac{\partial}{\partial t} + z\frac{\partial}{\partial x}\right)\left[\nabla^2\Psi + \frac{1}{S}\frac{\partial^2\Psi}{\partial z^2}\right] = 0, \tag{4.126}$$

subject to

$$\begin{aligned}\left(\frac{\partial}{\partial t} + z\frac{\partial}{\partial x}\right)\frac{\partial\Psi}{\partial z} - \frac{\partial\Psi}{\partial x} &= 0 \quad \text{on} \quad z = 0, 1, \\ \Psi &= 0 \quad \text{on} \quad y = \pm 1.\end{aligned} \tag{4.127}$$

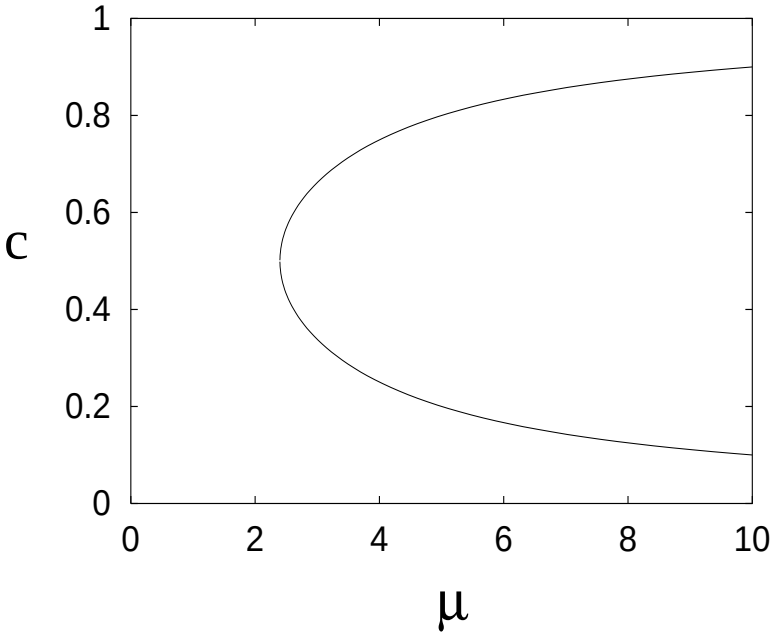

Figure 4.3: Wave speed of perturbations in the Eady model. Instability occurs where the wave speeds are complex conjugates, for $\mu \lesssim 2.4$.

We seek solutions as linear combinations of the form

$$\Psi = A(z)e^{\sigma t + ikx + il_n y}, \tag{4.128}$$

where $l_n = n\pi/2$, and n is an integer. The appropriate linear combination of the y-dependent part is $\sin l_n y$ for n even, and $\cos l_n y$ for n odd. Then

$$(ikz + \sigma)\left[A'' - \mu^2 A\right] = 0, \tag{4.129}$$

where

$$\mu^2 = (k^2 + l^2)S, \tag{4.130}$$

and

$$(ikz + \sigma)A' - ikA = 0 \quad \text{on} \quad z = 0, 1. \tag{4.131}$$

Smooth solutions of (4.129) are linear combinations of $\cosh \mu z$ and $\sinh \mu z$, and the dispersion relation which results from satisfaction of the boundary conditions in (4.129) is

$$c = -\frac{\sigma}{ik} = \frac{1}{2} \pm \frac{1}{\mu}\left[\left(\frac{\mu}{2} - \coth\frac{\mu}{2}\right)\left(\frac{\mu}{2} - \tanh\frac{\mu}{2}\right)\right]^{1/2}, \tag{4.132}$$

where c is the wave speed. Figure 4.3 shows the (real) value of c as a function of (positive) μ. Since $\mu/2 > \tanh(\mu/2)$, it is clear that c is complex for $\mu < \mu_c$, where

$$\frac{\mu_c}{2} = \coth\frac{\mu_c}{2}, \quad \mu_c \approx 2.399. \tag{4.133}$$

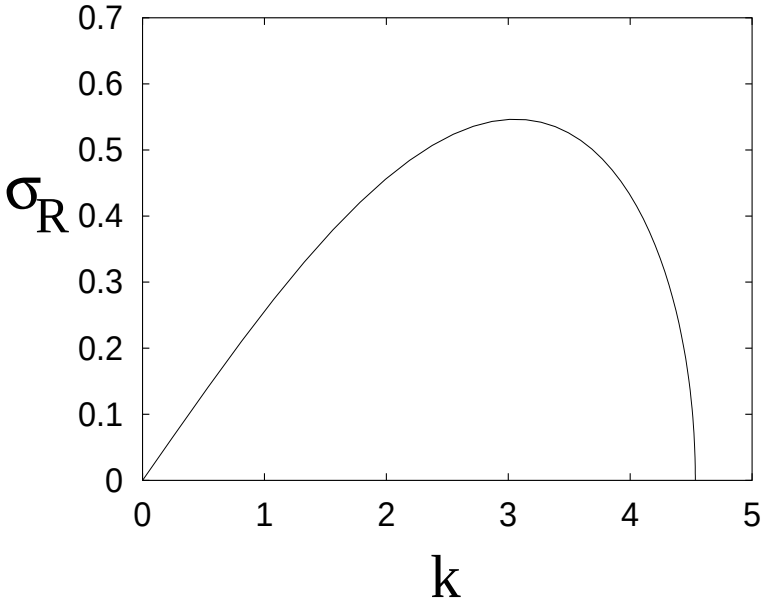

Figure 4.4: Growth rate σ_R of perturbations in the Eady model as a function of wavenumber k when the stratification $S = 0.25$. The growth rate is well approximated by $\sigma_R \approx 0.145k(k_c - k)^{1/2}$, where $k_c = \sqrt{\dfrac{\mu_c^2}{S} - \dfrac{\pi^2}{4}}$ is the maximum wavenumber for instability.

Complex conjugate values of c indicate instability, and this occurs for $\mu < \mu_c$. Instability occurs if $k^2 + l^2 < \mu_c^2/S$, and thus is effected by the minimum values $k = 0$, $l = \pi/2$, and the Eady instability criterion is

$$S < \frac{4\mu_c^2}{\pi^2} \approx 2.218; \tag{4.134}$$

this is readily satisfied in the Earth's atmosphere.

Evidently, the waves (stable or unstable) move to the east in the northern hemisphere, as is observed. The wave speed of unstable waves is 0.5, and the growth rate is

$$\sigma_R = \frac{k}{\mu}\left[\left(\coth\frac{\mu}{2} - \frac{\mu}{2}\right)\left(\frac{\mu}{2} - \tanh\frac{\mu}{2}\right)\right]^{1/2}. \tag{4.135}$$

The growth rate goes to zero as $k \to 0$, and also as $\mu \to \mu_c$. Since, for the fundamental mode ($n = 1$) $\mu^2 = \left(k^2 + \dfrac{\pi^2}{2}\right) S$ increases with k, it appears that the growth rate is maximum for an intermediate value of k. Indeed, figure 4.4 shows a typical graph of the growth rate plotted as a function of wave number k. Although linear stability gives us no information about the eventual form of the growing waves, it is plausible that the maximum growth rate at wavenumber k_m selects the preferred wavelength

of disturbances as $2\pi/k_m$. This appears to be consistent with actual synoptic scale waves in mid-latitudes.

4.12 Frontogenesis

What has all this to do with the weather? If we look at a weather map, or listen to a weather forecaster on a mid-latitude television station, we will hear about fronts and depressions, low pressure systems, cyclones and anti-cyclones. These are indeed the standard bearers of the atmosphere, bringing their associated good and bad weather, storms, rainfall and snow. We are now in a position at least to describe how these features occur.

The weather is described, at least in essence, by some form of the geostrophic or quasi-geostrophic equations. Dissipative effects due to eddy viscosity and eddy thermal conductivity have a short term (days) effect in the planetary boundary layer within a kilometre or so of the surface, but only control the mean temperature of the troposphere over much longer time scales. As a consequence, weather is effectively described by a conservative system, indeed certain approximate models can be written as a Hamiltonian system, and as a consequence it is subject to the same sort of large amplitude fluctuations as those which characterise instability in such systems.

The basic poleward gradient of surface temperature attempts to drive a zonal flow, which is linearly unstable in the presence of a sufficiently small stratification parameter S. The very simplest representation of this instability is found in the Eady model (4.122) and (4.123), which is a nonlinear hyperbolic equation for the potential vorticity. The consequence of the instability is that the steady, parallel characteristics of the zonal flow are distorted and intersect, forming a shock, as illustrated in figure 4.5. This is a front. It consists of a tongue of cold air intruded under warmer air, and the width of the front is typically of order 100 km.

As the front develops, the baroclinic instability also distorts the flow in a wave like pattern. The effect of this is to bend the front round, as illustrated in figure 4.6, forming a series of vortex-like rings. In the atmosphere, these are the cyclonic disturbances which form the mid-latitude low pressure storm systems, with typical dimensions of 2000 km. They also occur in the ocean, forming coherent rings of some 50 km diameter.

The description above is a little idealistic. On the Earth, fronts are an intrinsic consequence of the difference in properties between different air masses. The mid-latitude cells, for example, are bounded north and south by fronts across which the wind direction and the temperature changes. The warm mid-latitude westerlies are bounded polewards by the cold polar easterlies. The situation is complicated by continents and oceans. Continental air is dry, whereas oceanic air is moist. As a consequence of these geographic variations there are a number of different types of air masses, and the boundaries between these provide the seeds for frontal development. The fronts move and distort as shown in figure 4.6, but it is more sensible to think of the roll-up of a planar front and the formation of storm systems as a result of (Kelvin-

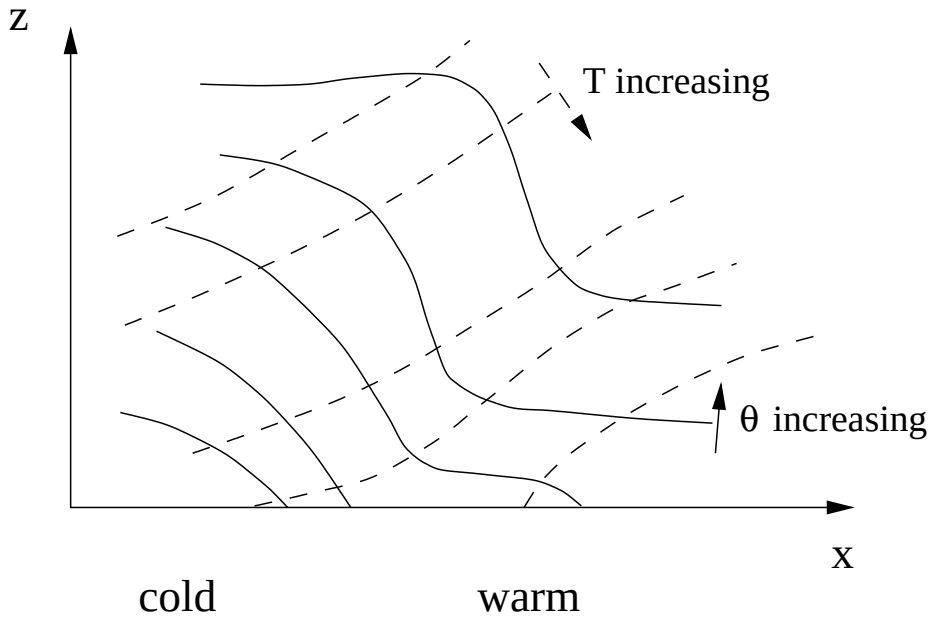

Figure 4.5: Contours of temperature (dashed lines) and potential temperature (solid lines) in a forming front.

Helmholtz like) instability of a linear vortex sheet, rather than as a consequence of shock formation in the nonlinear wave evolution of the quasi-geostrophic potential vorticity (QGPV) equation. In fact, the QGPV does not do a very good job of numerical weather front prediction.[11]

4.13 Depressions and hurricanes

The storm systems which develop as shown in figure 4.6 are called cyclones. They are like vortices which rotate anti-clockwise, and are associated with low pressure at their centres (thus they are also called depressions). Conversely, a high pressure vortex rotating clockwise is called an anti-cyclone. A severe storm with central pressure of 960 millibars represents a dimensionless amplitude of $0.04 \sim \varepsilon^2$, and is thus within the remit of the quasi-geostrophic scaling.

In the tropics, tropical cyclones occur, and the most severe of these is the hurricane, or typhoon. In essence, the hurricane is very similar to the mid-latitude depression, consisting of an anti-clockwise rotating vortex, with wind convergence at the surface, and divergence at the tropopause. It is, however, fuelled by convection, and can be thought of as the result of a strong convective plume interacting with the Coriolis force, which causes the rotation, and in fact organises it into a spiral wave structure, as can be seen in satellite images by the spiral cloud formations.

[11]This comment is due to Peter Lynch.

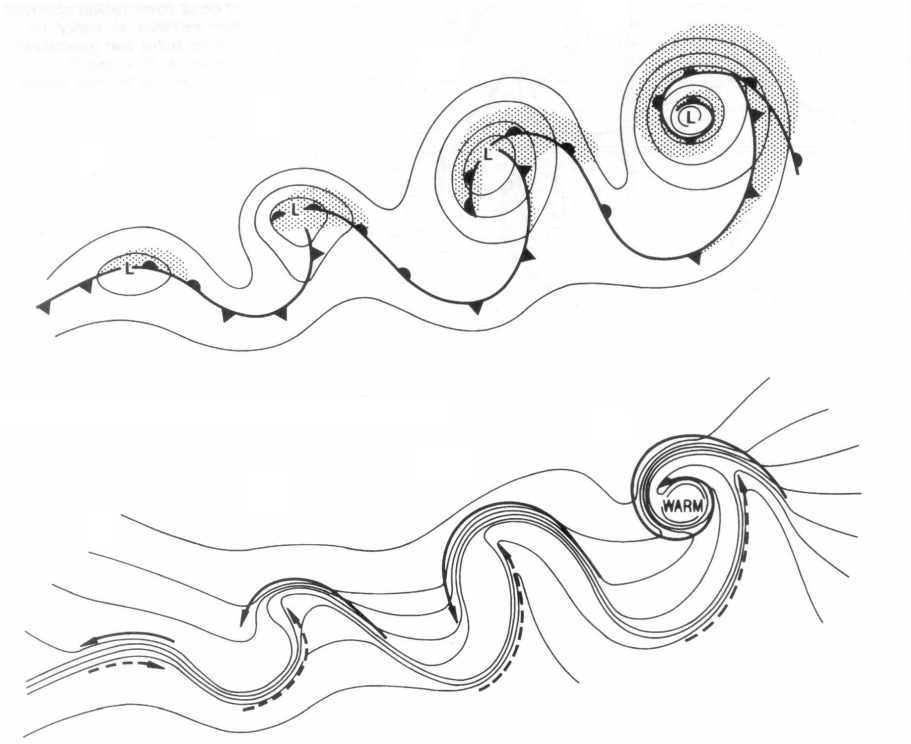

Figure 4.6: Two views of the formation of cyclonic depressions from a baroclinically unstable front. The illustration resembles the Kármán vortex street which forms at moderate Reynolds number in the flow past a cylinder. The upper diagram shows isobars, the front, and cloud cover (stippled); the lower diagram shows isotherms, and flow of cold air (solid arrows) and warm air (dashed arrows). From Barry and Chorley (1998), page 162.

The hurricane is distinguished by its high winds, high rainfall and relatively small size (hundreds rather than the thousands of kilometres of a mid-latiude cyclone). The strongest hurricane on record was hurricane Gilbert in 1988, where the central pressure fell to 888 mbar, and maximum windspeeds were in excess of 55 m s^{-1} (200 km hr^{-1}). The strong convection is a consequence of evaporation from a warm ocean, and it is generally thought that hurricane formation requires a sea surface temperature above 27° centigrade, or 300 K. Relative to a mean surface temperature of 288 K, this is an amplitude of 12 K, and dimensionlessly $12/288 \approx 0.04$, of $O(\varepsilon^2)$. In the tropics, the Rossby number ε is higher, and near the equator the quasi-geostrophic approximation breaks down, but hurricanes do not form in a band near the equator.

Hurricanes typically move westwards in the prevailing tropospheric winds, and dissipate as they move over land, where the fuelling warm oceanic water is not present, and surface friction is greater. They develop a central eye, which is relatively calm and cloud free, and in which air flow is downwards. In hurricanes, this eye is warm.

Notes and references

Exercises

4.1 Using values $d = 10$ km, $\kappa\rho d = 0.67$, $\sigma = 5.67 \times 10^{-8}$ W m^{-2} K^{-4}, $T = 290$ K, show that a representative value of the radiative conductivity k_R defined by

$$k_R = \frac{16\sigma T^3}{3\kappa\rho}$$

for an opaque atmosphere is $k_R \simeq 10^5$ W m^{-1} K^{-1}. Hence show that a typical value for the effective Péclet number

$$Pe = \frac{\rho c_p U d^2}{k_R l}$$

is about 20, if $\rho \approx 1$ kg m^{-3}, $c_p \approx 10^3$ J kg^{-1} K$^{-1}U \approx 20$ m s^{-1}, $l \approx 1000$ km. Explain the import of this in terms of the heat equation

$$\rho c_p \frac{dT}{dt} = \boldsymbol{\nabla}.[k_R \boldsymbol{\nabla} T].$$

4.2 Derive a reference state for a dry atmosphere (no condensation) by using the equation of state

$$p = \frac{\rho R T}{M_a},$$

the hydrostatic pressure

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

and the dry adiabatic temperature equation

$$\rho c_p \frac{dT}{dt} - \frac{dp}{dt} = 0.$$

Show that

$$\bar{T} = T_0 - \frac{gz}{c_p}, \quad \bar{p} = p_0 p^*(z),$$

where

$$p^*(z) = \left(1 - \frac{gz}{c_p T_0}\right)^{M_a c_p/R}.$$

Use the typical values $c_p T_0/g \approx 29$ km, $M_a c_p/R \approx 3.4$, to show that the pressure can be adequately represented by

$$\bar{p} = p_0 \exp(-z/H),$$

where here the scale height is defined as

$$H = \frac{RT_0}{M_a g} \approx 8.4 \text{ km}.$$

(A slightly better numerical approximation near the tropopause is obtained if the scale height is chosen as 7 km.)

4.3 Suppose that θ satisfies the equation

$$\frac{D\theta}{Dt} + \varepsilon W \frac{\partial \theta}{\partial z} = \varepsilon^2 \Gamma W + \varepsilon^2 H, \qquad (*)$$

where Γ and H are constants, $W = W(x, y)$ and the horizontal material derivative is given by

$$\frac{D}{Dt} = \frac{\partial}{\partial t} - \frac{\partial \psi}{\partial y}\frac{\partial}{\partial x} + \frac{\partial \psi}{\partial x}\frac{\partial}{\partial y},$$

where ψ is the geostrophic stream function.

The equation is to be solved in the region V: $-L < x < L$, $-1 < y < 1$, $0 < z < 1$, with the boundary condition $\theta = 1 + \varepsilon^2 \Theta_0(y)$ on $z = 0$, and an initial condition for θ. We can assume without loss of generality that the average of Θ_0 over y is zero. (Why?) Assume that $\psi = \pm 1$ on $y = \pm 1$, and that it is periodic in x (with period $2L$). Comment on the suitability of the initial and boundary conditions. Does it matter whether W is positive or negative??

If A is any horizontal section of V, show that

$$\int_A \frac{D\theta}{Dt}\, dS = \frac{\partial}{\partial t} \int_A \theta \, dS \,,$$

and deduce that the equation

$$\frac{D\theta}{Dt} = g$$

only has a bounded solution if $\bar{g}(z) = 0$, where $\bar{g}$ is the time average of $\displaystyle\int_A g \, dS$.

By expanding θ as $\theta_0 + \varepsilon\theta_1 + \varepsilon^2\theta_2 + \ldots$ and assuming that the solution remains regular, find the equations satisfied by θ_i, $i = 1, 2, 3$, and show that a solution exists in which $\theta_0 = \theta_0(z)$; whence also

$$\theta_0 = 1$$

and $\theta_1 = \theta_1(z)$, and θ_1 is given by

$$\theta_1 = \left(\Gamma + \frac{H}{\overline{\overline{W}}}\right) z;$$

whence

$$\frac{D\theta_2}{Dt} = H\left(1 - \frac{W}{\overline{\overline{W}}}\right). \qquad (\dagger)$$

Suppose now that $\theta_2 = \dfrac{\partial \psi}{\partial z}$; show that $\dfrac{D}{Dt}\left[\dfrac{\partial \theta_2}{\partial z}\right] = \dfrac{\partial}{\partial z}\left(\dfrac{D\theta_2}{Dt}\right)$, and deduce that a solution for θ_2 can be found in the form $\theta_2 = \bar{\theta}_2(z) + \Theta(x, y)$, where $\Theta(x, y)$ is a particular solution of $(\dagger)$, and show that the secularity constraint at $O(\varepsilon^3)$ implies that we can take $\bar{\theta}_2 = 0$. Deduce that $\psi = z\Theta(x, y)$.

Suppose now that a diffusion term $\varepsilon^2 \dfrac{\partial^2 \theta}{\partial z^2}$ is added to the right hand side of $(*)$. Show that the preceding discussion still applies, but now Θ represents an outer solution for θ_2 away from the boundary $z = 0$. By writing $\theta_2 = \Theta + \chi$ and $z = \varepsilon Z$, show that χ satisfies the approximate boundary layer equation

$$\frac{D\chi}{Dt} + W\frac{\partial \chi}{\partial Z} = \frac{\partial^2 \chi}{\partial Z^2},$$

with boundary conditions

$$\begin{aligned} \chi &\to 0 \quad \text{as} \quad Z \to \infty, \\ \chi &= \chi_0(x, y) = \Theta_0 - \Theta \quad \text{on} \quad Z = 0. \end{aligned}$$

For the particular case of a steady zonal flow in which $\dfrac{D}{Dt} = u\dfrac{\partial}{\partial x}$, $u = u(y)$, $W = W(y)$ and $\chi_0 = \displaystyle\sum_k \hat{\chi}_k(y) e^{ikx}$, show that

$$\chi = \sum_k \hat{\chi}_k(y) e^{ikx - \alpha Z},$$

where

$$\alpha = \left(\frac{W^2}{4} + iku\right)^{1/2} - \frac{W}{2}. \qquad (\ddagger)$$

By writing $\dfrac{W^2}{4} + iku = (p + iq)^2$, $p > 0$, and defining the square root in $(\ddagger)$ as having $p > 0$, show that $\operatorname{Re}\alpha > 0$ irrespective of the sign of W. How would you expect Θ to behave over long time scales in this case?

Chapter 5

Two-phase flows

Two-phase flow occurs in numerous situations in industry, as well as in nature. Two-phase flow refers to the coexistence of two phases of a substance in a flow. Typically (though not always) the phases are liquid and gas, as for example in the common occurrence of steam-water flows.

In boilers, water is heated as it flows through a bank of channels, until it starts to boil. This leads to a two-phase flow region until *dryout* occurs, and the flow is of superheated steam. The boundaries which physically divide the various régimes are called the *boiling boundary* and the *superheat boundary*.

A similar situation occurs in nuclear reactors where liquid sodium is commonly used as a coolant. Here it is important that dryout does not occur, since the insulating properties of vapour reduce the cooling efficiency of the flow. Two-phase flow also occurs in condensers, where superheated steam is cooled through the reverse sequence of two-phase and then sub-cooled regions.

Natural examples of two-phase flows include volcanic eruptions, where a variety of such flows can occur, for example ash flows (solid/liquid), and vesicular eruptions, where dissolved gases are exsolved as the magma rises (and loses pressure), so that the erupting flow is of a gas/liquid mixture.

5.1 Flow régimes

Modelling two-phase flow is complicated by a variety of factors. For a start, the flow is usually turbulent, so that some sort of averaging is necessary to model the mean flow. In addition, the distribution of phases means that averaging must also be done so that average variables such as void fraction can be defined. (This is analogous to the definition of variables such as porosity in permeable media.)

A further complication is that two-phase flows can exist in a variety of régimes, all of which will generally occur in a boiling flow. When boiling commences, small bubbles are nucleated at the wall, detach and are taken up by the fluid. Initially, the liquid away from the walls may still be *sub-cooled* (below boiling point), so that heat transfer to the vapour is predominantly at the wall. When the liquid reaches

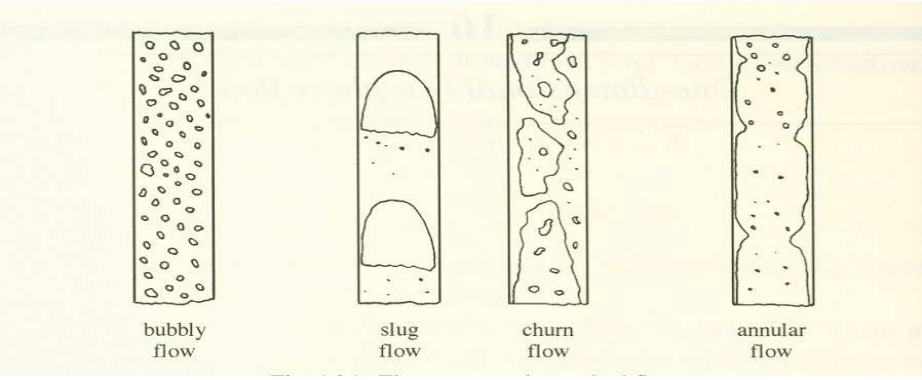

Figure 5.1: Flow patterns in vertical flow

saturation (and in fact becomes slightly superheated), then this régime of *bubbly flow* evolves, by virtue of bubble coalescence and evaporation at bubble interfaces, to the régime known as *slug flow*, in which plugs of gas filling the tube alternate with slugs of bubbly fluid. As the evaporation proceeds, the gas plugs become irregular, and one gets *churn flow*, which leads finally to *annular flow*, in which the liquid is confined to a film at the tube wall, and the gas flows in the core. Shearing between the gas and the liquid causes droplets to be eroded and entrained in the gas. The sequence of flows is portrayed in figure 5.1. Various experimentally based laws to determine parametric criteria for which type of régime a particular flow will adopt lead to the construction of *flow régime maps*, an example of which is shown in figure 5.2.

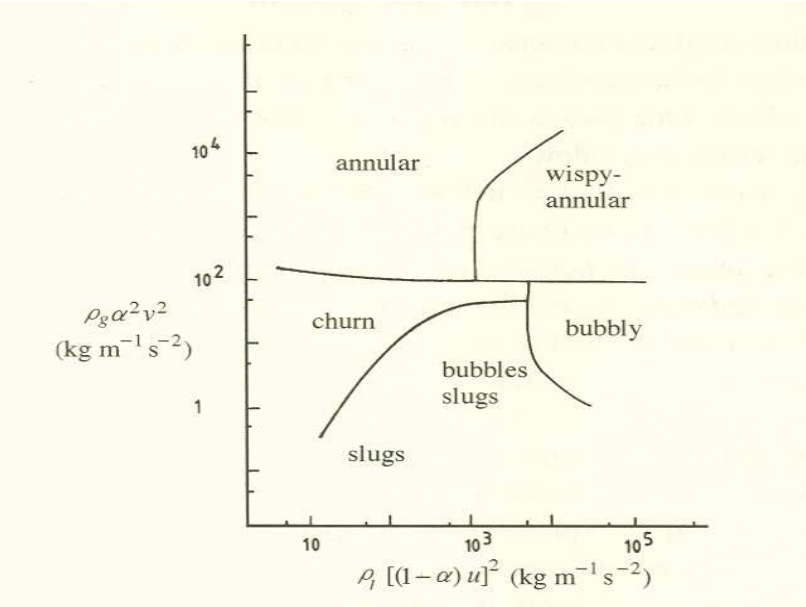

Figure 5.2: Flow régime map

5.2 A simple two-fluid model

Two-phase flow equations are averaged in various ways: in time, cross-sectionally, in space. For one-dimensional flow in a tube, we seek relations for cross-sectionally and time averaged variables representing the two fluids. These variables are the void fraction α, which is the gas volume fraction; u and v, which are the liquid and gas velocities; and averaged pressures p_l and p_g for each phase. We concentrate on steam-water or air-water flow, for example, for which the viscous stresses are manifested through the wall-friction, while the internal friction is largely due to *Reynolds stresses*; both of these terms must be constituted. In writing the simplest model (in order to examine its structure), we will in fact omit frictional terms for the moment.

Ignoring surface tension, it seems reasonable to take $p_g = p_l = p$, and then equations conserving mass and momentum of each phase (and without change of phase due to boiling or condensation) are

$$\begin{aligned}(\alpha\rho_g)_t + (\alpha\rho_g v)_z &= 0,\\ \{\rho_l(1-\alpha)\}_t + \{\rho_l(1-\alpha)u\}_z &= 0,\\ \rho_g[v_t + vv_z] &= -p_z,\\ \rho_l[u_t + uu_z] &= -p_z.\end{aligned} \tag{5.1}$$

These equations can be derived from first principles in the usual way. They represent four equations for the variables α, u, v, p, if we suppose ρ_g and ρ_l, the gas and liquid densities, are given by appropriate equations of state. These are simple generalisations of Euler's equations to the case of two-fluid motion.

For heated flows, one requires also two energy or (more usually) enthalpy equations for the gas and liquid enthalpies h_g and h_l. For *adiabatic* (unheated) flow, these equations are redundant.

Boundary conditions

From a physical consideration of the system, it seems we could prescribe inlet and outlet pressures, and the two inlet mass fluxes. The natural boundary conditions are those of α, u, v at the inlet and p at the outlet (say). Thus if we solve the system for given α_0 at $z = 0$, as well as u and v, then we will obtain the pressure drop Δp as a functional of α_0, $\Delta p = \Delta p(\alpha_0)$. Inversion of this relation determines the necessary α_0 to obtain the correct pressure drop.

5.3 Other models

There are two other commonly used two phase flow models. In the homogeneous model, both phases are assumed to move at the same velocity, thus $u = v$ and one considers momentum conservation for the mixture, in the (simplest) form

$$\rho[u_t + uu_z] = -p_z, \tag{5.2}$$

where

$$\rho = \alpha\rho_g + (1-\alpha)\rho_l \tag{5.3}$$

is the mixture density. The homogeneous model should work best for régimes such as bubbly flow, where there may be little relative motion between the phases.

The *drift-flux model* allows a relative motion, but rather than have separate momentum equations, it considers total momentum conservation in the form

$$[\alpha\rho_g v + (1-\alpha)\rho_l u]_t + [\alpha\rho_g v^2 + (1-\alpha)\rho_l u^2]_z = -p_z, \tag{5.4}$$

with v being related to u through the drift flux (basically, $v-u$ is constituted as a function of α). This is rather akin to the status of Darcy's law.

5.4 Characteristics

The form of the equations (5.1) suggests that the system should be hyperbolic. It can be written in the form

$$A\boldsymbol{\psi}_t + B\boldsymbol{\psi}_z = 0, \tag{5.5}$$

where $\boldsymbol{\psi} = (\alpha, u, v, p)^T$ and, if ρ_g and ρ_l are constant,

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & \rho_l & 0 & 0 \\ 0 & 0 & \rho_g & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} v & 0 & \alpha & 0 \\ -u & 1-\alpha & 0 & 0 \\ 0 & \rho_l u & 0 & 1 \\ 0 & 0 & \rho_g v & 1 \end{pmatrix}. \tag{5.6}$$

We seek characteristics $dz/dt = \lambda$ satisfying $\det(\lambda A - B) = 0$; we find that the eigenvalues λ must satisfy

$$\rho_g(1-\alpha)(\lambda - v)^2 + \rho_l\alpha(\lambda - u)^2 = 0, \tag{5.7}$$

whence

$$\lambda = \frac{u \pm isv}{1 \pm is}, \quad s = \left[\frac{\rho_g(1-\alpha)}{\rho_l\alpha}\right]^{1/2}. \tag{5.8}$$

It follows that there are two *complex* characteristics unless $u = v$, (the other two are infinite, corresponding to two infinite sound speeds). Consequently, the model is *ill-posed* as it stands. There is thus a fundamental logical inconsistency, and before we go on to consider more complicated models, it is worth pursuing this further. Notice that the ellipticity of the model is not due to the neglect of frictional terms: these are algebraic, and do not affect the characteristics.

To see the practical effect of complex characteristics, consider a uniform state $\boldsymbol{\psi} = \boldsymbol{\psi}_0$, subject to small perturbations proportional to $\exp(\sigma t + ikz)$. Such solutions exist if $\sigma = ik\lambda = \mp k\lambda_I + ik\lambda_R$, where $\lambda = \lambda_R \pm i\lambda_I$ represents (5.8). Thus if $\lambda_I \neq 0$, there are unstable solutions, moreover these grow arbitrarily fast at very short wavelengths. These grid scale instabilities are a practical sign of an ill-posed problem.

5.5 More on averaging

In order to resolve this dilemma, let us examine the process of averaging in greater detail. There are many different ways of approaching averaging, and here we follow that outlined by Drew and Wood (1985), or Drew (1983). The idea is to use indicator functions X_k for each phase (labelled by k) such that $X_k(\mathbf{x}, t) = 1$ if $\mathbf{x}$ is in phase k at time t, and $X_k = 0$ otherwise. Averaged equations can then be obtained by multiplying the pointwise conservation laws for phase k by X_k, and averaging. In so doing, we consider X_k as a generalised function, which allows us to use integration by parts.

A typical conservation law has the form

$$\frac{\partial}{\partial t}(\rho\psi) + \boldsymbol{\nabla}.(\rho\psi\mathbf{v}) = -\boldsymbol{\nabla}.\mathbf{J} + \rho f. \tag{5.9}$$

Multiplying by X_k and averaging (in practice, this is often a time and/or space average), yields (an overbar denoting the average)

$$\begin{aligned}\frac{\partial}{\partial t}(\overline{X_k\rho\psi}) + \boldsymbol{\nabla}.[\overline{X_k\rho\psi\mathbf{v}}] &= -\boldsymbol{\nabla}.[\overline{X_k\mathbf{J}}] + \overline{X_k\rho f} \\ &+\overline{\rho\psi\left\{\frac{\partial X_k}{\partial t} + \mathbf{v}_i.\boldsymbol{\nabla}X_k\right\}} + \overline{\{\rho\psi(\mathbf{v}-\mathbf{v}_i) + \mathbf{J}\}.\boldsymbol{\nabla}X_k},\end{aligned} \tag{5.10}$$

where we assume that $\overline{\boldsymbol{\nabla}f} = \boldsymbol{\nabla}\bar{f}$, $\overline{\partial f/\partial t} = \partial\bar{f}/\partial t$, which will be the case for sufficiently well-behaved f. In (5.10), $\mathbf{v}_i$ is the average interfacial velocity of the boundary of phase k, and derivatives of X_k are interpreted as generalised functions. In particular, $\partial X_k/\partial t + \mathbf{v}_i.\boldsymbol{\nabla}X_k = 0$, since if ϕ is a smooth test function, then

$$\begin{aligned}
&\int\int \phi \left[\frac{\partial X_k}{\partial t} + \mathbf{v}_i.\boldsymbol{\nabla} X_k\right] dV dt \\
&\quad = -\int\int X_k \left[\frac{\partial \phi}{\partial t} + \mathbf{v}_i.\boldsymbol{\nabla}\phi\right] dV dt \\
&\quad - \int_{-\infty}^{\infty}\int_{V_k(t)} \left[\frac{\partial \phi}{\partial t} + \mathbf{v}_i.\boldsymbol{\nabla}\phi\right] dV dt \\
&\quad = -\int_{-\infty}^{\infty} \frac{d}{dt}\int_{V_k(t)} \phi \; dV dt \\
&\quad = -\left[\int_{V_k(t)} \phi \; dV\right]_{-\infty}^{\infty} = 0, \qquad (5.11)
\end{aligned}$$

for ϕ vanishing at large values of $|t|$.

The last term in (5.10) is related to the surface average, since $\boldsymbol{\nabla} X_k$ picks out interfacial values. For a smooth test function vanishing at large $\mathbf{x}$, $\mathbf{j}.\boldsymbol{\nabla} X_k$ is defined via

$$\begin{aligned}
\int_V \phi \mathbf{j}.\boldsymbol{\nabla} X_k \; dV &= -\int_V X_k \boldsymbol{\nabla}.(\phi \mathbf{j}) \; dV \\
&= -\int_{V_k} \boldsymbol{\nabla}.(\phi \mathbf{j}) \; dV \\
&= -\int_{S_k} \phi j_n \; dS, \qquad (5.12)
\end{aligned}$$

where j_n is the normal component of $\mathbf{j}$ at the interface, pointing *away* from phase k. This suggests that $\overline{\mathbf{j}.\boldsymbol{\nabla} X_k}$ can be identified with the *surface* average of $-\mathbf{j}.\mathbf{n}$.

Now put $\psi = 1$, $\mathbf{J} = \mathbf{0}$, $f = 0$ in (5.9), corresponding to mass conservation. Then equations of conservation of mass of each phase are, from (5.10),

$$\frac{\partial}{\partial t}(\overline{X_k \rho}) + \boldsymbol{\nabla}.[\overline{X_k \rho \mathbf{v}}] = \overline{\rho(\mathbf{v} - \mathbf{v}_i).\nabla X_k}. \qquad (5.13)$$

The form of (5.13) suggests that we define the average phase volume, density and velocity as follows:

$$\alpha_k = \overline{X_k}, \; \rho_k = \overline{X_k \rho}/\alpha_k, \; \mathbf{v}_k = \overline{X_k \rho \mathbf{v}}/\alpha_k \rho_k, \qquad (5.14)$$

so that (5.13) gives

$$\frac{\partial}{\partial t}(\alpha_k \rho_k) + \boldsymbol{\nabla}.[\alpha_k \rho_k \mathbf{v}_k] = \Gamma_k, \qquad (5.15)$$

where $\Gamma_k = \overline{\rho(\mathbf{v} - \mathbf{v}_i).\boldsymbol{\nabla} X_k}$, and represents a mass source due to phase change (without which $\mathbf{v} = \mathbf{v}_i$ at the interface).

Next, consider momentum conservation. With appropriate interpretation of tensor notation, we put

$$\psi = \mathbf{v},\ \mathbf{J} = p\mathbf{I} - \boldsymbol{\tau},\ f = \mathbf{g}, \tag{5.16}$$

where $\boldsymbol{\tau}$ is the deviatoric stress tensor, $\mathbf{g}$ is gravity. Then

$$\begin{aligned}\frac{\partial}{\partial t}(\overline{X_k\rho\mathbf{v}}) + \boldsymbol{\nabla}.[\overline{X_k\rho\mathbf{v}\mathbf{v}}] = \boldsymbol{\nabla}.[\overline{X_k(-p\mathbf{I}+\boldsymbol{\tau})}] + \overline{X_k\rho\mathbf{g}} \\ +\overline{\{\rho\mathbf{v}(\mathbf{v}-\mathbf{v}_i) + (p\mathbf{I}-\boldsymbol{\tau})\}.\boldsymbol{\nabla}X_k}.\end{aligned} \tag{5.17}$$

Now $\overline{X_k\rho\mathbf{v}} = \alpha_k\rho_k\mathbf{v}_k$, and we would like to have $\overline{X_k\rho\mathbf{v}\mathbf{v}} = \alpha_k\rho_k\mathbf{v}_k\mathbf{v}_k$; but evidently the latter is not the case. Since the flow is normally turbulent, this can be circumvented by separating $\mathbf{v}$ (and, more generally ψ) into mean and fluctuating parts, thus $\mathbf{v} = \mathbf{v}_k + \mathbf{v}'_k$, so that

$$\overline{X_k\rho\mathbf{v}\mathbf{v}} = \alpha_k\rho_k\mathbf{v}_k\mathbf{v}_k + \overline{X_k\rho\mathbf{v}'_k\mathbf{v}'_k}; \tag{5.18}$$

The second term can be interpreted as the averaged Reynolds stress. The momentum equation can thus be written as

$$\begin{aligned}\frac{\partial}{\partial t}(\alpha_k\rho_k\mathbf{v}_k) + \boldsymbol{\nabla}.[\alpha_k\rho_k\mathbf{v}_k\mathbf{v}_k] = \\ \boldsymbol{\nabla}.[\alpha_k(\mathbf{T}_k + \mathbf{T}'_k)] + \alpha_k\rho_k\mathbf{g} + \mathbf{M}_k + \mathbf{v}^m_{ki}\Gamma_k,\end{aligned} \tag{5.19}$$

where

$$\begin{aligned}\alpha_k\mathbf{T}_k &= \overline{X_k(-p\mathbf{I}+\boldsymbol{\tau})}, \\ \alpha_k\mathbf{T}'_k &= \overline{X_k\rho\mathbf{v}'_k\mathbf{v}'_k}, \\ \mathbf{M}_k &= \overline{(p\mathbf{I}-\boldsymbol{\tau}).\boldsymbol{\nabla}X_k}, \\ \mathbf{v}^m_{ki} &= [\overline{\rho\mathbf{v}(\mathbf{v}-\mathbf{v}_i).\boldsymbol{\nabla}X_k}] \Big/ [\overline{\rho(\mathbf{v}-\mathbf{v}_i).\boldsymbol{\nabla}X_k}].\end{aligned} \tag{5.20}$$

Evidently, the average pressure in phase k is $p_k = \overline{X_k p}/\alpha_k$. If we neglect viscous stresses, we can write the interfacial momentum source as

$$\mathbf{M}_k = \overline{p\boldsymbol{\nabla}X_k} = p_{ki}\boldsymbol{\nabla}\alpha_k + \mathbf{M}'_k, \tag{5.21}$$

where

$$\mathbf{M}'_k = \overline{(p-p_{ki})\boldsymbol{\nabla}X_k}, \tag{5.22}$$

p_{ki} is the average interfacial pressure on phase k, and we use $\overline{\boldsymbol{\nabla}X_k} = \boldsymbol{\nabla}\alpha_k$. Thus the momentum equation can be written as

$$\frac{\partial}{\partial t}(\alpha_k\rho_k\mathbf{v}_k) + \boldsymbol{\nabla}.[\alpha_k\rho_k\mathbf{v}_k\mathbf{v}_k] = -\alpha_k\boldsymbol{\nabla}p_k - (p_k - p_{ki})\boldsymbol{\nabla}\alpha_k$$
$$+\boldsymbol{\nabla}.[\alpha_k\mathbf{T}'_k] + \alpha_k\rho_k\mathbf{g} + \mathbf{M}'_k + v^m_{ki}\Gamma_k. \tag{5.23}$$

Commonly, we assume $p_k = p_{ki}$ and the pressure term is explicitly $-\alpha_k\boldsymbol{\nabla}p_k$, an important point to consider.

One dimension: profile coefficients

In one-dimensional turbulent flow, it is common to constitute Reynolds stresses via an eddy viscosity, or equivalently, to define (cf. (5.18))

$$\overline{X_k\rho\mathbf{v}\mathbf{v}} = D_k\alpha_k\rho_k\mathbf{v}\mathbf{v}, \tag{5.24}$$

where D_k is known as a *profile coefficient*. It includes the effects of both the Reynolds stresses and also the cross-sectional non-uniformity of the flow. Let us consider the simplest modification to (5.1) which allows $D_k \neq 1$. A one-dimensional version of (5.1) is, with ρ_g and ρ_l constant,

$$\begin{aligned}
&\alpha_t + (\alpha v)_z = 0,\\
&-\alpha_t + [(1-\alpha)u]_z = 0,\\
&\rho_g(\alpha v)_t + \rho_g(D_g\alpha v^2)_z = -\alpha p_z,\\
&\rho_l[(1-\alpha)u]_t + \rho_l[D_l(1-\alpha)u^2]_z = -(1-\alpha)p_z.
\end{aligned} \tag{5.25}$$

It will usually be appropriate to choose $D_g = 1$, but we allow $D_l \neq 1$. Then the last two of (5.25) can be written as

$$\begin{aligned}
&\rho_g[v_t + vv_z] = -p_z,\\
&\rho_l\left[u_t + (2D_l - 1)uu_z - (D_l - 1)\left[\frac{u^2}{1-\alpha}\right]\alpha_z\right] = -p_z,
\end{aligned} \tag{5.26}$$

and the system can be written as (5.5) for $\boldsymbol{\psi} = (\alpha, u, v, p)^T$, with

$$A = \begin{pmatrix} 1 & 0 & 0 & 0\\ -1 & 0 & 0 & 0\\ 0 & \rho_l & 0 & 0\\ 0 & 0 & \rho_g & 0 \end{pmatrix},\ B = \begin{pmatrix} v & 0 & \alpha & 0\\ -u & (1-\alpha) & 0 & 0\\ -\frac{\rho_l(D_l-1)u^2}{(1-\alpha)} & \rho_l(2D_l-1)u & 0 & 1\\ 0 & 0 & \rho_g v & 1 \end{pmatrix}. \tag{5.27}$$

The characteristics $dz/dt = \lambda$ satisfy $\det(\lambda A - B) = 0$, hence with s defined in (5.8),

$$(\lambda - u)^2 = \delta[u^2 + 2u(\lambda - u)] - s^2(\lambda - v)^2, \tag{5.28}$$

where $\delta = D_l - 1$. For small values of s and δ, we see that λ is real provided

$$\delta > s^2(u-v)^2/u^2, \tag{5.29}$$

so that practically, a very small Reynolds stress (or profile coefficient above one) is sufficient to make the basic system have real characteristics, and hence be well-posed. Other possibilities to render the system well-posed can be chosen. In practice, any realistic model will (and *should*) be well-posed, otherwise it will be physically meaningless.

5.6 A simple model for annular flow

In many two-phase boiling flows, annular flows are significant. Because the gas flow is less impeded by the liquid, the flow velocities are high, and consequently, annular flow regions can occupy large parts of the tube. Moreover, since the gas and liquid velocities are very different, a two-fluid model is appropriate. If we denote α, β as gas and liquid volume fractions, u, v as liquid and gas velocities, and h_g and h_l as gas and liquid enthalpies, then a typical one-dimensional set of model equations is given by

$$\begin{aligned}
(\alpha\rho_g)_t + (\alpha\rho_g v)_z &= \Gamma, \\
(\beta\rho_l)_t + (\beta\rho_l u)_z &= -\Gamma, \\
(\beta\rho_l u)_t + (D\beta\rho_l u^2)_z &= -\beta p_z - F_{lw} + F_{li}, \\
(\alpha\rho_g v)_t + (\alpha\rho_g v^2)_z &= -\alpha p_z + F_{gi}, \\
\beta\rho_l[h_{lt} + uh_{lz}] &= \Gamma(h_l - h_{li}) + E_l + Q/A, \\
\alpha\rho_g[h_{gt} + vh_{gz}] &= \Gamma(h_{gi} - h_g) + E_g, \\
E_l + E_g + \Gamma(h_{gi} - h_{li}) &= 0.
\end{aligned} \tag{5.30}$$

The first two of these represent conservation of mass, as before. The next two represent conservation of momentum, and we have chosen to include the wall friction F_{lw} on the liquid, and the interfacial friction F_{li} on the liquid (and F_{gi} on the gas, $F_{gi} = -F_{li}$). The next two are enthalpy equations. In an annular flow, boiling takes place at the liquid-gas interface, so that the liquid in particular must be superheated. If the average phasic enthalpy h_k is different from the interfacial value h_{ki}, then there is a convective transfer of enthalpy $\Gamma(h_{ki} - h_k)$ to that phase associated with the phase change term. In addition, there will be a diffusive transport E_k due to heat conduction. Finally, Q is the external heat supply per unit length per unit time, and A is the cross-sectional area. The final relation then represents the volume average of the Stefan condition. There are various other terms which could be included but they can be treated in the same way as below, and are in any case often small.

Nondimensionalisation

Our aim is to derive appropriate scaling relationships for the variables and to show how suitable simplifications can often be made. To do this, we use values of the parameters typical to one particular application, that of a steam turbine.

Firstly, we must choose constitutive forms for the various terms which arise through the averaging process. We suppose

$$\begin{aligned} F_{lw} &= \frac{2}{d} f_{lw} \rho_l \left| u \right| u, \\ F_{li} &= -F_{gi} = \frac{2}{d} f_{li} \rho_g \left| v - \chi u \right| (v - \chi u) \end{aligned} \tag{5.31}$$

are the friction terms. The numbers f_{lw}, f_{li} are friction factors, and are themselves usually considered to be functions of the phasic Reynolds numbers. Here we take them to be constant. The coefficient χ in $(16.31)_2$ represents the speed of interfacial waves; a typical value is $\chi = 2$.

We will assume that the interface is in thermodynamic equilibrium, thus

$$h_{gi} = h_g^{sat}, \quad h_{li} = h_l^{sat}, \tag{5.32}$$

where these are the saturation values of the enthalpies, and $L = h_g^{sat} - h_l^{sat}$ is the latent heat. Enthalpies are related to temperature by

$$\begin{aligned} h_g &= h_g^{sat} + c_{pg}(T_g - T^{sat}), \\ h_l &= h_l^{sat} + c_{pl}(T_l - T^{sat}), \end{aligned} \tag{5.33}$$

where T^{sat} is the saturation (boiling) temperature, which we take as constant. In terms of temperature, the interfacial heat transfer terms are

$$\begin{aligned} E_g &= H_{gi}(T_{gi} - T_g)/L_s, \\ E_l &= H_{li}(T_{li} - T_l)/L_s, \end{aligned} \tag{5.34}$$

where H_{li} and H_{gi} are heat transfer coefficients, $T_{gi} = T_{li} = T^{sat}$ here, and L_s^{-1} is the (average) surface area per unit volume, which for annular flow we take as

$$L_s^{-1} = 4\alpha^{1/2}/d \approx 4/d. \tag{5.35}$$

In general, H_{ki} are complicated functions of Reynolds number, etc., but we take them as constant.

Now we scale the variables by writing

$$\begin{aligned} &z = lz^*, \quad u = Uu^*, \quad v = Vv^*, \quad p = p_0 + Pp^*, \quad \beta = B\beta^*, \\ &\Gamma = G\Gamma^*, \quad t = (l/U)t^*, \quad h_g = h_g^{sat} + Lh_g^*, \quad h_l = h_l^{sat} + Lh_l^*; \end{aligned} \tag{5.36}$$

we take l to be the tube length, and choose the unknown scales U, V, P, B, G by effecting the following balances in (5.30):

$$(\alpha\rho_g v)_z \sim \Gamma, \quad (\beta\rho_l u)_z \sim \Gamma, \quad F_{lw} \sim F_{li}, \quad \alpha p_z \sim F_{gi}, \quad \Gamma(h_l - h_{li}) \sim Q/A. \quad (5.37)$$

We take ρ_g and ρ_l as constants for simplicity, though in reality ρ_g will vary by a reasonable amount. Specifically, we choose (since $\alpha \sim 1$)

$$\begin{aligned} \rho_g V = lG, \quad \rho_l BU = lG, \quad & U/V = (f_{li}\rho_g/f_{lw}\rho_l)^{1/2}, \\ P = 2lf_{lw}\rho_l U^2/d, \quad & G = Q/AL \end{aligned} \quad (5.38)$$

(if Q is constant, or a typical value if it varies). Substitution of these values into the equations, and omitting the asterisks, yields the following:

$$\begin{aligned} \alpha &= 1 - c_1\beta, \\ c_2\alpha_t + (\alpha v)_z &= \Gamma, \\ \beta_t + (\beta u)_z &= -\Gamma, \\ c_3[(\beta u)_t + D(\beta u^2)_z] &= -c_1\beta p_z - u^2 + (v - \chi c_2 u)^2, \\ c_4[c_2(\alpha v)_t + (\alpha v^2)_z] &= -\alpha p_z - (v - \chi c_2 u)^2, \\ \beta(h_{lt} + uh_{lz}) &= \Gamma h_l - h_l/c_5 + q, \\ c_2\alpha h_{gt} + \alpha v h_{gz} &= \Gamma h_g - h_g/c_6, \\ \Gamma &= h_g/c_6 + h_l/c_5, \end{aligned} \quad (5.39)$$

where the parameters are defined by

$$\begin{aligned} c_1 = B, \quad c_2 = U/V, \quad & c_3 = Bd/2f_{lw}l, \\ c_4 = d/2f_{li}l, \quad c_5 = Gc_{pl}/H_{li}L_s^{-1}, \quad & c_6 = Gc_{pg}/H_{gi}L_s^{-1}, \end{aligned} \quad (5.40)$$

and $q = O(1)$ is the dimensionless heat supply.

To estimate the values of the parameters, we choose

$$\begin{aligned} l = 10\,\mathrm{m}, \quad \rho_g = 30\,\mathrm{kg\ m^{-3}}, \quad \rho_l = 760\,\mathrm{kg\ m^{-3}}, \quad & f_{lw} = .004, \\ f_{li} = .02, \quad d = .014\,\mathrm{m}, \quad A \sim 1.5\times10^{-4}\,\mathrm{m^2}, \quad & \dot{m} = .2\,\mathrm{kg\ s^{-1}}, \\ L_s^{-1} = 4/d, \quad c_{pl} = 5.2\,\mathrm{kJ\ kg^{-1}\,K^{-1}}, \quad & c_{pg} = 4.6\,\mathrm{kJ\ kg^{-1}\,K^{-1}}, \\ H_{li} \sim 1.7\times10^5\,\mathrm{W\ m^{-2}\,K^{-1}}, \quad & H_{gi} \sim .7\times10^5\,\mathrm{W\ m^{-2}\,K^{-1}}, \end{aligned} \quad (5.41)$$

where $\dot{m}$ is the inlet mass flux; this determines Q via $Ql \sim \dot{m}L$, so that G is equivalently determined from

$$G = \dot{m}/Al. \tag{5.42}$$

These parameters are relevant for steam/water flow at an ambient pressure of 60 bars. We find successively

$$G \sim 130 \text{ kg m}^{-3} \text{ s}^{-1}, \quad V \sim 43 \text{ m s}^{-1}, \quad U \sim 19 \text{ m s}^{-1},$$
$$B \sim .09, \quad P \sim 16 \text{ bars } (1.6 \times 10^6 \text{ Pa}), \tag{5.43}$$

and thus

$$c_1 \sim 0.09,\ c_2 \sim 0.44,\ c_3 \sim 0.016,\ c_4 \sim 0.035,\ c_5 \sim 0.014,\ c_6 \sim 0.03. \tag{5.44}$$

The fact that all the parameters are less than one indicates that the balances chosen in (5.38) are correct. In practice, this often has to be done through trial and error.

Analysis

Since all the parameters except c_2 in (16.44) are in fact small, the system lends itself to an asymptotic reduction. Specifically, $h_g, h_l \ll 1$ (thus the fluids are close to thermodynamic equilibrium); adding the two enthalpy equations then gives

$$\Gamma \approx q. \tag{5.45}$$

Also $\alpha \approx 1$, and the gas momentum equation is just a quadrature for p,

$$0 = -p_z - (v - \chi c_2 u)^2, \tag{5.46}$$

while the liquid momentum equation is simply a force balance:

$$u = \lambda v, \tag{5.47}$$

where

$$\lambda = (1 + \chi c_2)^{-1} \sim 0.5. \tag{5.48}$$

The equations thus reduce to

$$v_z \approx q,$$
$$\beta_t + \lambda(\beta v)_z = -q, \tag{5.49}$$

which can even be solved explicitly. Thus, if q is constant,

$$v = v_0 + qz, \tag{5.50}$$

and using characteristics, we find

$$(1 + \lambda\beta) = \frac{(1 + \lambda\beta_0)v_0}{v_0 + qz}, \tag{5.51}$$

if $v = v_0$, $\beta = \beta_0$ at $z = 0$. Thus β decreases and reaches zero at $z = \lambda\beta_0 v_0/q$, corresponding to dryout of the liquid film.

The dramatic collapse we have illustrated here is robust, in the sense that even if a more realistic system is algebraically more intractable (for example, because f_{lw} depends on u, f_{li} depends on v and β, etc.), nevertheless the approximations involved will continue to apply. The result is that numerical computations can be greatly simplified. Simple but accurate models are also of use in establishing parameter dependence of stability characteristics, for example.

In neglecting the parameters c_1, c_3, c_4, c_5, c_6, we reduce a sixth-order system to a third-order one, two of whose equations are quadratures. All these approximations are potentially singular, as they involve the neglect of small terms, and we must be concerned at whether we also lose the ability to satisfy initial/boundary conditions. Insofar as the system will be hyperbolic, we can afford not to worry about loss of time derivatives, since any resulting rapid transients will be washed out of the system.

Inspecting the reduced system, we see that we can specify v and β at the inlet, but not u, as we lose the acceleration terms in the liquid momentum equation. Loss of acceleration in the gas momentum equation does not matter, since the pressure gradient remains, and the equation acts as a quadrature for p. Inspection of the enthalpy equations shows that any initial/boundary condition for h_l or h_g is quickly relaxed. Since, in fact, it is natural to prescribe $h_l = h_g = 0$ at the entrance to the annular flow region, there will not actually be any enthalpy boundary layer. The only cause for worry is thus the inability to prescribe u, and we must forgo this luxury. Strictly, an inlet boundary layer analysis is then necessary to complete the solution.

Two important features about this two-phase flow are these: the pressure gradient disappears from the liquid momentum equation because $\beta \ll 1$ (in annular flow). Further, acceleration terms will always be small if $\rho u^2 \ll \Delta p$, which is inevitably the case. Thus, it will be a common feature that liquid momentum simply gives a force balance, so that the momentum equations can effectively be taken out of the system.

5.7 Density wave oscillations

Density wave oscillations in two phase flow have been of concern in the nuclear industry for a long time. In a steam generating boiler, water in an array of pipes is heated externally, and begins to boil as it flows along the pipes. In certain situations, the resulting two phase flow can be oscillatorily unstable, which is an undesirable feature in industrial systems. The instability mechanism is the same one that produces chugging in a domestic back boiler, if the pipework to the hot water tank is incorrectly installed (i. e., a section of pipe from fire to hot water tank is inclined downwards), and also the same that produces geysering in geothermal springs. A more direct analogy is in certain types of effusive volcanic eruptions, such as that at Villarrica volcano in Chile, where the magma flow in the vent appears to oscillate between a bubbly flow régime and a slug flow régime.

The simplest model to describe the instability was posed and analysed by Davies

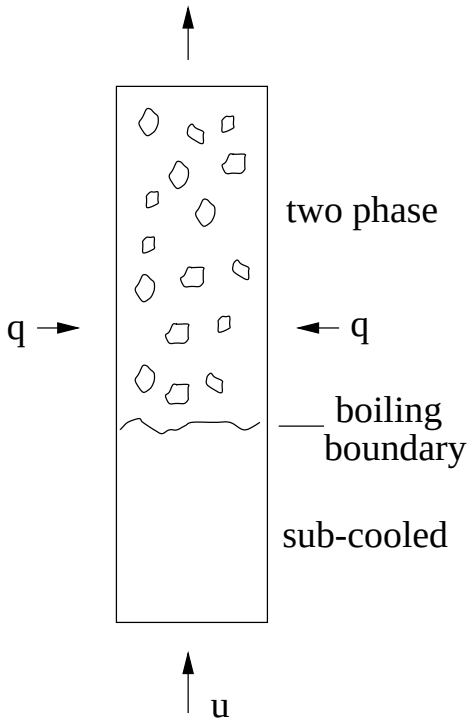

Figure 5.3: Geometry of heated flow in a boiler tube.

and Potter (1967). They studied a homogeneous two phase flow model, in which it is assumed that in the two phase region, the liquid and vapour phases move with the same velocity. This is the simplest assumption, though it is inaccurate, particularly in the annular flow régime. Nevertheless it appears to capture the essence of the instability, and serves as a useful starting point.

The geometry of the flow under consideration is shown in Figure 5.3. We denote the vertical coordinate along the tube as x, so that the tube occupies $0 < x < l$. The flow is supposed turbulent, as is generally the case, and we use a cross-sectionally averaged model of the turbulent flow, with the effects of Reynolds stresses modelled by an empirical wall friction factor. Heat is added to the flow at a rate q, and we suppose that the flow consists of two régimes, a sub-cooled region occupying the initial part of the tube $0 < x < r(t)$, and a two-phase region $r < x < l$, in which liquid and vapour coexist as a bubbly flow. In practice, such two phase flows pass through a succession of flow régimes, from bubbly to slug to churn to annular, as the vapour fraction increases. To model such different régimes at all requires a more sophisticated two fluid model, but it is not known with any theoretical confidence what controls the transition between the different régimes, so that the effort involved may not be appropriate.

We pose a homogeneous model of the two phase region. We suppose that the flow is driven by a pressure drop Δp, and that the inlet temperature and pressure (and thus also enthalpy) is prescribed. Equations describing the flow are those of

conservation of mass, momentum and energy, and they take the form

$$\begin{aligned} \rho_t + (\rho u)_x &= 0, \\ \rho(u_t + u u_x) &= -p_x - \rho g - \frac{4 f \rho u^2}{d}, \\ \rho \frac{dh}{dt} - \frac{dp}{dt} &= \frac{4q}{d}, \end{aligned} \tag{5.52}$$

where ρ is density (of the liquid in the single phase region, of the mixture in the two-phase region), u is velocity, p is pressure, and h is enthalpy; $\dfrac{d}{dt}$ is the material derivative $\dfrac{\partial}{\partial t} + u\dfrac{\partial}{\partial x}$, $f\rho u^2$ is the wall stress, q is the heat input per unit surface area, g is the gravitational acceleration, and d is the tube diameter. These are to be solved subject to the conditions

$$\begin{aligned} h = h_0 \quad \text{at} \quad x = 0, \\ -\int_0^l p_x \, dx &= \Delta p. \end{aligned} \tag{5.53}$$

One simplification which we immediately make is to suppose that

$$\frac{\Delta p}{\rho_g L} \ll 1, \tag{5.54}$$

where ρ_g is vapour density and L is latent heat. This is well satisfied in industrial contexts, and guarantees that the adiabatic pressure derivative term in the enthalpy equation can be ignored, which we henceforth do. For steam–water systems, this is a good approximation. For example, $\rho_g L \approx 270$ bars (1 bar $= 10^5$ Pa $= 10^5$ N m^{-2}) at an operating pressure of 30 bars, and $\rho_g L \approx 13$ bars at atmospheric pressure (where we take vapour density at the boiling point).

Sub-cooled region

In the single phase sub-cooled liquid near the inlet, we suppose the liquid is incompressible, so that $\rho = \rho_l$ is constant, and thus $u = U(t)$, and

$$h_t + U h_x = \frac{4q}{\rho_l d}. \tag{5.55}$$

With the inlet condition $(5.53)_1$, this is easily solved, to find

$$\begin{aligned} x &= \int_\tau^t U(\theta) \, d\theta, \\ h &= \frac{4q}{\rho_l d}(t - \tau) + h_0. \end{aligned} \tag{5.56}$$

We assume here that q is constant. A more realistic assumption is to have q depend on u, and it is possible to treat this case also, though less easily. We define the saturation enthalpy of the liquid at the boiling point to be h_l, and we define the inlet sub-cooling to be

$$\Delta h_0 = h_l - h_0; \tag{5.57}$$

then $h = h_l$ defines the location of the boiling boundary $x = R(t)$, and we find

$$R(t) = \int_{t-\tau_0}^{t} U(\theta)\, d\theta, \tag{5.58}$$

where

$$\tau_0 = \frac{\rho_l d \Delta h_0}{4q}. \tag{5.59}$$

Note that (5.58) introduces a delay τ_0 into the system: the boiling boundary position $R(t)$ depends on the history of the inlet velocity $U(t)$.

Two-phase region

In the two phase liquid–vapour region $x > R$, we define the void fraction α to be the volume fraction of vapour. We then have the definitions of two phase density and enthalpy:

$$\begin{aligned} \rho &= \rho_l(1-\alpha) + \rho_g \alpha, \\ \rho h &= \rho_l h_l (1-\alpha) + \rho_g h_g \alpha, \end{aligned} \tag{5.60}$$

where suffixes l and g indicate liquid and gas properties; note that the latent heat is

$$L = h_g - h_l. \tag{5.61}$$

Eliminating α yields h as a function of ρ, and we substitute this into (5.52) to find that ρ and u in $x > R$ satisfy the equations

$$\begin{aligned} \rho_t + u\rho_x &= -u_x \rho, \\ \frac{\rho_g \rho_l L}{\Delta \rho}\, u_x &= \frac{4q}{d}, \end{aligned} \tag{5.62}$$

where

$$\Delta \rho = \rho_l - \rho_g, \tag{5.63}$$

subject to

$$\rho = \rho_l, \quad u = U(t) \quad \text{on} \quad x = R. \tag{5.64}$$

When this pair of equations is solved, then the inlet velocity is found by requiring that

$$\begin{aligned} \Delta p &= \frac{4f}{d}\left[\rho_l U^2 R + \int_R^l \rho u^2\, dx\right] + g\left[\int_R^l \rho\, dx + \rho_l R\right] \\ &\quad + \rho u^2\big|_l - \rho_l U^2 + \frac{\partial}{\partial t}\int_R^l \rho u\, dx + \rho_l \dot{U} R. \end{aligned} \tag{5.65}$$

Non-dimensionalisation

At this point it is convenient to non-dimensionalise the model. We scale the variables as follows:

$$\rho \sim \rho_l, \quad x, R \sim l, \quad t \sim \tau_0, \quad u, U \sim u_0 = \frac{l}{\tau_0}; \tag{5.66}$$

we then find the dimensionless model to be

$$\begin{aligned} R &= \int_{t-1}^{t} U(\theta)\, d\theta, \\ \Delta p &= \Delta p_f \left[U^2 R + \int_R^1 \rho u^2 \, dx \right] + \Delta p_g \left[\int_R^1 \rho \, dx + R \right] \\ &\quad + \Delta p_i \left[\rho u^2 \big|_1 - U^2 + \frac{\partial}{\partial t} \int_R^1 \rho u \, dx + \dot{U} R \right]. \end{aligned} \tag{5.67}$$

where

$$\Delta p_i = \rho_l u_0^2, \quad \Delta p_f = \frac{4 f l \rho_l u_0^2}{d}, \quad \Delta p_g = \rho_l g l, \tag{5.68}$$

and ρ and u satisfy

$$\begin{aligned} \rho_t + u \rho_x &= -u_x \rho, \\ \varepsilon u_x &= 1 \end{aligned} \tag{5.69}$$

in $R < x < 1$, with

$$u = U, \quad \rho = 1 \quad \text{on} \quad x = R, \tag{5.70}$$

and

$$\varepsilon = \frac{\rho_g L}{\Delta \rho \Delta h_0}. \tag{5.71}$$

For reasons which will emerge, we define the pressure drop scale

$$\Delta p_0 = \frac{\Delta p_f}{\varepsilon} = \frac{4 f l \rho_l u_0^2 \Delta \rho \Delta h_0}{\rho_g L d}, \tag{5.72}$$

and then in terms of the dimensionless parameters

$$\gamma = \frac{\Delta p_g}{\Delta p_f} = \frac{g d}{4 f u_0^2}, \quad \delta = \frac{\Delta p_i}{\Delta p_f} = \frac{d}{4 f l}, \tag{5.73}$$

the dimensionless pressure drop $\Pi = \Delta p / \Delta p_0$ is

$$\begin{aligned} \frac{\Pi}{\varepsilon} &= \left[U^2 R + \int_R^1 \rho u^2 \, dx \right] + \gamma \left[\int_R^1 \rho \, dx + R \right] \\ &\quad + \delta \left[\rho u^2 \big|_1 - U^2 + \frac{\partial}{\partial t} \int_R^1 \rho u \, dx + \dot{U} R \right]. \end{aligned} \tag{5.74}$$

The solution for u is evidently

$$u = U + \frac{x - R}{\varepsilon}. \tag{5.75}$$

The characteristic solution for ρ is then found to be, after some rearrangement and integration by parts,

$$\begin{aligned} x &= R(t) + \varepsilon \int_0^s U_1(t - \varepsilon\zeta)\, e^{\zeta}\, d\zeta, \\ \rho &= e^{-s}, \end{aligned} \tag{5.76}$$

where

$$U_1(\theta) = U(\theta - 1), \tag{5.77}$$

and note that then

$$u = U + \int_0^s U_1(t - \varepsilon\zeta)\, e^{\zeta}\, d\zeta. \tag{5.78}$$

A reduced model

The model thus reduces to the pair of equations (5.74) and $(5.67)_1$, with the various integrals computed using (5.76) and (5.78). Writing them together, these equations are

$$\begin{aligned} R &= \int_{t-1}^{t} U(\theta)\, d\theta, \\ \frac{\Pi}{\varepsilon} &= \left[U^2 R + \int_R^1 \rho u^2\, dx\right] + \gamma \left[\int_R^1 \rho\, dx + R\right] \\ &\qquad + \delta \left[\rho u^2\big|_1 - U^2 + \frac{\partial}{\partial t}\int_R^1 \rho u\, dx + \dot{U} R\right], \end{aligned} \tag{5.79}$$

in which ρ and u are defined by

$$\begin{aligned} \rho &= e^{-s}, \\ u &= U + \int_0^s U_1(t - \varepsilon\zeta)\, e^{\zeta}\, d\zeta, \end{aligned} \tag{5.80}$$

and s is defined through

$$x = R(t) + \varepsilon \int_0^s U_1(t - \varepsilon\zeta)\, e^{\zeta}\, d\zeta. \tag{5.81}$$

The delay (scaled to be one) in this system is explicitly represented by the delayed function $U_1(t) = U(t - 1)$, but a second delay is manifested through the dependence of u and x on the delayed values of U_1.

The model depends on the four parameters Π, ε, γ and δ. It is common to describe the behaviour of the system in terms of two dimensionless parameters known as the sub-cooling number $N_{\rm sub}$ and the phase change number $N_{\rm pch}$. In terms of the parameters defined above, these are defined by

$$N_{\rm sub} = \frac{1}{\varepsilon}, \quad N_{\rm pch} = \frac{1}{\varepsilon U}. \tag{5.82}$$

Stability diagrams are often given in terms of these parameters. (Obviously, assuming that U is constant, i. e., under steady state conditions.)

It is generally the case that $\rho_g \ll \rho_l$, and we will suppose that $\Delta h_0 \sim L$, so that the inlet water is decently sub-cooled. It then follows that $\varepsilon \ll 1$, a typical value at an operating pressure of 30 bars being $\varepsilon \approx 0.02$. As illustrative values, we take $u_0 = 1$ m s^{-1}, $\rho_l = 10^3$ kg m^{-3}, $l = 10$ m, $d = 10^{-2}$ m, $f = 0.01$, $g \sim 10$ m s^{-2}. From these we find $\gamma \approx 2.5$, $\delta \approx 0.025$. Thus it seems reasonable to suppose that $\gamma \sim O(1)$, and $\delta \sim \varepsilon \ll 1$.

We see from (5.76) that if $0 < R < 1$ and is not close either to the inlet or outlet, then $s_1 \sim \ln(1/\varepsilon)$, where $s = s_1$ when $x = 1$. Since we then have that $\varepsilon\zeta \sim \varepsilon \ln(1/\varepsilon) \ll 1$, we can expand the integral terms in (5.74), and at leading order we then find that s_1 is given by

$$s_1 \sim \ln\left[\frac{1-R}{\varepsilon U_1}\right], \tag{5.83}$$

and the dimensionless pressure drop Π can be written as

$$\begin{aligned}\Pi \;\approx\; & \tfrac{1}{12}U_1(1-R)^2 + \varepsilon U^2 R + \varepsilon\gamma\left[\varepsilon U_1 \ln\left(\frac{1-R}{\varepsilon U_1}\right) + R\right] \\ & + \delta\left[U_1(1-R) - \varepsilon U^2\right] + \varepsilon\delta\left[\frac{d}{dt}\left\{U_1(1-R)\right\} + \dot{U}R\right].\end{aligned} \tag{5.84}$$

Steady states

Figure 5.4 shows steady states of U as a function of applied pressure drop. Evidently multiple steady states occur. The middle branch is unstable, this instability being known as Ledinegg instability. In the steady state $R = U$, so that the Ledinegg values $0.44 \lesssim U \lesssim 0.86$ (in Figure 5.4) correspond to situations where the boiling boundary approaches the outlet.

Instability and ill-posedness

Oscillatory instabilities can occur as well as the direct Ledinegg instability. To study these we begin by putting $\varepsilon = 0$ (now we can see why the factor ε was included in the denominator of the left hand side of (5.84)). At leading order, the model reduces to

$$\Pi \approx \tfrac{1}{2}U_1(1-R)^2, \quad R = \int_{t-1}^{t} U(\theta)\,d\theta. \tag{5.85}$$

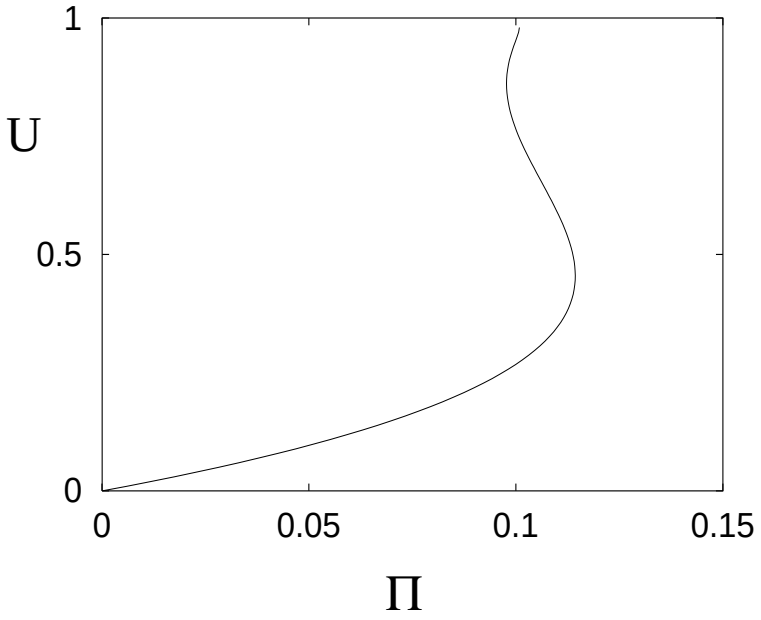

Figure 5.4: Multiple steady states of (5.84) when $\varepsilon = 0.03$, $\gamma = 2.5$, $\delta = 0.025$.

This innocuous-looking delay integral equation is ill-posed. (it also appears not to be accurate, at least when U approaches one, since in the steady state Π then approaches zero, unlike the full curve shown in Figure 5.4. We can see the ill-posedness as follows. If we define

$$W = \int^t U(\theta)\, d\theta, \tag{5.86}$$

then $R = W - W_1$, and (5.85) is

$$\dot{W}_1 = \frac{2\Pi}{(1 - W + W_1)^2}, \tag{5.87}$$

which is an equation of advanced type, i. e., a delay equation with negative delay. In assessing the instability of the steady state, we then find an infinite number of unstable states, whose (complex) growth rate σ tends to infinity in the complex plane with $\operatorname{Re}\sigma > 0$.

To be specific, denote the steady state as $R = U = U^*$, and linearise the reduced model (5.85). The solutions are exponential, thus

$$U - U^* = e^{\sigma t}, \quad R - U^* = \frac{1}{\sigma}\left(1 - e^{-\sigma}\right) e^{\sigma t}, \tag{5.88}$$

and we require σ to satisfy the transcendental equation

$$g_0(\sigma) = \tfrac{1}{2}\left(1 - U^*\right)^2 e^{-\sigma} - \frac{U^*\left(1 - U^*\right)}{\sigma}\left(1 - e^{-\sigma}\right) = 0. \tag{5.89}$$

It is straightforward to show that this equation has an infinite number of roots in the complex plane, and these tend to the essential singularity at ∞ in the right half plane, $\mathrm{Re}\,\sigma > 0$. This accumulation of rapidly growing modes is the signal of an ill-posed equation of advance.

Putting $\varepsilon = 0$ is thus a singular approximation. There are a number of small terms in (5.84) which might regularise the model. The principal suspect in this regard is the sub-cooled frictional pressure drop. Adding this yields the model

$$\Pi \approx \tfrac{1}{2}U_1(1-R)^2 + \varepsilon U^2 R. \tag{5.90}$$

This gives an equation of mixed type, and regularises the model if ε is large enough. Linear stability of the steady state yields the equation for the growth rate σ as

$$g_\varepsilon(\sigma) = g_0(\sigma) + \varepsilon U^{*2}\left[2 + \frac{1-e^{-\sigma}}{\sigma}\right] = 0. \tag{5.91}$$

Of concern is the sector where $\sigma \to \infty$. It is easy to show that $\mathrm{Re}\,\sigma$ cannot tend to $+\infty$. If $\mathrm{Re}\,\sigma$ is bounded, then we find

$$\sigma \approx \ln\left[\frac{(1-U^*)^2}{4\varepsilon U^*}\right] \pm (2n+1)i\pi \tag{5.92}$$

for large integer n, and thus the model is regularised if

$$\varepsilon > \frac{(1-U^*)^2}{4U^*}. \tag{5.93}$$

This conditional regularisation of the model is reminiscent of the conditional regularisation of two fluid models of bubbly flow for low enough void fraction, associated (perhaps) with flow régime transition boundaries.

The fact that the regularisation is conditional, and in particular does not apply for sufficiently small ε, suggests that a further regularisation is necessary. The correct term to include is the derivative term, thus we replace (5.90) by

$$\Pi \approx \tfrac{1}{2}U_1(1-R)^2 + \varepsilon U^2 R + \nu R\dot{U}, \tag{5.94}$$

where

$$\nu = \varepsilon\delta. \tag{5.95}$$

This unequivocally regularises the model. Linear stability of the steady state is determined by values of σ for which

$$g_{\varepsilon,\nu}(\sigma) = g_\varepsilon(\sigma) + \nu U^*\sigma = 0. \tag{5.96}$$

Now as $\sigma \to \infty$, we must have

$$\nu U^*\sigma + \tfrac{1}{2}(1-U^*)^2 e^{-\sigma} \approx 0, \tag{5.97}$$

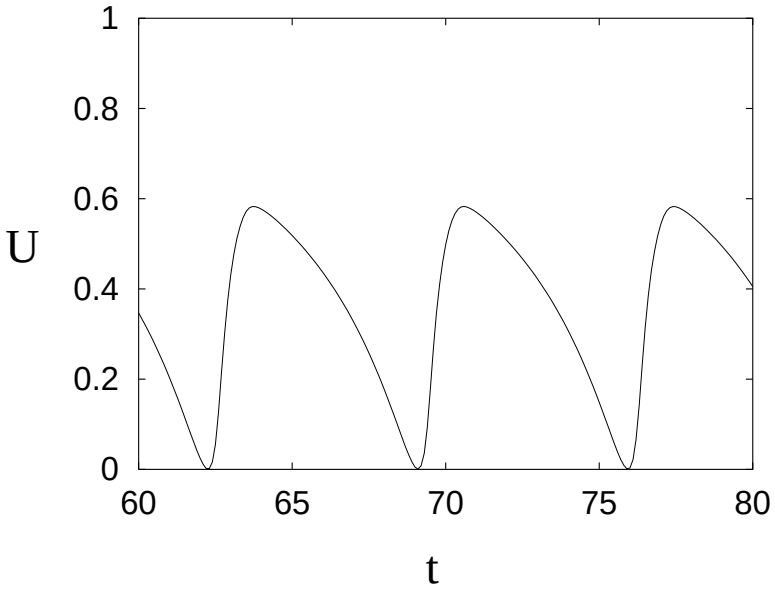

Figure 5.5: Solution of (5.99) with $\gamma = 2.5$, $\delta = 0.025$, $\varepsilon = 0.03$, $\Pi = 0.08$, and $\nu = 0.88$.

and it follows from this that

$$\sigma \approx 2ni\pi - \ln 2n\pi + \ldots \tag{5.98}$$

for large integer n. The problem is thus regularised, although of course there will be many unstable modes with smaller $|\sigma|$.

To write an approximate model, we retain the regularising term $\nu R\dot{U}$, but ignore the other term in ν, which allows us to write the model as a pair of delay differential equations:

$$\begin{aligned} \nu R\dot{U} &= \Pi - \tfrac{1}{2}U_1(1-R)^2 - \varepsilon U^2 R \\ &\qquad -\varepsilon\gamma\left[\varepsilon U_1 \ln\left(\frac{1-R}{\varepsilon U_1}\right) + R\right] - \delta\left[U_1(1-R) - \varepsilon U^2\right], \\ \dot{R} &= U - U_1. \end{aligned} \tag{5.99}$$

Although it appears that a useful further approximation can be made by putting the terms in ε and δ to zero, this is not numerically accurate, because in practice (see Figure 5.4) Π is small.

(5.99) has the appearance of a singularly perturbed delay differential equation, and it is challenging to find asymptotic solutions of (5.99). In addition, this system has an added twist. If we fix the parameters $\varepsilon = 0.03$, $\delta = 0.025$, $\gamma = 2.5$, $\Pi = 0.08$, and progressively reduce ν, we find that the steady state at $\Pi = 0.08$, $U = 0.181$ (see Figure 5.4) has a Hopf bifurcation to a stable limit cycle as ν is reduced through

0.984. As ν decreases, the amplitude grows until at $\nu = 0.88$ the inlet velocity decreases to zero (see Figure 5.5), and for $\nu < 0.88$, reversed flow occurs. The model is not applicable for reversed flow, and indeed the numerical solution breaks down in that case. This appears consistent with theoretical stability results of many authors, which indicate that instability is commonplace at small ε. The twist here is that the model itself precludes attainment of the relevant asymptotic limit.

Notes and references

Exercises

5.1

References

Aldridge and Fowler

Baines, P. G. and A. E. Gill 1969 On thermohaline convection with linear gradients. J. Fluid Mech. **37**, 289–306.

Baines, W. D. and J. S. Turner 1969 Turbulent buoyant convection from a source in a confined region. J. Fluid Mech. **37**, 51–80.

Balmforth, N. J., A. Provenzale and J. A. Whitehead 2001 The language of pattern and form. In: Geomorphological fluid mechanics, eds. N. J. Balmforth and A. Provenzale, pp. 3–33. Springer-Verlag, Berlin.

Batchelor, G. K. 1967 An introduction to fluid dynamics. C. U. P., Cambridge.

Davies, G. F. 1999 Dynamic Earth: plates, plumes and mantle convection. C. U. P., Cambridge.

Davies and Potter 1967

Delmastro

Drew 1983

Drew and Wood 1985

Holmes, A. 1978 Principles of physical geology. rd edition, revised by Doris Holmes. John Wiley and sons, New York.

Howard, L. N. 1966 Convection at high Rayleigh number. Proc. 11th Int. Cong. Appl. Mech., ed. H. Görtler, pp. 1109–1115. Springer, Berlin.

Jimenez, J. and J. A. Zufiria 1987 A boundary layer analysis of Rayleigh-Bénard convection at large Rayleigh number. J. Fluid Mech. **178**, 53–71.

Linden, P. F. 2000 Convection in the environment. In: Perspectives in fluid dynamics, eds. G. K. Batchelor, H. K. Moffatt and M. G. Worster, pp. 289–345. C. U. P., Cambridge.

McBirney, A. R. and R. M. Noyes 1979 Crystallisation and layering of the Skaergaard intrusion. J. Petrol. **20**, 487–554.

Morton, B. R., G. Taylor and J. S. Turner 1956 Turbulent gravitational convection from maintained and instantaneous sources. Proc. R. Soc. Lond. A**234**, 1–23.

Roberts, G. O. 1979 Fast viscous Bénard convection. Geophys. Astrophys. Fluid Dynamics **12**, 235–272.

Turcotte, D. L. and E. R. Oxburgh 1967 Finite amplitude convection cells and continental drift. J. Fluid Mech. **28**, 29–42.

Turcotte, D. L. and G. Schubert 1982 Geodynamics: applications of continuum physics to geological problems. John Wiley, New York.

Turner, J. S. 1979 Buoyancy effects in fluids. C. U. P., Cambridge.

Wager, L. R. and G. M. Brown 1968 Layered igneous rocks. Oliver and Boyd, Edinburgh.

Wegener, A. L. 1924 The origin of continents and oceans. rd ed., trans. J. G. A. Skerl. Methuen, London.