HYDRODYNAMICS AND MAGNETOHYDRODYNAMICS FOR SOLAR AND STELLAR APPLICATIONS

Dana Longcope
Montana State University, Bozeman

June 29, 2017

Contents

Chapter 1

Fluid Equations

1A State variables for a fluid

A fluid is a material continuum capable of flowing. Its state at any time is fully described by the following functions of space and time

$$\begin{aligned} \rho(\mathbf{x},t) &: \text{mass density} \quad [\text{g/cm}^3] \\ \mathbf{u}(\mathbf{x},t) &: \text{fluid velocity} \quad [\text{cm/sec}] \\ p(\mathbf{x},t) &: \text{pressure} \quad [\text{erg/cm}^3] \ . \end{aligned}$$

There are five functions in all, counting separately u_x, u_y and u_z.

The fluid state variables are basically densities, intended to be integrated over volume in order to obtain traditional physical quantities. A *fluid parcel* occupying a volume $\mathcal{V}$ is characterized by the following physical properties:

$$\text{mass} \ : \ M = \int_{\mathcal{V}} \rho(\mathbf{x})\, d^3x \tag{1.1}$$

$$\text{momentum} \ : \ \mathbf{P} = \int_{\mathcal{V}} \rho\mathbf{u}\, d^3x \tag{1.2}$$

$$\text{kinetic energy} \ : \ E_K = \int_{\mathcal{V}} \tfrac{1}{2}\rho|\mathbf{u}|^2\, d^3x \tag{1.3}$$

$$\text{thermal energy} \ : \ E_T = \frac{1}{\gamma-1}\int_{\mathcal{V}} p(\mathbf{x})\, d^3x \tag{1.4}$$

The first three should be very familiar from standard Newtonian mechanics. The final expression is more common in thermodynamics. Its form assumes, as we *always* will, that the fluid is an *ideal gas*; γ is known as the *ratio of specific heats* or the adiabatic index.[1] For a gas of monatomic particles, such as ions and electrons (i.e. a plasma) $\gamma = 5/3$, so its internal energy is an integral of $(3/2)p$. Earth's air is composed primarily of diatomic molecules which have two classical rotational degrees of freedom in addition to their three

[1] Under an adiabatic process, the pressure p and volume V of an ideal gas change so as to conserve the product pV^γ.

translational dimensions; this gives air an adiabatic index $\gamma \simeq 7/5$. These relations show that the fluid's state variables, ρ, $\mathbf{u}$, and p, describe its mass, momentum and energy.

In treating a fluid we will almost *never* refer to the basic particles which compose it. While a fluid is composed of individual particles, it turns out to be diabolically difficult to understand its properties in those terms. One of the most profound thinkers in history, Isaac Newton, tried this and failed rather miserably. We return at the end of these lectures to explore the very complicated relation between the particles and the fluid they compose. In the mean time we work exclusivel in terms of a continuum described by the smooth functions of space and time $\rho(\mathbf{x}, t)$, $\mathbf{u}(\mathbf{x}, t)$, and $p(\mathbf{x}, t)$.

Rather than particles we will refer to *fluid elements* — infinitesimally small fluid parcels. Consider a fluid element with center of mass at $\mathbf{x} = \mathbf{r}$. This is a parcel whose linear dimension $\delta\ell$ is small enough that all the fluid properties are approximately constant over that scale. This means that the integrands in eqs. (1.1)–(1.4) are constant and can be pulled out of the integral to yield

$$\begin{aligned} \text{mass} \quad &: \quad M \;\simeq\; \rho(\mathbf{r}, t)\, \delta V \\ \text{momentum} \quad &: \quad \mathbf{P} \;\simeq\; \rho(\mathbf{r}, t)\mathbf{u}(\mathbf{r}, t)\, \delta V \\ \text{thermal energy} \quad &: \quad E_T \;\simeq\; \frac{p(\mathbf{r}, t)}{\gamma - 1}\, \delta V \end{aligned}$$

where δV is the fluid element's volume.

The center-of-mass velocity of the fluid element is (Marion, 1970, §2.6)

$$\mathbf{v}_{\rm cm} \;=\; \frac{\mathbf{P}}{M} \;=\; \mathbf{u}(\mathbf{r}, t) \quad .$$

This is the time derivative of the center of mass position, so the fluid element's trajectory $\mathbf{r}(t)$ satisfies the equation

$$\frac{d\mathbf{r}}{dt} \;=\; \mathbf{u}[\,\mathbf{r}(t),\, t\,] \quad . \tag{1.5}$$

The velocity function $\mathbf{u}(\mathbf{x}, t)$ gives the velocity of *fluid elements*, *not* the velocity of particles composing the fluid. It is to make this distinction excessively clear that we choose the variable $\mathbf{u}$ to represent fluid velocity; at a much later point we will discuss particles composing the fluid and represent *their* velocities using $\mathbf{v}$.

The *specific energy* of the fluid element (energy per unit mass) is

$$\varepsilon \;=\; \frac{E_T}{M} \;=\; \frac{1}{\gamma - 1}\frac{p}{\rho} \quad . \tag{1.6}$$

1B The advective derivative

Consider a generic fluid property $f(\mathbf{x}, t)$, such as density or temperature or salinity. The value of this property within a fluid element following trajectory $\mathbf{r}(t)$,

$$f_{\mathbf{r}}(t) \;=\; f[\,\mathbf{r}(t),\, t\,] \;=\; f[\,r_x(t),\, r_y(t),\, r_z(t),\, t\,] \quad ,$$

depends on time in two ways: through intrinsic temporal dependence of f and through motion of the element across spatial variations. The time-rate-of-change of the property, as observed by the fluid element, is found using the chain rule.

$$\frac{df_{\mathbf{r}}}{dt} \;=\; \frac{dr_x}{dt}\frac{\partial f}{\partial x} + \frac{dr_y}{dt}\frac{\partial f}{\partial y} + \frac{dr_z}{dt}\frac{\partial f}{\partial z} + \frac{\partial f}{\partial t} \;=\; \frac{\partial f}{\partial t} + \mathbf{u}\cdot\nabla f \quad . \tag{1.7}$$

This motivates us to define a differential operator known as the **advective derivative**,

$$\frac{D}{Dt} \;=\; \frac{\partial}{\partial t} + \mathbf{u}\cdot\nabla \quad . \tag{1.8}$$

Other names for this operator are *material derivative* (Batchelor, 1967; Meyer, 1971; Priest, 1982; Thompson, 2006), *Lagrangian derivative* (Moffatt, 1978), *convective derivative* (Meyer, 1971), and *substantive derivative* (Tritton, 1977). Based on its appearance in constructions like eq. (1.7), some authors (Landau & Lifshitz, 1959; Fetter & Walecka, 1980; Choudhuri, 1998; Foukal, 2004; Kulsrud, 2005) use the same symbol for the advective derivative as for the total derivative: d/dt. This can lead to confusion. The advective derivative makes a very specific choice which bears constant reminding: the quantity is being time-differentiated in the reference frame of a fluid element. For this reason I and many others (Lamb, 1932; Batchelor, 1967; Moffatt, 1978; Tritton, 1977; Priest, 1982; Thompson, 2006) prefer a unique symbol for it.

Regardless of its complex definition (and unfamiliar symbol) the advective derivative is a traditional first derivative in all respects

$$\begin{aligned}
&\frac{D}{Dt}(f+g) = \frac{Df}{Dt} + \frac{Dg}{Dt} && \text{linear}\\
&\frac{D}{Dt}(fg) = g\frac{Df}{Dt} + f\frac{Dg}{Dt} && \text{product rule}\\
&\frac{D}{Dt}\Psi[f(\mathbf{x},t)] = \Psi'(f)\,\frac{Df}{Dt} && \text{chain rule}
\end{aligned}$$

as can be easily checked using its definition, eq. (1.8). It is always possible to replace it by this definition in terms of traditional partial derivatives. Using the properties above can often, however, save some effort. Still, one must be careful about moving it around other derivatives. For example.

$$\frac{\partial}{\partial t}\left(\frac{Df}{Dt}\right) \;=\; \frac{\partial^2 f}{\partial t^2} + \mathbf{u}\cdot\nabla\frac{\partial f}{\partial t} + \frac{\partial \mathbf{u}}{\partial t}\cdot\nabla f \;=\; \frac{D}{Dt}\left(\frac{\partial f}{\partial t}\right) + \frac{\partial \mathbf{u}}{\partial t}\cdot\nabla f \;\neq\; \frac{D}{Dt}\left(\frac{\partial f}{\partial t}\right) \quad .$$

In mathematical jargon, the operators $\partial/\partial t$ and D/Dt do not commute. Nor does it commute with any spatial derivatives,

$$\frac{\partial}{\partial x}\left(\frac{Df}{Dt}\right) \;=\; \frac{D}{Dt}\left(\frac{\partial f}{\partial x}\right) + \frac{\partial \mathbf{u}}{\partial x}\cdot\nabla f \;\neq\; \frac{D}{Dt}\left(\frac{\partial f}{\partial x}\right) \quad .$$

1C The fluid equations

The state of a fluid at any instant is completely described by the five fields, $\rho(\mathbf{x},t)$, $\mathbf{u}(\mathbf{x},t)$, and $p(\mathbf{x},t)$. These evolve in time according to five governing equations: the fluid equations. Hydrodynamics is nothing more than a study of the solutions to those governing equations.

Perhaps the most confusing thing about hydrodynamics is the wide variety of forms the fluid equations can take. Here are all the fluid equations in one place in their most general form as we will most often use them

$$\frac{\partial \rho}{\partial t} \; + \; \nabla \cdot (\rho \mathbf{u}) \; = \; 0 \quad , \qquad \text{(C) continuity}$$

$$\rho \frac{D\mathbf{u}}{Dt} \; = \; \rho \frac{\partial \mathbf{u}}{\partial t} \; + \; \rho(\mathbf{u} \cdot \nabla)\mathbf{u} \; = \; -\nabla p \; + \; \mathbf{f} \qquad \text{(M) momentum}$$

$$\frac{Dp}{Dt} \; = \; \frac{\partial p}{\partial t} \; + \; \mathbf{u} \cdot \nabla p \; = \; -\gamma p (\nabla \cdot \mathbf{u}) + (\gamma - 1)\dot{Q} \qquad \text{(E) energy}$$

The functions $\mathbf{f}$ and $\dot{Q}$ are, respectively, the force density and volumetric heating rate.[2] They are place-holders for various expressions we will use at different times, depending on the physical problem being tackled. In many cases one or both will be absent altogether; this common scenario, $\mathbf{f} = 0$, $\dot{Q} = 0$, is known as **ideal hydrodynamics**.

The most significant feature of this set of equations is that it provides an explicit recipe for time-advancing each of the state variables. Furthermore the recipe depends only on those same state variables, provided $\mathbf{f}$ and $\dot{Q}$ can be found from them. Thus they constitute a complete and closed system of equations for the time-evolution of the fluid.

In addition to a menu of different choices for the place-holders $\mathbf{f}$ and $\dot{Q}$, the equations can be cast into different forms using manipulations of vector calculus. We discuss now a few common alternatives derived through such manipulation.

1C.1 Continuity

Version (C) of the continuity equation is known as its *conservative form*. An alternative, called its *Lagrangian form*, is found by using the product rule of the divergence

$$\frac{D\rho}{Dt} \; = \; -\rho(\nabla \cdot \mathbf{u}) \quad . \qquad \text{(C-L) Lagrangian continuity}$$

The advective derivative tells us that this equation gives the density change for a particular fluid element — i.e. the change observed as one follows the fluid element. It is evident from the expression that the density will increase wherever $\nabla \cdot \mathbf{u} < 0$ (convergence) and decrease wherever $\nabla \cdot \mathbf{u} > 0$ (divergence); where $\nabla \cdot \mathbf{u} = 0$ the density of a fluid element remains constant.

There are circumstances, known as *incompressible flows*, where $\nabla \cdot \mathbf{u} = 0$ everywhere and therefore ρ never changes. It is tempting to think that incompressible flow occurs *because* density cannot be changed. The most typical reason that $\nabla \cdot \mathbf{u} = 0$, however, is that the flow is very subsonic. Subsonic flow, it turns out, "lacks sufficient inertia" to change the density (or pressure). We will return to this limit often, since many flows are subsonic.

The specific volume of a fluid element (volume per unit mass), $v = \delta V / M = 1/\rho$, satisfies a variant form of the continuity equation

$$\frac{D}{Dt}\left(\frac{1}{\rho}\right) \; = \; \left(\frac{1}{\rho}\right) \nabla \cdot \mathbf{u} \quad . \tag{1.9}$$

[2] The dot on $\dot{Q}$ is not really a time derivative but a notational reminder that it is a rate of heating. Its units are $\mathrm{erg\,cm^{-3}\,s^{-1}}$.

1C.2 Momentum

The momentum equation most often appears in its Lagrangian form (M). It is obviously related to Newton's second law given that $D\mathbf{u}/Dt$ is the acceleration of a fluid element. The volumetric form of $m\mathbf{a}$ is therefore $\rho D\mathbf{u}/Dt$. This then is set equal to the volumetric forces acting on the fluid element: the rhs of eq. (M).

The conservative form is found combining (C) and (M)

$$\frac{\partial(\rho u_i)}{\partial t} + \nabla\cdot(\rho\mathbf{u}u_i) \;=\; -\frac{\partial p}{\partial x_i} \;+\; f_i \;\;, \qquad \text{(M-c) conservative momentum}$$

where $i = 1, 2$ or 3.

1C.3 Energy

The conservative form of the energy equation comes from adding $p(\nabla\cdot\mathbf{u})$ to both sides of eq. (E) and collecting terms

$$\frac{\partial p}{\partial t} \;+\; \nabla\cdot(\mathbf{u}p) \;=\; -(\gamma-1)p(\nabla\cdot\mathbf{u}) \;+\; (\gamma-1)\dot{Q} \;\;, \qquad \text{(E-c) cons've energy}$$

Other forms of the equation make more direct reference to the specific energy, derived in (1.6). For an ideal gas of particles with average mass m this is

$$\varepsilon \;=\; \frac{1}{\gamma-1}\frac{p}{\rho} \;=\; \frac{k_{\rm B}T}{(\gamma-1)m} \;\;, \tag{1.10}$$

where $k_{\rm B}$ is Boltzman's constant,[3] and T is the temperature. The first law of thermodynamics, written per unit mass, is

$$d\varepsilon \;=\; -\,p\;dv + T\,ds \;=\; -\,p\,d\left(\frac{1}{\rho}\right) + T\,ds \;\;, \tag{1.11}$$

where the specific entropy for an ideal gas is

$$s \;=\; \frac{k_{\rm B}/m}{\gamma-1}\ln(p/\rho^\gamma) \;\;, \tag{1.12}$$

up to an additive constant. In terms of advective derivatives the first law becomes

$$\frac{D\varepsilon}{Dt} \;=\; -\,p\frac{D}{Dt}\left(\frac{1}{\rho}\right) + T\,\frac{Ds}{Dt} \;=\; -\,p\frac{D}{Dt}\left(\frac{1}{\rho}\right) + \frac{\dot{Q}}{\rho} \tag{1.13}$$

where we have introduced the volumetric (per unit volume) heating rate[4]

$$\dot{Q} \;=\; \rho\,T\,\frac{Ds}{Dt} \;\;. \tag{1.14}$$

[3] In order to avoid using properties of individual particles it is common to introduce their ratio, $R = k_{\rm B}/m$, as the so-called *gas constant*, which is independent of their size. Since we are twenty-first century Physicists, rather than nineteenth century Chemists, we continue to use the per-particle quantities $k_{\rm B}$ and m even though the fluid equations do not involve particles directly.

[4] Some authors (Choudhuri, 1998), viewing the gain of energy as a negative loss, use the notation $\mathcal{L} = -\dot{Q}$. This strikes me as just too many negative signs.

The menu of possible terms which might contribute to $\dot{Q}$ lists the number of things which might change the entropy of a fluid element. When its entropy does not change, a situation called *adiabatic*, then we set $\dot{Q} = 0$.

Using specific energy of an ideal gas, eq. (1.10), to replace ε with the ratio of pressure and density gives

$$\frac{D}{Dt}\left(\frac{p}{\rho}\right) \;=\; \frac{1}{\rho}\frac{Dp}{Dt} + p\frac{D}{Dt}\left(\frac{1}{\rho}\right) \;=\; -(\gamma-1)p\frac{D}{Dt}\left(\frac{1}{\rho}\right) \;+\; (\gamma-1)\frac{\dot{Q}}{\rho} \quad .$$

Using (1.9) to eliminate advective derivatives of specific volume leads to the form of energy equation (E) explicitly in terms of pressure.

A useful alternate form of the energy equation comes form using the ideal gas relation, (1.10), to replace ε with T. This leads to a version of the energy equation prescribing the temperature evolution

$$\frac{DT}{Dt} \;=\; -(\gamma-1)T(\nabla\cdot\mathbf{u}) \;+\; (\gamma-1)(m/k_{\rm B})\,\frac{\dot{Q}}{\rho} \quad . \tag{1.15}$$

Temperature can be time-advanced, rather than p, provided p can be eliminated from $\dot{Q}$ — usually easy. Pressure is still required for the momentum equation and must be found using the ideal gas law

$$p \;=\; \frac{k_{\rm B}}{m}\,\rho T \quad , \tag{1.16}$$

which follows from equating the two right expression in (1.10). I do not include this among the basic fluid equations because in their most common form they make no specific use of T. Equation (1.16) is then an auxiliary equation defining T in terms of fluid state variables.

It is worth remarking on one peculiar feature of eq. (1.15), to which we will return on occasion. If we make the formal assignment $\gamma = 1$ we find the temperature does not change: $DT/Dt = 0$. If the fluid is also *isothermal*, $\nabla T = 0$, then it will remain so. Physically the limit $\gamma \to 1$ corresponds to a gas of particles with a huge number of internal degrees of freedom. Energy is shared between all of these and adding more energy does little or nothing to change the temperature. Since a plasma has $\gamma = 5/3$ it seems like we are never able to use this trick. We will, however, find occasion we can use it nevertheless.

1D Conservation laws

Conservation of mass, momentum and energy can be used, and often *are* used, to derive the fluid equations. I have chosen instead to introduce the equations and will now demonstrate that they satisfy the expected conservation laws. This tack is, in fact, far more useful in practice that the bottoms-up approach. There is frequent call to modify the governing fluid equations for particular applications. Incompressible hydrodynamics is one widely known example. After making such a modification it is sometimes important to establish that the modified equations satisfy conservation laws — perhaps modified ones. Each time one is faced with this task, one follows the same approach we present here: begin with the governing PDEs and derive *from them* laws for the conservation of integral quantities.

1D.1 Conservation of mass

Consider the conservation for a volume of finite (i.e. non-infinitesimal) size. There are two obvious possibilities for the time evolution of this volume. In the simplest case the volume $\mathcal{R}$ is fixed in space and fluid flows through its imaginary boundaries $\partial\mathcal{R}$. Time derivative of an integral, such as in the mass defined by eq. (1.1), will not involve a derivative of the limits of integration (i.e. $\partial\mathcal{R}$) since these are fixed. All spatial coordinates of the integrand, $\rho(x, y, z, t)$, are dummy variables of integration so the time derivative of the integrand is a partial (Eulerian) derivative which holds them fixed

$$\frac{dM_{\mathcal{R}}}{dt} \;=\; \frac{d}{dt}\int_{\mathcal{R}} \rho(\mathbf{x},t)\, d^3x \;=\; \int_{\mathcal{R}} \frac{\partial\rho}{\partial t}\, d^3x \quad . \tag{1.17}$$

Using eq. (C) this becomes

$$\frac{dM_{\mathcal{R}}}{dt} \;=\; -\int_{\mathcal{R}} \nabla\cdot(\rho\mathbf{u})\, d^3x \;=\; -\oint_{\partial\mathcal{R}} \rho\mathbf{u}\cdot d\mathbf{a} \quad , \tag{1.18}$$

where $d\mathbf{a}$ is the outward-directed area element of $\partial\mathcal{R}$. This is the sense in which mass is "conserved": it cannot appear or disappear from within $\mathcal{R}$. The mass of the stationary volume will, however, change as fluid flows across its boundaries with mass flux $\rho\mathbf{u}$.

In the second case $\mathcal{V}$ is a fluid parcel and moves with the flow. The derivative of the integral, eq. (1.1), now includes the time-derivative of the limits of integration

$$\frac{dM}{dt} \;=\; \frac{d}{dt}\int_{\mathcal{V}} \rho(\mathbf{x},t)\, d^3x \;=\; \int_{\mathcal{V}} \frac{\partial\rho}{\partial t}\, d^3x \;+\; \oint_{\partial\mathcal{V}} \rho v_n\, da \quad , \tag{1.19}$$

where $\partial\mathcal{V}$ moves at velocity v_n increasing differential volume at a rate $v_n\, da$. Since $\partial\mathcal{V}$ moves with the fluid, its normal velocity is $v_n = \mathbf{u}\cdot\hat{\mathbf{n}}$, where $\hat{\mathbf{n}}$ is the outward normal. Along with $d\mathbf{a} = \hat{\mathbf{n}}\, da$, this gives

$$\frac{dM}{dt} \;=\; \int_{\mathcal{V}} \frac{\partial\rho}{\partial t}\, d^3x + \oint_{\partial\mathcal{V}} \rho\mathbf{u}\cdot d\mathbf{a} \;=\; \int_{\mathcal{V}} \left[\, \frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{u}) \,\right] d^3x \quad , \tag{1.20}$$

after using Gauss's theorem to rewrite the surface integral as the volume integral of a divergence. The continuity equation, eq. (C), immediately leads to the vanishing of the right hand integrand, meaning $dM/dt = 0$: the mass of the fluid parcel does not change. This is genuine mass conservation.

1D.2 Conservation of momentum

The same technique can be used to time-differentiate the momentum of a fluid parcel, eq. (1.2),

$$\frac{dP_i}{dt} \;=\; \int_{\mathcal{V}} \left[\, \frac{\partial(\rho u_i)}{\partial t} + \nabla\cdot(\rho u_i\mathbf{u}) \,\right] d^3x \quad , \tag{1.21}$$

by simply substituting ρu_i for ρ in eq. (1.20). Using the conservative version of the momentum equation, eq. (M-c), leads to

$$\frac{dP_i}{dt} \;=\; -\int_{\mathcal{V}} \frac{\partial p}{\partial x_i}\, d^3x + \int_{\mathcal{V}} f_i\, d^3x \quad . \tag{1.22}$$

In order to cast the pressure gradient as a divergence we introduce the Kronecker δ-function

$$\frac{\partial p}{\partial x_i} \;=\; \frac{\partial}{\partial x_j}(p\,\delta_{ij}) \quad \text{or} \quad \nabla p \;=\; \nabla\cdot(p\,\underline{\underline{\mathrm{I}}}) \;\;,$$

where $\underline{\underline{\mathrm{I}}}$ is the identity matrix, and repeated indices are to be summed over. Using Gauss's law for this divergence (you may check that it also works for the divergence of a tensor)

$$\frac{d\mathbf{P}}{dt} \;=\; -\oint_{\partial\mathcal{V}} p\underline{\underline{\mathrm{I}}}\cdot d\mathbf{a} \;+\; \int_{\mathcal{V}} \mathbf{f}\,d^3x \;=\; -\oint_{\partial\mathcal{V}} p\,d\mathbf{a} \;+\; \underbrace{\int_{\mathcal{V}} \mathbf{f}\,d^3x}_{\mathbf{F}} \;\;. \tag{1.23}$$

According to Newton's second law both terms on the right of eq. (1.23) are forces acting on the fluid parcel to change its momentum. Naturally the integral of the force density $\mathbf{f}$ is a force, $\mathbf{F}$. Because $\mathbf{f}$ acts of each bit of material inside the parcel it is often referred to as a *body force*.[5] Body forces include all the fundamental forces of nature, electric, gravitational, etc., which the particles composing the fluid elements feel directly.

Pressure is the prototypical example of a force which is *not* a body force. It acts only on the surface of the fluid parcel, and is not felt by the fluid inside. Indeed, individual particles do not feel anything like the pressure force — it is a force only felt by *fluids*. Each outward-directed surface element, $d\mathbf{a}$, experiences an inward force $-p\,d\mathbf{a}$, from the surrounding fluid.

The momentum change of a stationary volume, $\mathcal{R}$,

$$\frac{d\mathbf{P}_{\mathcal{R}}}{dt} \;=\; -\oint_{\partial\mathcal{R}} \rho\mathbf{u}\mathbf{u}\cdot d\mathbf{a} \;-\; \oint_{\partial\mathcal{R}} p\,d\mathbf{a} \;+\; \int_{\mathcal{R}} \mathbf{f}\,d^3x \;\;, \tag{1.24}$$

contains one additional term (first on the right) due to momentum carried by fluid across its imaginary boundary, $\partial\mathcal{R}$. Even though the volume itself is not moving, it contains moving fluid and will therefore have non-zero momentum in general. Note that the momentum flux appearing in the first surface integral, is a second-rank tensor. We express this in dyadic notation, $\rho\mathbf{u}\mathbf{u}$, and its contraction with the area element is self-evident: $\rho\mathbf{u}\mathbf{u}\cdot d\mathbf{a} = \rho\mathbf{u}(\mathbf{u}\cdot d\mathbf{a})$.

1D.3 Conservation of energy

The time derivative of the kinetic energy integral, (1.3), is

$$\begin{aligned}
\frac{dE_K}{dt} \;&=\; \tfrac{1}{2}\frac{d}{dt}\int_{\mathcal{V}} \rho u_i u_i\,d^3x \;=\tfrac{1}{2}\int_{\mathcal{V}}\left[\frac{\partial(\rho u_i u_i)}{\partial t} + \nabla\cdot(\rho u_i u_i\mathbf{u})\right] d^3x \\
&=\; \tfrac{1}{2}\int_{\mathcal{V}}\left[u_i\frac{\partial(\rho u_i)}{\partial t} + \rho u_i\frac{\partial u_i}{\partial t} + u_i\nabla\cdot(\rho u_i\mathbf{u}) + \rho u_i\mathbf{u}\cdot\nabla u_i\right] d^3x \;\;, \\
&=\; \int_{\mathcal{V}}\left\{\tfrac{1}{2}u_i\left[\frac{\partial(\rho u_i)}{\partial t} + \nabla\cdot(\rho u_i\mathbf{u})\right] + \tfrac{1}{2}u_i\left[\rho\frac{\partial u_i}{\partial t} + \rho(\mathbf{u}\cdot\nabla)u_i\right]\right\} d^3x \;\;,
\end{aligned}$$

using the standard product rule and re-arranging some terms. Taking the dot product of $\mathbf{u}$ with both the conservative and Lagrangian forms of the momentum equations, eqs. (M-c)

[5] We will include at least one term, namely the viscosity, which is a divergence and therefore is not strictly speaking a body force.

and (M), we can rewrite this as

$$\frac{dE_K}{dt} \;=\; -\int_{\mathcal{V}} \mathbf{u}\cdot\nabla p\, d^3x \;+\; \int_{\mathcal{V}} \mathbf{u}\cdot\mathbf{f}\, d^3x \quad . \tag{1.25}$$

Next we time-differentiate the thermal energy and use the conservative energy equation, (E-c) to get

$$\frac{dE_T}{dt} \;=\; \frac{1}{\gamma-1}\int_{\mathcal{V}}\left[\frac{\partial p}{\partial t}+\nabla\cdot(p\mathbf{u})\right] d^3x \;=\; -\int_{\mathcal{V}} p(\nabla\cdot\mathbf{u})\, d^3x \;+\; \int_{\mathcal{V}} \dot{Q}\, d^3x \quad . \tag{1.26}$$

Summing the two expressions above gives the change in total energy $E = E_K + E_T$ as

$$\frac{dE}{dt} \;=\; -\oint_{\partial\mathcal{V}} p\mathbf{u}\cdot d\mathbf{a} \;+\; \int_{\mathcal{V}} \mathbf{u}\cdot\mathbf{f}\, d^3x \;+\; \int_{\mathcal{V}} \dot{Q}\, d^3x \quad . \tag{1.27}$$

Thus the energy of a fluid parcel is not really conserved: it is changed by work or heating done on it. The first term is the work done on the fluid parcel by the pressure of the surrounding fluid, the second is work done by the other forces and the third is the net heating.

The change in total energy of a stationary volume, $\mathcal{R}$, can be worked out with similar algebraic steps, but omitting contributions from changes to the boundaries $\partial\mathcal{R}$. The result is very similar,

$$\frac{dE}{dt} \;=\; -\oint_{\partial\mathcal{R}} \left(\tfrac{1}{2}\rho|\mathbf{u}|^2\mathbf{u}+\frac{\gamma}{\gamma-1}p\mathbf{u}\right)\cdot d\mathbf{a} \;+\; \int_{\mathcal{R}} \mathbf{u}\cdot\mathbf{f}\, d^3x \;+\; \int_{\mathcal{R}} \dot{Q}\, d^3x \quad , \tag{1.28}$$

except that the surface integral now represents a *flux* of energy across the imaginary boundary. The first term is the flux of kinetic energy carried by the moving fluid, and the second is a kind of thermal energy flux. The fluid's thermal energy density is $p/(\gamma-1)$, according to eq. (1.4). This will be carried with the moving fluid, providing an energy flux, $p\mathbf{u}/(\gamma-1)$. In addition to carrying its own energy the flow *does work*, at a rate $p\mathbf{u}$, on the internal fluid as it crosses the boundary – this is sometimes called "$p\,dV$ work". This work is also present in eq. (1.27), even though the flow there does not cross the boundary (the fluid parcel moves with the flow). These combined contributions appears as a flux of the sum $w = p + p/(\gamma-1) = \gamma p/(\gamma-1)$, known as *enthalpy*.[6] The energy changes through an enthalpy flux $w\mathbf{u}$ carrying energy across the boundary.

Other energies: gravitational

The foregoing showed that combined kinetic plus thermal energy of the fluid could change due to several contributions. Among these was work done by the catch-all force density term $\mathbf{f}(\mathbf{x},t)$. There are certain cases where this force is itself conservative and its contribution to the energy change can be replaced, or partially replaced, by a change in some other form of potential energy. This is exactly the kind of argument used in Freshman Physics to define potential energies for gravity, springs etc.

[6] Thermodynamics more often uses the extensive energy, $E = pV/(\gamma-1)$. The total enthalpy adds to this the mechanical work to get $W = E + pV = pV\gamma/(\gamma-1)$. In fluid mechanics we work with volumetric densities: $w = W/V$.

As an example consider gravity whose force density (to be presented shortly) is $\mathbf{f}^{(\mathrm{g})} = \rho\mathbf{g} = -\rho\nabla\Psi$, where $\Psi(\mathbf{x})$ is the gravitational potential. This will change the fluid's energy at a rate

$$\int_{\mathcal{R}} \mathbf{u}\cdot\mathbf{f}^{(\mathrm{g})}\, d^3x \;=\; -\int_{\mathcal{R}} \rho\mathbf{u}\cdot\nabla\Psi\, d^3x \quad , \tag{1.29}$$

found by using $\mathbf{f}^{(\mathrm{g})}$ in the second term on the rhs of eq. (1.28). We can integrate this expression by parts and use the continuity equation (C) to rewrite it

$$\begin{aligned}\int_{\mathcal{R}} \mathbf{u}\cdot\mathbf{f}^{(\mathrm{g})}\, d^3x &= -\int_{\mathcal{R}} \nabla\cdot(\Psi\rho\mathbf{u})\, d^3x \;+\; \int_{\mathcal{R}} \Psi\nabla\cdot(\mathbf{u}\rho)\, d^3x \\ &= -\oint_{\partial\mathcal{R}} \Psi\rho\mathbf{u}\cdot\, d\mathbf{a} \;-\; \int_{\mathcal{R}} \Psi\frac{\partial\rho}{\partial t}\, d^3x \\ &= -\oint_{\partial\mathcal{R}} \Psi\rho\mathbf{u}\cdot\, d\mathbf{a} \;-\; \underbrace{\frac{d}{dt}\int_{\mathcal{R}} \Psi\rho\, d^3x}_{dE_G/dt} \quad ,\end{aligned} \tag{1.30}$$

after assuming the gravitational potential is static: $\partial\Psi/\partial t = 0$.[7] The final expression shows that the total energy used above, kinetic plus thermal, can be supplemented with a gravitational potential energy

$$E_G \;=\; \int_{\mathcal{R}} \rho\Psi\, d^3x \quad , \tag{1.31}$$

in order to remove the gravitational force from the work term. The augmented total energy, $E = E_K + E_T + E_G$, would then change only due to work done by other (non-gravitational) forces, the heating and flux terms already present in eq. (1.28), and a new flux of gravitational energy on the right hand side of (1.30). That final contribution occurs as mass enters and leaves the stationary region, $\mathcal{R}$, with its own gravitational energy. If the derivation is repeated for a fluid parcel $\mathcal{V}$ the flux term will not appear (try it yourself).

1E A catalog of terms for fluid equations

1E.1 Force density terms:

- **Gravity:** $\mathbf{f}^{(\mathrm{g})} = \rho\mathbf{g}$, where $\mathbf{g}$ is gravitational acceleration. This is a classic example of a body force: everything inside a fluid parcel will feel the pull of gravity.
  For self-gravitating fluid, $\mathbf{g} = -\nabla\Psi$, where the gravitational potential, $\Psi(\mathbf{x}, t)$, satisfies the Poisson equation

  $$\nabla^2\Psi \;=\; 4\pi G\,\rho \quad . \tag{1.32}$$

  On the Sun, a star or a planet we assume the gravitational field is fixed in time and write

  $$\begin{aligned}\mathbf{g} &= -g(r)\hat{\mathbf{r}} \quad \text{spherical} \\ \mathbf{g} &= -g\hat{\mathbf{z}} \quad \text{planar}\end{aligned}$$

  where $g = 980\,\mathrm{cm/s^2}$ on the Earth and $g_\odot = 2.7\times10^4\,\mathrm{cm/s^2}$ at the solar photosphere.

[7] This restriction can be relaxed using the relation $\nabla^2\Psi = 4\pi G\,\rho$ for self-gravitation, and some further integration by parts.

- **Lorentz force** (acts on plasma — MHD):

$$\mathbf{f}^{(B)} \;=\; \frac{1}{c}\mathbf{J}\times\mathbf{B} \tag{1.33}$$

This is another body force, which acts on the particles in a conducting fluid capable of carrying a current. The single-particle version of the Lorentz force involves the particle velocity. Expression (1.33) is the net force felt by the entire fluid which may consist (almost always *will* consist) of particles of both charge signs. If they move together the individual Lorentz forces will cancel out. Only when their motion leads to an electric current is there a net force on the composite fluid. We will return to this later.

- **Viscosity:** This is the most complex contribution, but is so small in general that it can be neglected in most astrophysical plasmas. When it is used in the astrophysical literature, it most commonly appears as the force density

$$\mathbf{f}^{(\mathrm{visc})} \;=\; \mu\nabla^2\mathbf{u} \quad . \tag{1.34}$$

where μ is the *viscosity*.[8] The viscosity μ is a property of the fluid and will in general depend on temperature. For concreteness, and possible future reference, a few examples of viscosity are listed in Table 1.1.

The Laplacian of a vector should be understood to work on *Cartesian* components[9] separately

$$\nabla^2\mathbf{u} \;=\; \hat{\mathbf{x}}\nabla^2 u_x \;+\; \hat{\mathbf{y}}\nabla^2 u_y \;+\; \hat{\mathbf{z}}\nabla^2 u_z \quad .$$

If you're of a mind to use curvilinear unit vectors you should use the vector identity

$$\nabla^2\mathbf{u} = \nabla(\nabla\cdot\mathbf{u}) - \nabla\times(\nabla\times\mathbf{u}) \quad ,$$

and do the tedious math.

Expression (1.34) is a simplified form of the true viscous force density. It is useful, convenient and relatively correct in most circumstances. The most accurate form of force density from viscosity takes the form

$$\mathbf{f}^{(\mathrm{visc})} \;=\; \nabla\cdot\underline{\underline{\sigma}} \quad , \quad f_i^{(\mathrm{visc})} \;=\; \frac{\partial\sigma_{ij}}{\partial x_j} \tag{1.35}$$

where σ_{ij} is the *viscous stress tensor*.[10] Since $\mathbf{f}^{(\mathrm{visc})}$ is a perfect divergence, this force

[8] It is sometimes called the *dynamic viscosity* to distinguish it from the *kinematic viscosity* $\nu = \mu/\rho$. To distinguish it from a non-traceless contribution to the stress tensor, called *bulk viscosity*, the full name for μ is the *shear viscosity*. Since bulk viscosity occurs only when $\gamma \neq 5/3$, we will never need to refer to it and the generic term *viscosity* will always refer to either μ or ν.

[9] It is, however, possible to use curvilinear *coordinates* to compute the Laplacians.

[10] Some authors (Choudhuri, 1998) use the negative of this tensor $\pi_{ij} = -\sigma_{ij}$, in order to emphasize its relation to pressure. Stress is a general term for a surface force — recall the *Maxwell stress tensor*. We focus here on stress arising from fluid velocity, called *viscous stress*, but the same form of force arises in *elasticity*. Then the stress is due to the local deformation of the fluid element, characterized by a *strain tensor*. Pressure contributes an isotropic, negative term, $-p\,\delta_{ij}$ to the total stress tensor — negative stress is by convection compressive (Landau & Lifshitz, 1986). This is the reason for also including the negative sign in the definition of π_{ij}.

viscosity ..... [units]	dynamic μ [g cm^{-1} sec^{-1}]	kinematic $\nu = \mu/\rho$ [cm^2 sec]
dry air[a]	1.8×10^{-4}	0.15
water[a]	1.0×10^{-2}	1.0×10^{-2}
gas of solid spheres [b]	$0.04\,(mk_{\rm B}T)^{1/2}/R^2$	$0.55\,(k_{\rm B}T/m)^{1/2}/4\pi R^2 n$
SAE 40 motor oil[c,d]	6.5 – 9.0	7.2 – 10
Peanut butter[a,d]	2.5×10^4	2.5×10^4
unmagnetized plasma[e]	$0.12\ T_6^{5/2}$	$7.3 \times 10^{13}\ T_6^{5/2}\, n_9^{-1}$
magnetized[e,f] plasma	$5 \times 10^{-14}\, T_6^{-1/2}\ n_9^2\, B_2^{-2}$	$27\ T_6^{-1/2}\ n_9\, B_2^{-2}$

[a] at T=20° C and standard atmospheric pressure.

[b] A gas of identical perfectly elastic spheres, radius R, mass m, at temperature T and number density n, is an idealized model of a diffuse ($nR^3 \ll 1$) neutral gas in which particles interact by physically bumping into one another (Lifshitz & Pitaevskii, 1981, §10).

[c] The designation10W-40 means an SAE rating of 10 ($\mu \sim 1\,{\rm g\,cm^{-2}\,s^{-2}}$) for cold starting (winter) and 40 at high temperatures.

[d] Some complex fluids like petroleum oil, shampoo or peanut butter transport momentum through long molecules (polymers) instead of random particle motion. This endows them with considerably higher viscosity.

[e] Properties of the fully-ionized Hydrogen plasma (trace elements will have little effect on viscosity) are denoted by $T_6 = T/10^6\,{\rm K}$ and $n_9 = n_e/10^9\,{\rm cm^{-3}}$, $B_2 = B/100$ G, which will all be roughly order one in the solar corona.

[f] For strong magnetization, $\Omega_i \tau_i \gg 1$, the viscous stress tensor is very complicated. The term most relevant to the simplified form in eq. (1.34) is the coefficient coupling rates of strain to stresses both perpendicular to the magnetic field. Coupling between terms parallel to the field are the same as for the unmagnetized case.

Table 1.1: Viscosities in some commonly encountered fluids.

is closely related to the pressure force: it acts on the surface of a fluid parcel

$$\mathbf{F}^{(\rm visc)} \;=\; \int_{\mathcal{V}} \mathbf{f}^{(\rm visc)}\, d^3x \;=\; \int_{\mathcal{V}} \nabla\cdot\underline{\underline{\sigma}}\, d^3x \;=\; \oint_{\partial\mathcal{V}} \underline{\underline{\sigma}}\cdot d\mathbf{a} \quad .$$

Like pressure, viscosity is not a body force, and no individual particles ever feel viscosity. While pressure always contributes a force *normal* to the surface, off-diagonal terms in the matrix σ_{ij} permit viscous stress *tangent* to the surface (probably your intuitive expectation for viscosity). The matrix $\underline{\underline{\sigma}}$ is defined to be *traceless* ($\sigma_{ii} = 0$) and *symmetric* ($\sigma_{ij} = \sigma_{ji}$).

The viscous stress tensor is proportional to first derivatives of the velocity[11] with a constant of proportionality,[12] μ,

$$\sigma_{ij} \;=\; \mu\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \tfrac{2}{3}\delta_{ij}\nabla\cdot\mathbf{u}\right) \quad . \tag{1.36}$$

[11] The tensor in parentheses, called the *rate of strain tensor*, is the only combination of first derivatives which is both symmetric and traceless.

[12] An astute and mathematically inclined reader might note that the most general linear relationship between two second-rank tensors, the rate of strain and the stress, is through a *fourth-rank* tensor. This worst-case scenario actually applies to some cases, such as when a strong magnetic field breaks the isotropy.

Equation (1.36) is a formidable expression,[13] seemingly unrelated to the simpler version of eq. (1.34). In almost every case where the viscosity might conceivably matter the flow is slow enough to be very sub-sonic and therefore incompressible: $\nabla \cdot \mathbf{u} = 0$. In this case the momentum equation contains the terms

$$\frac{\partial \mathbf{u}}{\partial t} \;=\; \text{stuff } \cdots \; + \frac{\mu}{\rho}\nabla^2\mathbf{u} \;=\; \text{stuff } \cdots \; + \nu\nabla^2\mathbf{u} \quad . \tag{1.37}$$

where the "stuff" will include terms proportional to $\nabla\mu$ when the viscosity has spatial variations. Expression (1.37), sketchy as it is, serves to show that viscosity leads to *velocity diffusion* with coefficient $\nu = \mu/\rho$, the kinematic viscosity in cm^2/sec.

The most common reason to include viscosity in astrophysical fluid equations at all is for the numerical stability contributed by diffusion. The simplified viscosity force, eq. (1.34), comes from assuming incompressibility and omitting terms proportional to $\nabla\mu$. This provides diffusion as effectively as the full expression, and much more clearly. In such cases the viscosity μ is assigned a value chosen for numerical stability rather than its physical value — which would be many orders of magnitude smaller. The physical value of μ is typically so small that it would have no noticable effect on the fluid's dynamics. So when viscosity is included its value it is assigned for stability and it is not important that the most accurate form of the force density be retained.

1E.2 Heating rates:

- **Adiabatic:** In many cases we will consider, fluid evolution will occur rapidly enough that irreversible processes are not effective so $\dot{Q} \simeq 0$. This means, according to eq. (1.14), that the specific entropy of any fluid element is preserved: $Ds/Dt = 0$. Using expression (1.12), or setting $\dot{Q} = 0$ in eq. (E), leads to a relation between pressure and density

$$\frac{D}{Dt}\left(\frac{p}{\rho^\gamma}\right) \;=\; 0 \quad . \tag{1.38}$$

  It is tempting to propose the solution $p = K\rho^\gamma$, where K is some constant. The problem is that K is constant for each fluid element ($DK/Dt = 0$) but it will probably not be the same constant for different fluid elements — it will vary in space. This raises the important distinction between a fluid which is *adiabatic* (s constant in time for any fluid element) and *isentropic* (s uniform in space at some time)

$$\begin{aligned} \text{adiabatic} \quad &: \quad \frac{Ds}{Dt} \;=\; 0 \\ \text{isentropic} \quad &: \quad \nabla s \;=\; 0 \quad . \end{aligned}$$

  In those particular cases which are *both* adiabatic and isentropic can we replace pressure with $p = K\rho^\gamma$ where K is a single constant parameter.

- **Radiation:** If electromagnetic radiation is absorbed or emitted by the fluid there will be a positive or negative contribution to $\dot{Q}$ respectively. This is by far the most

[13] A second term involving the bulk viscosity has been dropped since bulk viscosity vanishes for a gas of particles without internal degrees of freedom: $\gamma = 5/3$.

complicated contribution to fluid dynamics. Attempts to include a radiation field fully consistent with the fluid, called *radiative hydrodynamics*, are the most sophisticated astrophysical fluid computations today. We will not go down this daunting path.

The simplest, and therefore most common, treatments of radiative heating are either in the limit where photon mean free paths are very short, called *optically thick*, or very long, called *optically thin*. Stellar interiors are optically thick and photons are quickly absorbed and re-emitted causing them to undergo a random walk. The resulting heating is identical to thermal conduction and is discussed below.

Plasmas of low density, such as the solar corona, are optically thin. They absorb virtually no radiation and all they emit disappears with no real chance for re-absorption. Radiation is therefore a simple loss term ($\dot{Q}^{(\mathrm{rad})} < 0$). Its actual value depends on the rate of radiative emission by the particles composing the plasma. The primary loss at temperatures below about 10^7 K is from atomic transitions of partially ionized trace elements. At temperatures above this continuum emission mechanisms such as thermal Brehmsstrahlung (a.k.a. free-free emission) dominates (Rybicki & Lightman, 1979).

Regardless of the mechanism, virtually every emission is triggered by a collision with a single electron. One radiator collides with surrounding electrons at a rate proportional to their local number density, n_e. The number of radiators per unit volume, of any type, is also proportional to the number density of electrons. The constant of proportionality will involve the overall ionization level and the abundance of that type of radiator. In any event the radiation rate per unit volume will be proportional to n_e^2

$$\dot{Q}^{(\mathrm{rad})} \;=\; -\,n_e^2\Lambda(T;\mathcal{A}) \;=\; -\,\rho^2\Lambda(T;\mathcal{A})/m_p^2 \quad , \tag{1.39}$$

with a factor Λ, called the *radiative loss function*, incorporating all the types of radiators and the average energy per emission. It is notable that the quadratic pre-factor is the only density dependence[14] for radiative losses; the radiative loss function, Λ *does not* depend on density.

The radiative loss function depends entirely temperature since that determines the speed at which electrons collide, their cross sections for various radiative processes, the ionization states of all ionic species, and the spectrum of the overall emission. Due to the complications and uncertainties in quantifying these factors the literature contains different estimates for $\Lambda(T)$; a few are shown in fig. 1.1.

Different elements contribute to $\Lambda(T)$ at different temperatures so their relative abundances, designated schematically here by $\mathcal{A}$, are critical to the form of the loss function. Even though iron constitutes only one of every 30,000 ions in the solar atmosphere, it accounts for an overwhelming majority of the radiation from million degree plasmas (see fig. 1.1). Its coronal abundance can be enhanced in some circumstances (Feldman, 1992) leading to a significant increase in radiative losses. Such an enhancement, by a factor of three to the iron abundance, produces the enhancement in the MKB curve (dash-dot) over the CCJA (solid) in the figure.

[14] We substituted $n_e = \rho/m_p$, where m_p is the mass of the proton. For cases other than a fully ionized Hydrogen plasma there would be a factor, close to unity, incorporated into Λ.

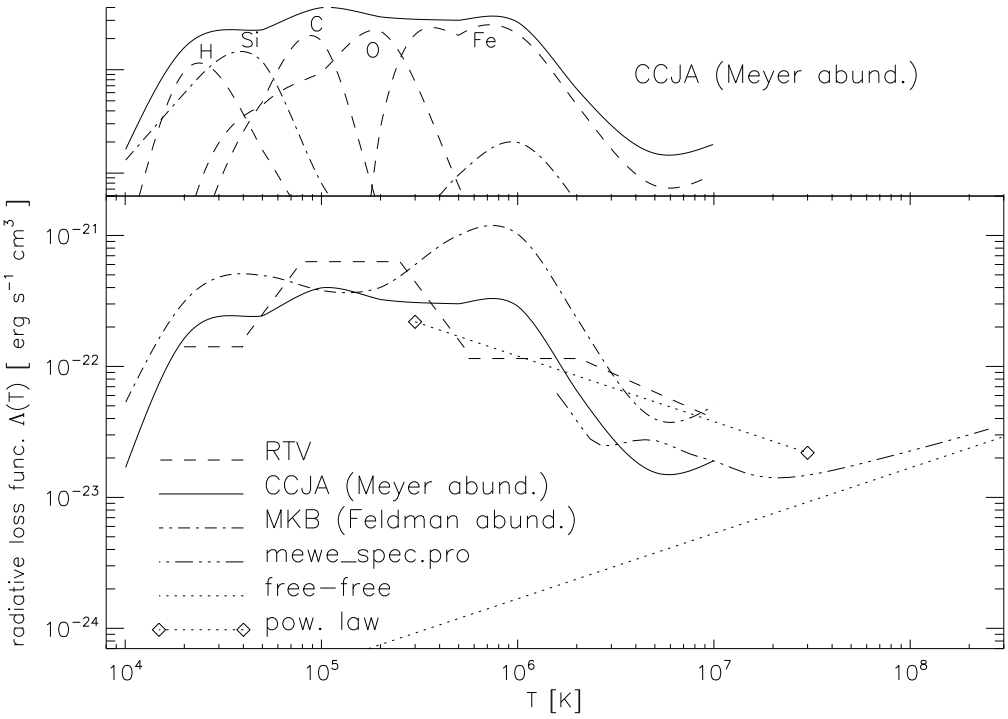

Figure 1.1: The radiative loss function $\Lambda(T)$ as reported in the astrophysical literature. The bottom panel shows approximations published by Rosner et al. (1978, RTV, dashed), Cook et al. (1989, CCJA, solid) and Martens et al. (2000, MKB, dash-dot). The last two use the same emission spectra but assume different elemental abundances — from Meyer (1985) and Feldman (1992) respectively. High temperature values (dash-triple-dot) are found from a Solarsoft program and from the formula for thermal Brehmsstrahlung (free-free emission, dotted). A second dotted line, between diamonds, shows a power-law approximation, $\Lambda(T) = 1.2 \times 10^{-19}\, T^{-1/2}$. The top panel shows the contribution from each of the five elements making major contributions to the CCJA curve.

- **Thermal conduction:** This is a term which moves heat around without adding or removing any — it is *conservative*. The most general form of such a contribution is the divergence of a vector

$$\dot{Q}^{(\mathrm{cond})} \;=\; -\,\nabla\cdot\mathbf{q} \quad . \tag{1.40}$$

The net contribution of this heating to a fluid parcel, as in eq. (1.27),

$$\int_{\mathcal{V}} \dot{Q}^{(\mathrm{cond})}\, d^3x \;=\; -\oint_{\partial\mathcal{V}} \mathbf{q}\cdot d\mathbf{a} \quad , \tag{1.41}$$

involves only flux through the boundary. The vector $\mathbf{q}$ is therefore called the conductive heat flux and has units $\mathrm{erg\,cm^{-2}\,s^{-1}}$.

The second law of thermodynamics dictates that $\mathbf{q}$ be proportional to the temperature gradient,

$$\mathbf{q} \;=\; -\,\kappa\nabla T \tag{1.42}$$

with a constant κ called the *thermal conductivity*[15] which must be non-negative: heat flows from hot to cold. Using this in (1.40) gives the heating function

$$\dot{Q}^{(\mathrm{cond})} \;=\; \nabla\cdot(\kappa\nabla T) \quad . \tag{1.43}$$

Using this in the temperature evolution, eq. (1.15), gives a form

$$\frac{\partial T}{\partial t} \;=\; \text{stuff}\;\cdots\;+\frac{\gamma-1}{k_{\mathrm{B}}(\rho/m)}\,\kappa\,\nabla^2 T \;=\; \text{stuff}\;\cdots\;+\;\tilde{\kappa}\nabla^2 T \quad , \tag{1.44}$$

where the "stuff" can include a term proportional to $\nabla\kappa$ when the conductivity is spatially varying. This illustrates that thermal conductivity produces a diffusion of temperature. The coefficient $\tilde{\kappa}$ is a diffusion coefficient, with units of $\mathrm{cm}^2/\mathrm{sec}$, sometimes called *thermal diffusivity*, and is closely related to thermal conductivity. Indeed, the so-called *heat equation.*, in which the "stuff" term entirely vanishes,[16] is the prototypical diffusion equation.

The heat flux $\mathbf{q}$ can take the form of energy carried by particles, such as electrons, or by photons. Random walks produce diffusive transport, so both cases lead to the same basic form of heating, eq. (1.43), which we collectively term conduction. In the air around us, for example, heat is transported by randomly moving particles: molecules. The rate of transport naturally depends on the speed of the random walkers, so κ will depend on temperature. When the step size is independent of temperature, as for solid spheres or neutral molecules, then $\kappa\sim T^{1/2}$. Charged particles in a plasma interact electrostatically. This results in a temperature-dependent mean-free-path[17] and a conductivity, $\kappa\sim T^{5/2}$.

Heat is carried by randomly-walking photons in some optically thick fluids, for example inside stars. The sequential emission, absorption, or scattering of photons constitutes a random walk. The step size, $\ell_{\mathrm{mfp},\gamma}$, is assumed to be far smaller than the length scales over which fluid properties vary in order to qualify as being optically thick. This means photons are being emitted and absorbed by matter of roughly the same temperature, and the radiation field will come into thermal equilibrium with the matter. In other words it will be a *blackbody spectrum* characterized by that temperature: black body spectra occur most often under optically thick conditions.

To see how the black body spectrum of photons produces thermal conduction, consider a case where the temperature varies only along one direction, say x. The radiation crossing the $x=0$ plane going rightward will have originated from the material to the left, $x<0$. The radiative flux it carries will be characteristic of the temperature of the material about one mean free path left of the plane: $\sigma_{\mathrm{SB}}T^4(-\ell_{\mathrm{mfp},\gamma})$, where σ_{SB} is the Stefan-Boltzmann constant. There is also a flux of photons leftward originating from $x=+\ell_{\mathrm{mfp},\gamma}$. The *net* flux crossing the $x=0$ plan is the difference between the rightward and leftward fluxes

$$q_x \;=\; \sigma_{\mathrm{SB}}T^4(-\ell_{\mathrm{mfp},\gamma})-\sigma_{\mathrm{SB}}T^4(+\ell_{\mathrm{mfp},\gamma}) \;\simeq\; -\,8\sigma_{\mathrm{SB}}T^3\ell_{\mathrm{mfp},\gamma}\,\frac{\partial T}{\partial x} \quad . \tag{1.45}$$

[15] The most general linear relation between the two vectors is a matrix. In some cases the scalar κ is replaced by a symmetric matrix with non-negative eigenvalues (positive semi-definite).

[16] . . . as it would, for example, in a static fluid with uniform conductivity

[17] This dependence is derived with more rigor in Chapter 15.

Higher order terms in the Taylor expansion are dropped because $T(x)$ varies on scales $\gg \ell_{\rm mfp,\gamma}$. The final expression exactly matches the form of eq. (1.42) with a conductivity

$$\kappa^{(\rm rad)} \;=\; 8\sigma_{\rm SB} T^3 \ell_{\rm mfp,\gamma} \quad . \tag{1.46}$$

This is the radiative conductivity, representing the heat carried by photons. At high enough temperatures radiative conductivity dominates electron conductivity, as shown in fig. 1.2. This is the case throughout stellar interiors.

Small-scale random fluid motions, called *turbulence*, can move heat in a random walk analogous to the processes just described. The fluid motion is itself governed by fluid equations, but it is sometimes useful to consider it akin to a "microscopic process" acting on large-scale motions obeying modified fluid equations. The small-scale turbulence diffuses heat down the large-scale temperature gradients according to eq. (1.43), except with a conductivity κ due to the turbulent motions rather than electron or photon scattering. The modified coefficient is called the *turbulent conductivity*. While turbulent motions are relatively slow (especially compared to photon speeds) they produce much larger random steps and can therefore dominate the conductivity. This is the case in the turbulent outer envelope of stars, as shown in fig. 1.2.

- **Viscous heating:** When the viscous force density of eq. (1.35) is included in the momentum equation it will sap kinetic energy from the fluid (see eq. [1.25]). In order to conserve total energy, eq. (1.27), this kinetic energy loss must be offset by a heating

$$\dot{Q}^{(\rm visc)} \;=\; \sigma_{ij} \frac{\partial u_i}{\partial x_j} \quad , \tag{1.47}$$

where σ_{ij} is the viscous stress tensor and repeated indices are summed over. In verifying this assertion it is necessary to integrate by parts the work from eq. (1.27)

$$\int_{\mathcal{V}} \mathbf{u}\cdot\mathbf{f}^{(\rm visc)}\, d^3x \;=\; \int_{\mathcal{V}} \mathbf{u}\cdot(\nabla\cdot\underline{\underline{\sigma}})\, d^3x \;=\; \oint_{\partial\mathcal{V}} \mathbf{u}\cdot\underline{\underline{\sigma}}\cdot d\mathbf{a} \;-\; \int_{\mathcal{V}} \underline{\underline{\sigma}} : \nabla\mathbf{u}\, d^3x \quad .$$

The final integrand is the same as (1.47) and will thus be cancelled by the proposed viscous heating. The surface term is work done on the fluid parcel by external fluid via viscous stressing. It is a surface contribution, similar to that from pressure (the first term of the rhs of eq. [1.27]), hinting once more at the underlying connection between pressure and viscosity.

The symmetry and traceless-ness of σ_{ij} allows eq. (1.47) to be rewritten

$$\dot{Q}^{(\rm visc)} \;=\; \sigma_{ij}\left(\tfrac{1}{2}\frac{\partial u_i}{\partial x_j} + \tfrac{1}{2}\frac{\partial u_j}{\partial x_i} + \alpha\delta_{ij}\right) \quad ,$$

for any constant α (contraction with the Kronecker δ-function is equivalent to the trace of the matrix, which produces zero in the present case). Choosing $\alpha = -(\nabla\cdot\mathbf{u})/3$ makes the tensor in parentheses traceless as well as symmetric, akin to the one it is contracted with. The full form of viscous stress, eq. (1.36) results in heating

$$\dot{Q}^{(\rm visc)} \;=\; \tfrac{1}{2}\mu \left\| \frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} - \tfrac{2}{3}\delta_{ij}\nabla\cdot\mathbf{u} \right\|^2 \quad , \tag{1.48}$$

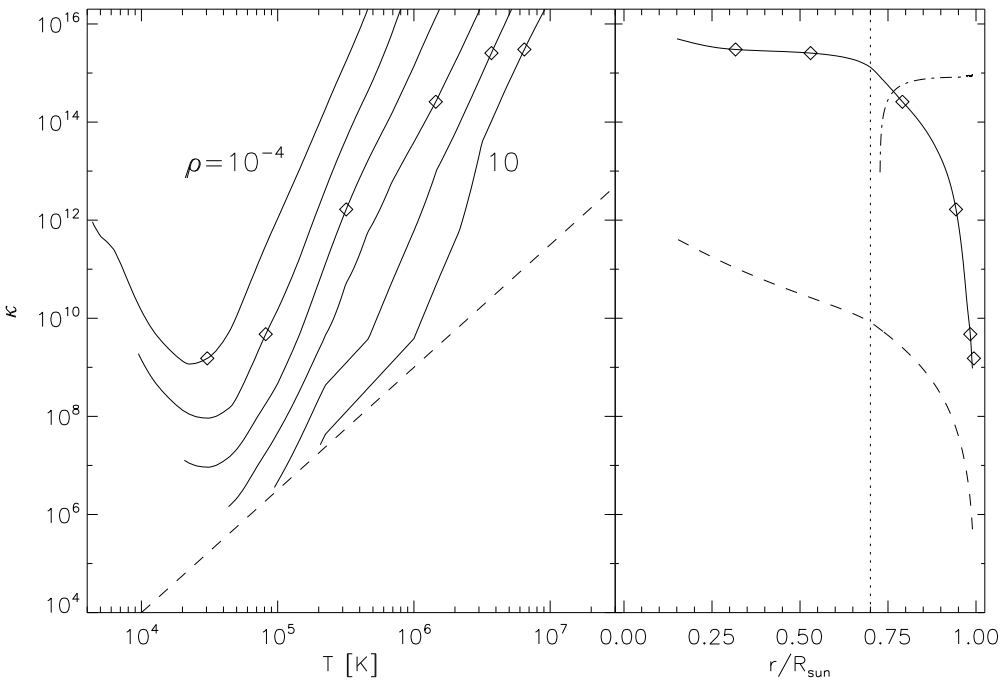

Figure 1.2: Thermal conductivity, κ ($\mathrm{erg\,cm^{-1}\,s^{-1}\,K^{-1}}$) at various temperatures and densities. The left panel shows electron conductivity (dashed), independent of density, and radiative conductivity (solid) at densities of $\rho = 10^{-4}$, 10^{-3}, 10^{-2}, 0.1, 1, $10\,\mathrm{g/cm^3}$, from left to right. The right panel shows the two versus the radius throughout the solar interior. The broken curve shows the turbulent conductivity within the convection zone (vertical dotted line).

where

$$\|M_{ij}\|^2 \;=\; M_{ij}M_{ij} \quad ,$$

is the norm of matrix M_{ij}. Since the norm is never negative, we see that choosing a stress tensor of the form eq. (1.36) assures $Ds/Dt \geq 0$ when viscosity is the only source of heating, as demanded by the second law of thermodynamics.

When the simpler form of viscosity, eq. (1.34), is used the same line of reasoning demands that

$$\dot{Q}^{(\mathrm{visc})} \;=\; \mu \left\| \frac{\partial u_i}{\partial x_j} \right\|^2 \;=\; \mu \frac{\partial u_i}{\partial x_j}\frac{\partial u_i}{\partial x_j} \quad . \tag{1.49}$$

1F How to scale an equation

The equations of fluid dynamics include more terms than many other PDEs encountered in Physics. A typical form of the momentum equation, eq. (M), for example,

$$\underbrace{\rho\frac{\partial \mathbf{u}}{\partial t}}_{M.1} + \underbrace{\rho(\mathbf{u}\cdot\nabla)\mathbf{u}}_{M.2} = -\underbrace{\nabla p}_{M.3} + \underbrace{\rho\mathbf{g}}_{M.4} + \underbrace{\rho\nu\nabla^2\mathbf{u}}_{M.5} \quad , \tag{1.50}$$

includes five different terms identified here as $M.1$ — $M.5$. Compare this to Faraday's law which expresses the equality of two terms. The large number of terms provides the opportunity for making many different approximations which can sometimes seem perplexing — terms are neglected in some circumstances and not others.

It is generally true for an equation of many terms that two of the terms are larger than the rest and that the equation expresses an approximate balance between them; the remaining terms do little more than nudge this balance. If we knew ahead of time which terms would be the dominant ones we could simplify our lives considerably by neglecting the "nudging" terms. In addition to making the equation easier to solve, this also helps us understand the basic nature of the Physics we face — there is a balance between two effects. This approach is adopted very often in fluid mechanics leading to the sense that the fluid equations are a kind of moving target.

The process of assessing which terms are dominant in a particular equation is an extremely useful skill in Physics. The trick is to imagine that one has solved all the equations and has found functions $\rho(\mathbf{x},t)$, $u_x(\mathbf{x},t)$ etc., which satisfy them. Each function takes on values which have a *characteristic* magnitude which we can denote with a *single* characteristic value; we will denote the characteristic value with the same variable — say ρ. Since we do not know the solution (yet) we do not know the exact value, but we should be able to predict ahead of time whether $\rho \sim 1\,\mathrm{g/cm^3}$, like water, or $\rho \sim 10^{-23}\,\mathrm{g/cm^3}$, typical of the solar wind. This characteristic value is typically set by initial conditions or boundary conditions, which you would know ahead of time. This is a game played to "order-of-magnitude" accuracy, so one should not fret over factors of two or three in the characteristic value.

The functions composing the solution will also vary in space and time with characteristic scales, which we denote L and τ respectively. These are also set by the initial conditions and boundary conditions and should be known ahead of time — at least at the "order-of-magnitude" level. The flow of air across your car will vary on a spatial scale of approximately $L \sim 1$ m, characteristic of the features on your car — say the windshield. A derivative with respect to space or time will thus have a scale given by a ratio,

$$\nabla\rho \sim \frac{\rho}{L} \quad , \quad \frac{\partial \mathbf{u}}{\partial t} \sim \frac{u}{\tau} \quad .$$

The $\sim$ is used here to indicate that the positive scalar value on the right is "characteristic" of the quantity on the left; in both examples these are vectors for which each component should have the same characteristic scale. This can be done for each term in a differential equation to estimate its approximate size.

It was stated above that viscosity is typically a small effect in Astrophysical fluids. This statement really means that term $M.5$ is much, much smaller than most other terms in eq.

(1.50). To see this we estimate the magnitude of its ratio with $M.2$,

$$\frac{M.5}{M.2} \;=\; \frac{\nu\nabla^2\mathbf{u}}{(\mathbf{u}\cdot\nabla)\mathbf{u}} \;\sim\; \frac{\nu u/L^2}{u^2/L} \;=\; \frac{\nu}{uL} \quad . \tag{1.51}$$

This shows that the importance of viscous stress to the momentum equation depends on the size of the kinematic viscosity, ν, compared to the product of the characteristic speed, u, and size L. When I ride my bike, for example, the air flows around me with characteristic speed $u \sim 300$ cm/sec and scale $L \sim 10$ cm (my arms, for example) giving $uL \sim 3\times10^3\,\mathrm{cm}^2/\mathrm{sec}$. The viscosity of air, $\nu = 0.15\,\mathrm{cm}^2/\mathrm{sec}$ (see Table1.1), is *twenty-thousand* times smaller than this and so cannot possibly affect me as I ride.

This example is typical of such scalings in several respects: the ratios are larger or smaller than unity by many orders of magnitude. For this reason it did not matter whether I chose $L \sim 10$ cm or $L \sim 100$ cm (my torso) — viscosity is negligible no matter how I construct my estimate. Only when ratios are so very large or very small can we feel justified completely neglecting a term in the governing equation. If the ratio is close to one — even within two orders of magnitude — you may need to retain both terms: they might each be important in some part of the solution.

The ratio of any two terms in an equation will yield a dimensionless number characterizing the problem. Such dimensionless numbers are important in Physics and have generally been assigned names honoring Physicists who worked on that equation. The ratio above, actually its reciprocal uL/ν, is known as the *Reynolds number* after British physicist Osborne Reynolds, and is typically denoted Re. The huge list of such named dimensionless ratios can seem overwhelming. Do not worry about them, it is far more important to understand the significance of a particular ratio than to know its name. The ratio uL/ν, for example, expresses the relative importance of advection (term $M.2$) to viscosity (term $M.5$) in the momentum equation.

We have just learned that flow around a bicyclist is characterized by $Re \sim 10^4$. The granulation flow at the solar surface, $u \sim 10^5$ cm/sec, $L \sim 10^8$ cm and $\nu \sim 10^3\,\mathrm{cm}^2/\mathrm{sec}$, has $Re \sim 10^{10}$. In either case viscosity is far too small to play a significant role and term $M.5$ may be discarded from the equation.[18]

There are also circumstances where $Re \ll 1$ and we can neglect $M.2$ in favor of $M.5$. When a dust grain, $L \sim 10^{-3}$ cm, falls though still air, $u \sim 10^{-1}$ cm/sec, we find $Re \sim 10^{-3}$ — viscosity is dominant and inertia is negligible. Low Reynolds number flow, known as *Stokes flow*, is entirely different from the fluid world we experience as humans. Table 1.1 shows that it is accessible on large scales ($L \sim 100$ cm, $u \sim 1$ cm/sec) when the viscosity is correspondingly large: $Re \sim 10^{-2}$ in peanut butter. So the life of a dust grain in air resembles what our life would be like if we lived in peanut butter. Because $M.2$ is the most non-linear term in the system, Stokes flow is much easier to treat mathematically than flow at high Reynolds number.

Terms $M.1$ and $M.2$ are both part of the advective derivative so we might expect them

[18] There is an important mathematical consequence of doing this. Discarding the $\nu\nabla^2\mathbf{u}$ term changes eq. (1.50) from a parabolic equation to a hyperbolic equation. This changes the nature of the boundary conditions necessary for a unique solution. There is a very long and interesting literature concerning this which we do not have the luxury of exploring here.

to be of comparable size. Their ratio has magnitude

$$\frac{M.2}{M.1} \;=\; \frac{(\mathbf{u}\cdot\nabla)\mathbf{u}}{\partial\mathbf{u}/\partial t} \;\sim\; \frac{u\tau}{L} \;=\; \frac{\tau}{L/u} \quad . \tag{1.52}$$

L/u is the time taken for a fluid element to traverse the characteristic scale L; we call this the *advective time scale*. Terms $M.2$ and $M.1$ are equally important when the flow changes on the advective time scale. We can, however, neglect $M.1$ for relatively *steady flows* — flows that appear steady over advective times.

Ratios of all other terms in eq. (1.50) can be easily found:

$$\begin{aligned}
\frac{M.2}{M.5}: &\quad \frac{uL}{\nu} \;\equiv\; Re && \text{Reynolds number}\\
\frac{M.4}{M.3}: &\quad \frac{L\rho g}{p} \;=\; \frac{L}{H_p} && H_p = \frac{p}{g\rho} = \text{Pressure scale height}\\
\frac{M.2}{M.3}: &\quad \frac{u^2}{p/\rho} \;\sim\; \left(\frac{u}{c_s}\right)^2 \equiv M^2 && M = \text{Mach number} \quad .
\end{aligned}$$

The second one, $M.4/M.3$, shows that gravity is only important on scales comparable to, or larger than, the pressure scale height H_p. The Earth's atmosphere, $\rho \simeq 10^{-3}\,\mathrm{gm/cm^3}$, $p \simeq 10^6\,\mathrm{gm\,cm^{-1}\,sec^{-2}}$, and $g \sim 10^3\,\mathrm{cm/sec^2}$ has a scale height $H_p \simeq 10^6\,\mathrm{cm} = 10$ km. Any flow with smaller scale, such as wind around a building, can be solved without including gravity.

From the considerations above we see that flow of air around human-scale objects ($10^{-2}\,\mathrm{cm} \ll L \ll 10^4$ cm) can be treated without including gravity or viscosity: $M.4$ and $M.5$. The resulting momentum equation

$$\frac{\partial\mathbf{u}}{\partial t} \;+\; (\mathbf{u}\cdot\nabla)\mathbf{u} \;=\; -\,\frac{\nabla p}{\rho} \quad , \tag{1.53}$$

is a common starting place for fluid dynamics in familiar settings. This is obviously simpler than the five terms in eq. (1.50).

Treatment of low Mach number, $M^2 \ll 1$, is a bit more subtle than the previous cases. At first blush it would seem that we could discard $M.2$ (as well as $M.1$ which has roughly the same size) in favor of $M.3$. This leaves the extremely simple equation

$$\nabla p \;=\; 0 \quad . \tag{1.54}$$

whose only solution is uniform pressure: $p(\mathbf{x}) = p_0$. What this actually tells us is that subsonic flow cannot make significant changes to the ambient pressure. The result is known as *incompressible* flow: flow at very small Mach number is almost always incompressible. If we subtract the ambient atmospheric pressure from the solution $p(\mathbf{x}, t)$, we find that all three terms in eq. (1.53) must have comparable magnitude. In other words the pressure difference[19] $(p - p_0)$ must have magnitude roughly matching the so-called *ram pressure* of the flow, $\rho_0 u^2$. The relative change in the absolute pressure is $\sim M^2 \ll 1$. The relative density change will be similar and can thus be ignored: density is relatively unchanged by sub-sonic flow.

[19] known as *gauge pressure*.

1G Incompressible fluid dynamics

For the reasons just described, subsonic flows are often treated by replacing the energy equation (E) by the constraint of incompressibility: $\nabla \cdot \mathbf{u} = 0$. The reason is that subsonic flow lacks the "oomph" necessary to change the pressure or density significantly, and thus both are carried around with approximately their initial values. Consistency with eq. (E) requires that $\nabla \cdot \mathbf{u} \simeq 0$ so that $Dp/Dt = 0$ (absent sources of heat such as flame).

Incompressibility is an approximation whose validity depends on the smallness (compared to unity) of the square of the Mach number — the ratio of velocity to sound-speed. As that becomes vanishingly small it becomes increasingly difficult[20] to use (E) for finding pressure, and increasingly easy (and accurate) to use incompressibility.

It was remarked after eq. (C-L) that incompressible flow does not change the density: if $\rho(\mathbf{x}, 0) = \rho_0$ initially, then $\rho = \rho_0$ forever after and a continuity equation is no longer necessary — this is the first advantage.

The complete set of **incompressible fluid equations** is therefore

$$\rho_0 \frac{D\mathbf{u}}{Dt} \;=\; \rho_0 \frac{\partial \mathbf{u}}{\partial t} \;+\; \rho_0 (\mathbf{u}\cdot\nabla)\mathbf{u} \;=\; -\nabla p \;+\; \mathbf{f} \qquad \text{(M-I) incompressible momentum}$$

$$\nabla \cdot \mathbf{u} \;=\; 0 \qquad \text{(IC) incompressibility}$$

The absence of an energy equation to time-advance the pressure p is puzzling at first. From a mathematical perspective, however, introducing another equation would over-constrain the system. Equations (M-I) and (IC), comprise four equations for four unknowns, u_x, u_y, u_z and p, so they are complete as they stand. Incompressibility is a constraint on the velocity field meaning that only two of the three functions are truly free to vary. Pressure plays the role of *Lagrange multiplier* used to enforce incompressibility.

Following this logic, the force density $-\nabla p$ is therefore a "force of constraint" required to maintain the constraint of incompressibility. Taking the divergence of (M-I) and using (IC) gives an elliptic equation for pressure

$$\nabla^2 p \;=\; \nabla \cdot \mathbf{f} \;-\; \rho_0 \nabla \cdot [\,(\mathbf{u}\cdot\nabla)\mathbf{u}\,] \quad . \tag{1.55}$$

Instead of an energy equation, eq. (1.55) must be solved at every time in order to determine the pressure. This means that *in incompressible fluid dynamics pressure is not related to temperature or density but to the global velocity field.*

Since the pressure p is required only to maintain a constraint (in the manner of a Lagrange multiplier) it is possible to find a formulation from which it is absent. Taking the curl of eq. (M-I) will formally eliminate it. In so doing it creates a new variable $\boldsymbol{\Omega} = \nabla \times \mathbf{u}$, called the *vorticity*. Making liberal use of vector identities and the fact that $\nabla \rho_0 = 0$ leads to alternative equations for incompressible fluid dynamics

$$\frac{\partial \boldsymbol{\Omega}}{\partial t} - \nabla \times (\mathbf{u} \times \boldsymbol{\Omega}) \;=\; \frac{D\boldsymbol{\Omega}}{Dt} - (\boldsymbol{\Omega}\cdot\nabla)\mathbf{u} \;=\; \frac{1}{\rho_0}\nabla \times \mathbf{f} \qquad \text{(V-I) vorticity}$$

[20] One difficulty stems from following dynamics on two very different time scales. Incompressible motions occur on advective time scales $\sim L/u$, where u is the typical flow velocity. Compressible motions occur on faster time scales, $\sim L/c_s$, where c_s is the speed of sound. When $M = u/c_s \ll 1$ these time scales are very different and it is difficult to "follow" the fast motions long enough for the slow evolution to unfold.

$$\nabla \times \mathbf{u} \;=\; \boldsymbol{\Omega} \quad , \quad \nabla \cdot \mathbf{u} \;=\; 0 \qquad \text{(DV) def. of vorticity}$$

In this formulation one advances the vorticity field using eq. (V-I). At each time the velocity field is found from "uncurling" the vorticity field,[21] according to eq. (DV). It is never necessary to know the pressure.

If this seems complicated or counter-intuitive... it is. Incompressible fluid dynamics is a vast and rich subject all unto itself. In many respects it is more mathematically challenging than the compressible version (C), (M), and (E), sometimes called *gas dynamics*. Incompressible motion is tackled using tools, vorticity and stream functions, very different from those in gas dynamics. Unfortunately it is a subject we cannot do justice to. Moreover, astrophysical flows tend to approach super-sonic speeds more commonly than do terrestrial air flows. This makes incompressibility less useful in general. We will, however, appeal to it on occasion.

[21] The equation is formally identical to Ampère's law and can be solved using a Biot-Savart integral or any other technique familiar from undergraduate magnetostatics.

Chapter 2

Hydrostatic Equilibria

2A Hydrostatic Atmospheres

An equilibrium is a fluid state satisfying the time-independent version of the fluid equations (C), (M) and (E). The simplest version of an equilibrium is the *static equilibrium*, in which the fluid itself is at rest: $\mathbf{u} = 0$. Setting to zero all terms in (C), (M) and (E) with either a time derivative or a velocity leaves only two equations (the continuity equation is trivial for a static equilibrium)

$$\frac{1}{\rho}\nabla p - \mathbf{g} \;=\; \frac{1}{\rho}\nabla p + \nabla\Psi \;=\; 0 \quad , \tag{2.1}$$

$$\dot{Q} \;=\; 0 \quad , \tag{2.2}$$

where we have used the gravitational potential Ψ to write the gravitational acceleration, $\mathbf{g} = -\nabla\Psi$. The second equation, $\dot{Q} = 0$, requires that heating and cooling sources, whatever they might be, balance at every point in the fluid lest the pressure at some point begin to increase or decrease. For an ideal fluid, which lacks heating and cooling, this condition is trivial. In other cases it turns out to be very complex indeed.

2A.1 Pressure-density relation

Equation (2.1), known as the condition of *hydrostatic balance*, expresses the requirement that the upward pressure gradient exactly balance the downward gravitational force at every point. Taking the curl of the expression shows that satisfying this condition requires the pressure gradient and density gradient to be parallel

$$\nabla \times \left(\frac{1}{\rho}\nabla p\right) \;=\; -\,\frac{1}{\rho^2}\,\nabla\rho \times \nabla p \;=\; 0 \quad . \tag{2.3}$$

Except in the very special case that $\nabla\rho = 0$ everywhere, the parallel gradient requirement means that one of the fields, say density, must depend on position only through some kind of dependence on pressure — density must be a function of pressure

$$\rho(\mathbf{x}) \;=\; \rho\Big[\,p(\mathbf{x})\,\Big] \quad . \tag{2.4}$$

Hydrostatic balance requires there be some density-pressure relationship, but it says nothing about what that relationship might be. It is after all, one equation for two unknowns, ρ and p, so we must look to an independent equation. Nor can we hope some kind of thermodynamic equation-of-state will provide this relationship. An equation of state is a relation between *three* thermodynamic variables, ρ, p and T, so it will provide a new equation *and* a new unknown — no real help. The new relationship must either come from the second equilibrium equation, $\dot{Q} = 0$, or it must depend on the details of how the atmosphere was prepared — say the initial conditions and time evolution leading up to it.

A common example of the first occurs when thermal conductivity is very large (i.e. $\tilde{\kappa} \gg c_s L$, for some length scale L). In this case $\dot{Q} = 0$ requires that $\nabla T \simeq 0$ in order that $\tilde{\kappa}\nabla T$ be comparable to sources and sinks. This means that $T(\mathbf{x}) = T_0$ is constant, a case known as as a *isothermal* atmosphere. For an isothermal atmosphere the dependence in eq. (2.4) is a simple linear one

$$\rho(p) \;=\; \frac{m}{k_{\rm B}T_0}\,p \;=\; \frac{p}{c_{\rm s}^2} \quad , \qquad \text{isothermal} \quad , \tag{2.5}$$

where m is the mean mass of a particle, and $c_s = \sqrt{k_{\rm B}T_0/m}$ is the isothermal sound speed.[1]

An example of the second case would be one where every fluid element evolves adiabatically and therefore retains the value of entropy s it first began with. If the fluid settles into equilibrium from a state where every fluid element started with the same entropy s_0, then every fluid element in the final equilibrium will have the same entropy as every other fluid element: the fluid is *isentropic*. In that case, the dependence in eq. (2.4) is a power-law

$$\rho(p) \;=\; \exp\left[\frac{(\gamma-1)s_0}{\gamma k_{\rm B}/m}\right] p^{1/\gamma} \quad , \qquad \text{isentropic} \quad . \tag{2.6}$$

These two may generalized into a kind of *ad hoc* relationship, called *polytropic*

$$\rho(p) \;=\; Kp^{1/\Gamma} \quad , \quad \text{polytropic} \quad , \tag{2.7}$$

where K and Γ are constants. The case $\Gamma = \gamma$ is isentropic and $\Gamma = 1$ is isothermal. A polytrope is often introduced to produce a complete set of solutions, parameterized by the free parameter Γ, which is swept over a range of values, not just γ and 1.

2A.2 Enthalpy

Regardless of what functional form appears in relationship (2.4), the equation of hydrostatic balance, eq. (2.1), is solved immediately by introducing the integral of its inverse. This indefinite integral

$$w(p) \;=\; \int^p \frac{dp}{\rho(p)} \quad , \tag{2.8}$$

will be called the *enthalpy*, although it corresponds to the thermodynamic quantity of that name only in the isentropic case. The integral is readily performed for the simple functions

[1] We return in Chapter 4 to show that this is the phase speed for certain perturbations to the fluid. In the mean time just consider this a quantity with units of cm/sec, defined by this expression.

discussed above

$$w = \begin{cases} \dfrac{\Gamma}{\Gamma-1}\dfrac{p}{\rho} = \dfrac{\Gamma p_0}{\rho_0(\Gamma-1)}\left(\dfrac{\rho}{\rho_0}\right)^{\Gamma-1} = \dfrac{\Gamma p_0}{\rho_0(\Gamma-1)}\left(\dfrac{p}{p_0}\right)^{(\Gamma-1)/\Gamma} & , \quad \text{polytropic} \\ c_s^2 \ln(\rho/\rho_0) \;=\; (k_{\rm B}T/m)\,\ln(p/p_0) & , \quad \text{isothermal} \end{cases} \tag{2.9}$$

where ρ_0 and p_0 are values assumed at some reference position. It is necessary to distinguish the isothermal case because the general polytropic integral becomes a logarithm as $\Gamma \to 1$.[2]

The utility of the enthalpy in solving the hydrostatic balance equation is readily appreciated from its gradient

$$\nabla w \;=\; \frac{dw}{dp}\nabla p \;=\; \frac{1}{\rho}\nabla p \quad , \tag{2.10}$$

after using the chain rule and eq. (2.8). Using this in eq. (2.1) leads quickly to the solution

$$w(\rho) + \Psi(\mathbf{x}) \;=\; \text{const.} \;=\; w(\rho_0) + \Psi_0 \quad . \tag{2.11}$$

The density structure of the equilibrium atmosphere, $\rho(\mathbf{x})$, can then be found by simply inverting the function $w(\rho)$. One attraction of the polytropic relationship is that it leads to a power-law enthalpy function which can be easily inverted. Any more complicated relationship (i.e. more realistic) would undoubtedly not be readily invertible.

2A.3 Atmospheric structures

In the isothermal case, where $w = c_s^2 \ln(\rho/\rho_0)$ and $w(\rho_0) = 0$, the solution to (2.11) is an exponential hydrostatic density

$$\rho_{\rm hs}(\mathbf{x}) \;=\; \rho_0 \,\exp\left[-\frac{\Psi(\mathbf{x}) - \Psi_0}{c_s^2}\right] \quad . \tag{2.12}$$

A planar approximation valid close to the surface, $\Psi = gz$, gives the classic exponential atmosphere

$$\rho(z) \;=\; \rho_0 e^{-z/H_p} \quad , \tag{2.13}$$

where $H_p = c_s^2/g = p/\rho g$ is the scale height.

In the broader perspective the gravitational potential is spherical, $\Psi = -GM/r$, and $\Psi_0 = -GM/R = -v_{\rm esc,o}^2/2$ where R is the stellar radius and $v_{\rm esc,o} = \sqrt{2GM/R}$ is the escape speed from there. In terms of these the density profile is

$$\rho_{\rm hs}(r) \;=\; \rho_0 \,\exp\left[-\frac{v_{\rm esc,o}^2}{2c_s^2}\,\left(1 - \frac{R}{r}\right)\right] \quad . \tag{2.14}$$

This expression returns eq. (2.13) for $z = r - R \ll R$ after noting that $v_{\rm esc,o}^2/c_s^2 = 2R/H_p$.

[2] This can be recovered from the general expression using the fact that $x^\epsilon/\epsilon \to \epsilon^{-1} + \ln x$ as $\epsilon \to 0$. This returns the isotropic expression plus a (divergent) constant term which can be discarded since only derivatives or differences of enthalpy will be relevant.

A notable, and surprising, feature of the isothermal, hydrostatic density in eq. (2.14) is that it remains finite even at distances arbitrarily far from the stellar surface. Taking $r \to \infty$ in the expression yields

$$\rho_\infty = \rho_0 \, \exp\left(-\frac{v_{\rm esc,o}^2}{2c_s^2}\right) \;=\; \rho_0 \, \exp\left(-\frac{T_{\rm esc,o}}{2T_0}\right) \quad , \tag{2.15}$$

where

$$T_{\rm esc,o} = \frac{2GM(m_p/2)}{Rk_{\rm B}} \tag{2.16}$$

is the temperature of a gas (average particle mass $m_p/2$) whose thermal velocity equals the photospheric escape velocity; for the Sun $T_{\rm esc,o} = 2.3 \times 10^7 \, K$, which is at least 8 times greater than the temperature of its outer atmosphere (corona).

For the polytropic case, we write the enthalpy function using the expression

$$\frac{p}{\rho} = \frac{p_0}{\rho_0} \, \frac{T}{T_0} \quad ,$$

where T_0 is the surface temperature. This allows eq. (2.11) to be written

$$T \;=\; T_0 \left[1 - \frac{(\Gamma - 1)(\Psi - \Psi_0)}{\Gamma p_0/\rho_0}\right] \quad . \tag{2.17}$$

A planar approximation yields the constant temperature gradient, also known as the lapse rate

$$\frac{dT}{dz} \;=\; -\left(1 - \frac{1}{\Gamma}\right) \frac{mg}{k_{\rm B}} \quad , \tag{2.18}$$

familiar in polytropic atmospheres (where $\Gamma > 1$ need not be related to the adiabatic exponent).[3] The temperature falls off linearly with height up to the point $T = 0$: the top of the atmosphere.

In the spherical case the atmosphere can also have a top provided

$$\Gamma \frac{p_0}{\rho_0} < (\Gamma - 1)|\Psi_0| \;=\; \tfrac{1}{2}(\Gamma - 1) v_{\rm esc,o}^2 \quad . \tag{2.19}$$

The maximum surface temperature for a bounded polytropic atmosphere is

$$T_0 < \frac{1}{2}\left(1 - \frac{1}{\Gamma}\right) T_{\rm esc,o} \quad . \tag{2.20}$$

When the surface temperature satisfies this inequality then the temperature falls to zero at some altitude: this is formally the top of the atmosphere.

[3]The Earth's atmosphere is mixed by turbulence which very effectively mixes its entropy making it approximately uniform. This means the atmosphere is nearly isentropic and $\Gamma \simeq \gamma \simeq 7/5$, for dry air. The lapse rate is $dT/dz = -9.8\,^\circ$C/km; moist air can have a lapse rate down to one half this value.

2A.4 Conductive planar atmosphere

The polytropic pressure-density relation was adopted, as readily admitted, purely for expediency: it allows a simple closed form expression for the enthalpy integral. The legitimate way of determining the pressure-density relation, eq. (2.4), is by solving the second governing equation eq. (2.2). One scenario which illustrates how this might work is an atmosphere with no sources of heating or cooling except thermal conduction. In this case $\dot{Q} = -\nabla\cdot\mathbf{q} = 0$, according to eq. (1.40). For the case of a planar atmosphere this means

$$\mathbf{q} \;=\; F\,\hat{\mathbf{z}} \;=\; -\,\kappa\frac{dT}{dz}\,\hat{\mathbf{z}} \quad , \tag{2.21}$$

where F is the uniform heat flux being carried upward by the atmosphere, and κ is the thermal conductivity. This heat flux might be carried by a planetary atmosphere, originating at the Sun-warmed ground, or by a layer of a stellar atmosphere, originating from nuclear reactions at the core. Since conduction is neither a source nor a sink, our assumption that it is the only contributor cannot hold over all space — a conductive atmosphere must exist between boundaries where heat is added and removed.

If we further assume that κ is uniform then the temperature gradient is also uniform,

$$\frac{dT}{dz} \;=\; -\,\frac{F}{\kappa} \quad . \tag{2.22}$$

This is a generic property of polytropic atmospheres, so it seems that a planar atmosphere dominated by conductive heat flux with constant conductivity is actually a polytrope. Equating the temperature gradient to that of the polytrope, given in eq. (2.18), allows us to determine the polytropic exponent

$$\Gamma \;=\; \left(1 - \frac{k_{\rm B}F}{mg\kappa}\right)^{-1} \quad . \tag{2.23}$$

It is not a property of the fluid but is, in this case, related to the heat flux and the thermal conductivity. If the heat flux F is very small, or the conductivity κ is very large, then $\Gamma \simeq 1$ and the atmosphere is approximately isothermal. This confirms previous assertions that large thermal conductivity can lead to isothermal behavior.

Alternatively, increasing the flux, F, raises the polytropic exponent until $\Gamma \to \infty$ at

$$F_{\rm cr} \;=\; \frac{mg\kappa}{k_{\rm B}} \quad . \tag{2.24}$$

This divergence means the atmosphere has uniform density, $\rho(z) = \rho_0$ up to the top of the atmosphere at $z = k_{\rm B}T_0/mg$; a situation similar to a layer of water.[4] For fluxes exceeding $F_{\rm cr}$, $\Gamma < 0$ which is not actually unphysical, but it does mean that density *increases* with height even as pressure decreases. In fact, it is shown in Chapter 5 that polytropic atmospheres with $\Gamma > \gamma$ are unstable to convective flows: while the atmosphere is in equilibrum, it is not a stable equilibrium and would not persist.

[4]The fancy term for this is *isochoric*. Some people refer to such a fluid by the term "incompressible", but we reserve that to describe flows rather than fluids.

2B Self-gravitating equilibria: Stellar structure

The previous section tacitly assumed that the gravitational potential $\Psi(\mathbf{x})$ was a known function, independent of density and therefore independent of the solution. This is appropriate for atmospheres of planets or stars, where gravity is provided by the much more massive underlying body. In considering the internal structure of a planet or star, however, the gravitational potential depends on the distribution of mass and must be solved for simultaneously. The hydrostatic equation, (2.1), involves the unknown gravitational field $\nabla\Psi$, which is related to ρ through eq. (1.32). The potential may therefore be eliminated by taking the divergence of (2.1)

$$\nabla\cdot\left(\frac{\nabla p}{\rho}\right) \;=\; -\nabla^2\Psi \;=\; -4\pi G\rho \quad , \tag{2.25}$$

where G is the gravitational constant.

Equation (2.25) expresses the requirement for mechanical force balance between outward pressure force and inward gravitational force. It is a single, second-order PDE, for two unknown fields $p(\mathbf{x})$ and $\rho(\mathbf{x})$, and must be supplemented by an independent relationship between pressure and density. This is the same problem encountered in the previous hydrostatic atmosphere. Once again we cannot appeal to evolution, since fluid elements remain fixed. Nor does an equation of state, say $p=(k_\mathrm{B}/m)\rho T$, help us since that simply introduces a third unknown field $T(\mathbf{x})$ along with the new equation.

Of course the legitimate approach would be to use a static version of eq. (E), equivalent to $\dot{Q}=0$, including various forms of heating: radiation, thermal conduction and also nuclear reactions near the core. The most difficult of these is the conduction, so-called *radiative conductivity*, primarily due to photons moving through the optically thick interior. This contrasts with the simpler case of the optically thin coronal plasma where radiation is a simple loss term and conductive flux is carried by electron subject to Coulomb scattering — Spitzer conductivity $\kappa\sim T^{5/2}$. Photons, on the other hand, scatter by numerous mechanisms characterized by complicated cross sections. The vast majority of any course on stellar structure (see Clayton, 1983, for example) is concerned with deriving the form of the heat transport in the interior in order to derive a relation between p and ρ. Delving into such details here would leave no time for the rest of fluid dynamics, so we will once again use an *ad hoc* work-around to explore stellar structure.

Even before considering the p–ρ relation, we can use the requirement of hydrostatic balance (force balance) expressed in eq. (2.25) to obtain one relation between the mass of a star and its central pressure — at least its rough magnitude. The solution will have spatial scale $L\sim R_*$, the stellar radius, and density $\rho\sim M_*/R_*^3$, where M_* is the star's total mass. Scaling of eq. (2.25) thus demands a pressure and temperature

$$p \;\sim\; \frac{GM_*^2}{R_*^4} \quad , \quad T \;\sim\; \frac{p}{(k_\mathrm{B}/m_p)\rho} \;\sim\; \frac{GM_*m_p}{k_\mathrm{B}R_*} \quad , \tag{2.26}$$

where m_p is the mass of a proton. This scaling predicts the temperature of the Sun to be $T\sim 2.3\times10^7$ K, which turns out to be very close to its actual central temperature. This result has nothing to do with radiation or nuclear reactions — it is required to provide pressure to counter gravity. It is notable that the expression for central temperature exactly

matches $T_{\rm esc,0}$ in eq. (2.16). There is no link between these other than dimensional analysis — it is the only temperature one can obtain using $M_\odot$, $R_\odot$ and constants of nature.

2B.1 Polytropic stellar structure

To go beyond these simple estimates and solve eq. (2.25) would require a relation between p and ρ whose actual form is, as explained above, far more complicated than we can hope to provide at the moment. A popular alternative is to introduce an artificially simple relationship to facilitate a solution. The simplest such relation is the *polytropic* relationship $p \sim \rho^\Gamma$, from eq. (2.7). Such a relation can be justified in certain circumstances — for example $\Gamma = \gamma$ corresponds to an isentropic star. In the end the choice is really made to allow a solution — it is simply expedient.

Having chosen a polytropic relation we may then use the enthalpy, from eq. (2.9), to recast the equation

$$\nabla^2 w \;=\; -4\pi G\rho \;=\; -4\pi G\rho_0 \left(\frac{w}{w_0}\right)^{1/(\Gamma-1)} \quad , \tag{2.27}$$

since $\nabla p/\rho = \nabla w$ by definition. The solution we seek should be spherically symmetric leading to the second-order ODE

$$\frac{1}{r^2}\frac{d}{dr}\left[r^2\frac{d}{dr}\left(\frac{w}{w_0}\right)\right] \;=\; -\frac{4\pi G\rho_0}{w_0}\left(\frac{w}{w_0}\right)^{1/(\Gamma-1)} \quad . \tag{2.28}$$

This non-linear ODE can be integrated from the center ($r = 0$) with "initial conditions" $w(0) = w_0$ and $w'(0) = 0$. The second condition is equivalent to requiring $\nabla p = 0$ at the center where $\mathbf{g} = -\nabla\Psi$ also vanishes. After all the center must be the point of maximum pressure.

Introducing the dimensionless variables, $\tilde{r} = r/R_*$, and $\phi = w/w_0$, normalized to the stellar radius and central enthalpy, the equation becomes

$$\frac{1}{\tilde{r}^2}\frac{d}{d\tilde{r}}\left[\tilde{r}^2\frac{d\phi}{d\tilde{r}}\right] \;=\; -\frac{4\pi G\rho_0 R_*^2}{w_0}\,\phi^{1/(\Gamma-1)} \;=\; -\lambda\,\phi^{1/(\Gamma-1)} \quad . \tag{2.29}$$

This is an eigenvalue problem for which the value of λ must be found in order to satisfy the three conditions, namely $\phi(0) = 1$ and $\phi'(0) = 1$ at the center ($\tilde{r} = 0$), as well as $\phi(1) = 0$ at the outer surface ($\tilde{r} = 1$). Having found the eigenvalue, λ, it is then possible to determine the central properties,

$$\frac{w_0}{\rho_0} \;=\; \frac{4\pi G R_*^2}{\lambda} \quad , \tag{2.30}$$

and the central pressure using eq. (2.9),

$$p_c \;=\; \frac{\Gamma-1}{\Gamma}\,w_0\,\rho_0 \;=\; \frac{\Gamma-1}{\Gamma}\,\frac{4\pi G\rho_0^2 R_*^2}{\lambda} \quad . \tag{2.31}$$

The central density may be related to the radius and total mass of the star through the integral

$$\begin{aligned} M_* \;&=\; 4\pi\rho_0\int_0^{R_*}\frac{\rho}{\rho_0}\,r^2\,dr \;=\; 4\pi\rho_0 R_*^3\int_0^1 \phi^{1/(\Gamma-1)}\,\tilde{r}^2 d\tilde{r} \\ &=\; -\frac{4\pi\rho_0 R_*^3}{\lambda}\int_0^1 \frac{d}{d\tilde{r}}\left[\tilde{r}^2\frac{d\phi}{d\tilde{r}}\right]d\tilde{r} \;=\; -\frac{4\pi\rho_0 R_*^3}{\lambda}\,\phi'(1) \quad , \end{aligned} \tag{2.32}$$

after using eq. (2.29) to replace $\phi^{1/(\Gamma-1)}$. Using this to eliminate ρ_0 from eq. (2.31) yields

$$p_c \;=\; \frac{(\Gamma-1)\lambda}{4\pi\Gamma[\phi'(1)]^2}\,\frac{GM_*^2}{R_*^4} \quad . \tag{2.33}$$

This expression matches the scaling of (2.26) — as it must by simple dimensional arguments — but shows the constant of proportionality dependant on the eigenvalue, λ, and eigenfunction, ϕ, satisfying eq. (2.29). It seems we must solve the full eigenfunction equation in order to find the exact constant in front of the central pressure.

2B.2 The Lane-Emden equation

Most eigenvalue equations require an ODE be solved repeatedly, while candidate eigenvalues are tried in turn, until the third, outer boundary condition is satisfied. Equation (2.29), however, has a very special property which makes this iteration unnecessary. In particular, the left hand side is homogeneous in the spatial variable $\tilde{r}$, which makes it possible to absorb the unknown eigenvalue, λ, into a different rescaling of the radius

$$\xi \;=\; \lambda^{1/2}\tilde{r} \;=\; \frac{\lambda^{1/2}}{R_*}\,r \quad . \tag{2.34}$$

In terms of this, eq. (2.29) takes the form of the simple non-linear ODE

$$\frac{1}{\xi^2}\frac{d}{d\xi}\left(\xi^2\frac{d\phi}{d\xi}\right) \;=\; \phi'' + \frac{2}{\xi}\phi' \;=\; -\,\phi^{1/(\Gamma-1)} \;=\; -\,\phi^n \tag{2.35}$$

known as the *Lane-Emden equation*. The equation has only one parameter, $n = 1/(\Gamma-1)$, and intial conditions $\phi(0) = 1$ and $\phi'(0) = 0$. The unknown eigenvalue has been traded for unknown location where the outer boundary condition $\phi = 0$ would be imposed. Instead the equation is integrated from $\xi = 0$ to a point $\xi_{\rm max}$ where ϕ first vanishes. This position is the outer boundary, $\tilde{r} = 1$, and gives the eigenvalue using eq. (2.34),

$$\lambda \;=\; \xi_{\rm max}^2 \quad . \tag{2.36}$$

The derivative of the eigenfunction at the outer boundary becomes

$$\phi'(\tilde{r}=1) \;=\; \xi_{\rm max}\,\phi'(\xi_{\rm max}) \quad , \tag{2.37}$$

leading to a central pressure

$$p_c \;=\; \frac{\Gamma-1}{4\pi\Gamma[\phi'(\xi_{\rm max})]^2}\,\frac{GM_*^2}{R_*^4} \quad . \tag{2.38}$$

An intuitive sense of this system can be obtained from the two cases for which eq. (2.35) can be exactly solved. The first, $n = 0$, corresponds to $\Gamma \to \infty$. In this limit the density must be uniform, $\rho = \rho_0$, — a limit sometimes called the *water moon*. The solution is readily found to be

$$\phi(\xi) \;=\; 1 \;-\; \frac{\xi^2}{6} \quad , \quad n = 0 \tag{2.39}$$

from which we see $\xi_{\max} = \sqrt{6}$, and $\phi'(\xi_{\max}) = -\sqrt{2/3}$. Using this in eq. (2.38) gives the central pressure of the water moon,

$$p_c \;=\; \frac{3}{8\pi}\frac{GM_*^2}{R_*^4} \quad . \tag{2.40}$$

This conforms to the dimensional result, eq. (2.26), with a factor $3/8\pi$ provided by the exact solution for $n = 0$.

The only other choice which renders eq. (2.35) linear, and thus analytically solvable,[5] is $n = 1$ corresponding to $\Gamma = 2$ (i.e. $p \sim \rho^2$). For this choice eq. (2.35) becomes the spherical Bessel equation whose solution is

$$\phi(\xi) \;=\; \frac{\sin\xi}{\xi} \quad , \quad n = 1 \tag{2.41}$$

and $\xi_{\max} = \pi$ and $\phi'(\xi_{\max}) = -1/\pi$. Using this in eq. (2.38) gives a centrlal pressure

$$p_c \;=\; \frac{\pi}{8}\frac{GM_*^2}{R_*^4} \quad . \tag{2.42}$$

has the same dimensional term multiplied by $\pi/8$ — larger than the water moon by $\pi^2/3 = 3.29$.

For all other values of the polytropic index Γ, and thus of n, eq. (2.35) will be non-linear and must be solved numerically. Such solutions, several shown in fig. 2.1, all decrease monotonically just like eqs. (2.39) and (2.41).[6] Increasing n leads to increasing values of $\xi_{\max}$ and also to increasing central concentration — i.e. increasing $\rho_0/\bar{\rho}$. In conjunction with an increasing $\xi_{\max}$, there is evidently a decreasing derivative there, $\phi'(\xi_{\max})$, which in eq. (2.38) gives an increasing central pressure with increasing n (decreasing Γ).

While the polytropic relation between p and ρ was introduced entirely for expedience, a sophisticated solution of the Sun's interior structure (Christensen-Dalsgaard et al., 1996) appears to match the $n = 4.6$ solution below the convection zone (see fig. 2.1). It seems that the detailed radiative transport, at least above $T = 2$ million Kelvins, yields $p \simeq \rho^{1.22}$, approximately.

2C Coronal loops: one-dimensional atmospheres

The tenuous atmosphere above the visible surface of the Sun, called the solar corona, is an atmosphere dominated by its magnetic field. Except during dynamic events, such as solar flares, the corona evolves very slowly and can be approximated as an equilibrium — but one dominated by magnetic forces. It will be shown later that magnetic fields are capable of exerting forces perpendicular to $\mathbf{B}$, and that if magnetic energy density $B^2/8\pi \gg p$ as it is the corona,[7] then the magnetic force will greatly exceed any from pressure or

[5] $n = 0$ and $n = 1$ are the only choices which yield an equation linear in the unknown and therefore solvable by standard techniques. By pure chance there is one more case, $n = 5$, which happens to have a closed-form solution. *To wit* $\phi(\xi) = (1 + \xi^2/3)^{-1/2}$ can be seen to satisfy the equation when $n = 5$. This never vanishes so $\xi_{\max} \to \infty$. In fact, $n = 5$ is the upper limit on solutions with finite $\xi_{\max}$.

[6] Indeed, they must all match the first two terms in the Taylor series of either solution.

[7] This limit is referred to as $\beta \ll 1$.

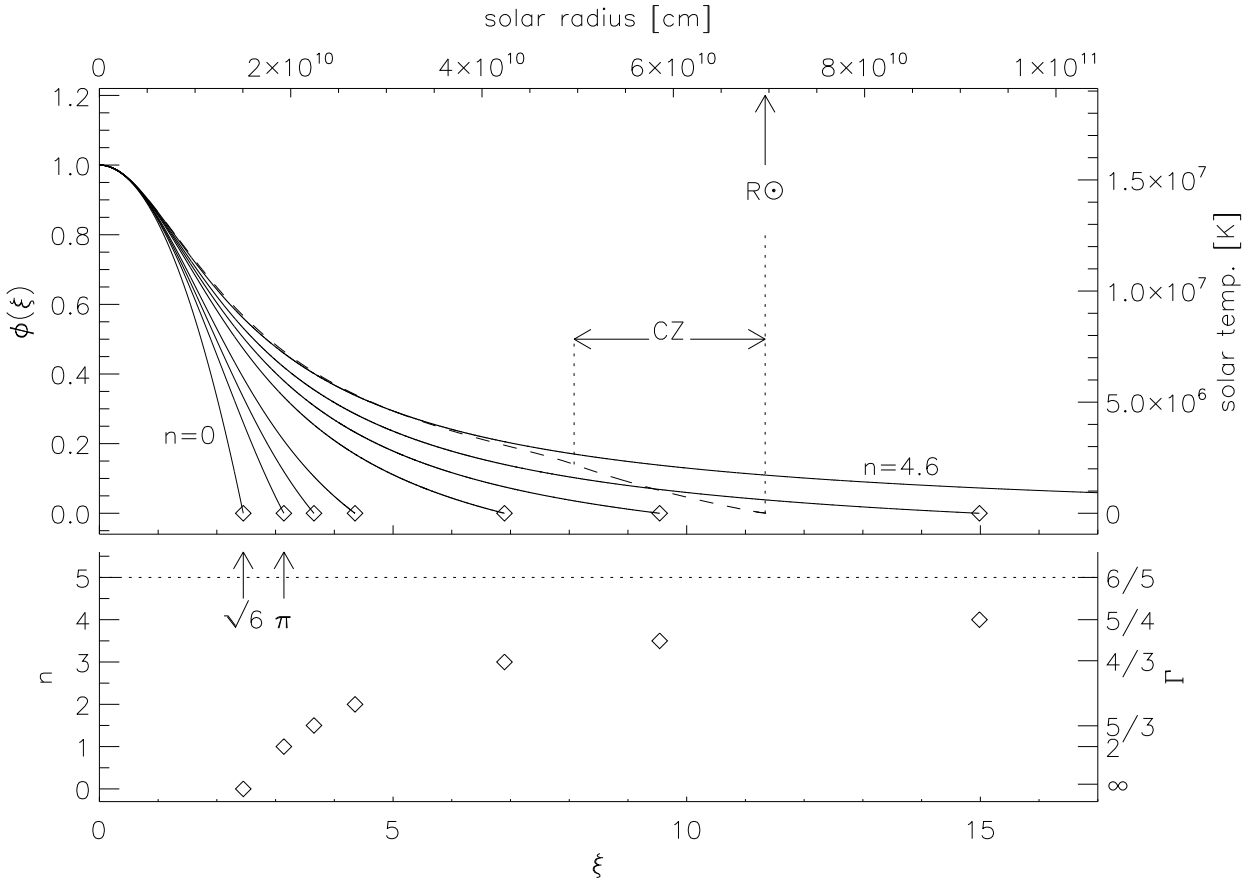

Figure 2.1: Solutions to the Lane-Emden equation, eq. (2.35), for $n = 0$, 1, 1.5, 2, 3, 3.5, 4 and 4.6 (solid curves in the top panel, from left to right). A diamond marks $\xi_{\max}$ on each curve, and its value of n is plotted agains this position in the bottom panel. The axis on the right gives the corresponding value of Γ. The horizontal dotted line is the upper limit $n = 5$, above which $\phi(\xi)$ does not vanish. The dashed curve in the upper panel shows the normalized temperature of the Sun against a re-scaled axis shown along the top (actual temperature is shown on the right). Below the convection zone (labeled CZ), it matches the $n = 4.6$ ($\Gamma = 1.22$) polytrope reasonably well.

gravity. Since there are no magnetic forces *parallel* to $\mathbf{B}$ the force balance in that direction is purely hydrostatic: pressure balancing gravity. Thus the corona consists of numerous *one dimensional hydrostatic atmospheres* confined to magnetic field lines. Images made at EUV or X-ray wavelengths show the corona to comprise many distinct *loops* believed to be these one-dimensional hydrostatic atmospheres exceeding one million Kelvins. Each follows a bundle of magnetic field lines known as a *flux tube*.

A magnetic field line is a curve, $\mathbf{x}(\ell)$, which is everywhere tangent to the local magnetic field vector,

$$\frac{d\mathbf{x}}{d\ell} \equiv \hat{\mathbf{b}} = \frac{\mathbf{B}[\mathbf{x}(\ell)]}{\left|\mathbf{B}[\mathbf{x}(\ell)]\right|} \quad ; \tag{2.43}$$

where ℓ, the length along the curve, serves to parametrize it. Taking the dot product of (2.1) with the tangent vector $\hat{\mathbf{b}}$, yields

$$\frac{1}{\rho}\hat{\mathbf{b}}\cdot\nabla p = \frac{1}{\rho}\frac{\partial p}{\partial \ell} = -\frac{\partial \Psi}{\partial \ell} \quad , \tag{2.44}$$

an equation for force balance *parallel* to the magnetic field.

2C.1 Energetics of loop equilibria

In order to solve eq. (2.44), we need a relation between the pressure and density, just as we did in the previous two sections. As before an equation of state, $p = (k_{\rm B}/m)\rho T$, while valid, does not fit the bill since it introduces another unknown field $T(\ell)$. The previous analysis resorted to *ad hoc* relations, such as polytropes, to serve this role. The coronal loop is one of the few cases where the correct physics can be used to construct an atmosphere.

The relation between p and ρ is naturally provided by one of the remaining fluid equations. The static ($\mathbf{u} = 0$) version of energy equation, (E) reduces to

$$\frac{Dp}{Dt} + \gamma p(\nabla\cdot\mathbf{u}) = 0 = (\gamma-1)\dot{Q} \quad . \tag{2.45}$$

This apparently trivial relation, $\dot{Q} = 0$, states that to remain in equilibrium the sources and sinks of heat must balance at every point.

The absence of bulk flows leaves only two obvious contributions to $\dot{Q}$: thermal conduction and radiative loss. Both of these will contribute *negatively* and thus cannot sum to zero by themselves. In an optically thin plasma, like the corona, radiation is an energy loss expressed by eq. (1.39) using the loss function $\Lambda(T)$. Thermal conduction will transport heat from the hot corona to the cold feet at its ends, so it is a loss from the loop's center. If these were the only two contributions we could never achieve an equilibrium. Thus there must be an additional heat *source* of some kind to add the heat which conduction and radiation carry away. This could be Ohmic heating from electric currents or viscous damping of motions, perhaps turbulent, on small scales within the loop. The details of the heating remain a mystery, frequently referred to as "the coronal heating problem". What is certain is that no equilibrium is possible unless *something* does heat the corona. Until the coronal heating problem is definitively solved we are forced to introduce an *ad hoc* function $h(\ell)$ to represent the volumetric heating responsible for creating the million-plus-degree corona.

Following the logic above the energy equation becomes

$$0 \;=\; \dot{Q} \;=\; \nabla\cdot\left(\underline{\underline{\kappa}}\cdot\nabla T\right) \;-\; n_e^2\Lambda(T) \;+\; h \quad . \tag{2.46}$$

Thermal conduction is expressed using eq. (1.43), but with a conductivity tensor, κ_{ij}, instead of the more common scalar. Magnetic field strongly affects the electrons carrying the bulk of the heat,[8] causing the heat flux

$$\mathbf{q} \;=\; -\,\underline{\underline{\kappa}}\cdot\nabla T \quad , \tag{2.47}$$

to be everywhere parallel to $\hat{\mathbf{b}}$. Because the heat is carried along the field line by electrons which move ever faster at higher temperatures, the conductivity is a strongly increasing function of temperature,

$$\underline{\underline{\kappa}} \;=\; \kappa_\parallel\hat{\mathbf{b}}\hat{\mathbf{b}} \;=\; \kappa_0 T^{5/2}\,\hat{\mathbf{b}}\hat{\mathbf{b}} \quad , \tag{2.48}$$

where

$$\kappa_0 \;\simeq\; 0.053\frac{k_{\rm B}^{7/2}}{m_e^{1/2}e^4} \;=\; 10^{-6}\,{\rm erg\,cm^{-1}\,sec^{-1}\,K^{-7/2}} \quad , \tag{2.49}$$

is derived later as eqs. (15.32) and (15.76) and known as *Spitzer conductivity*.

Once the final component, $h(\ell)$, has been specified it is possible to solve the pair of equations, (2.44) and (2.46) to find the distributions $p(\ell)$, $\rho(\ell)$ and $T(\ell)$ along any field line $\mathbf{x}(\ell)$. This must be done numerically. In practice, it is very often done after making three simplifying assumptions. Firstly one assumes the cross section of the loop is uniform in order to simplify the divergence in the conduction term. Secondly one takes the volumetric heating h to be uniform along the entire loop; h thus becomes a parameter of the problem. Finally one assumes an *isothermal* hydrostatic balance owing to the very large conductivity. This assumption, which can be verified after the fact, permits the use of eq. (2.12) in the form

$$p(\mathbf{x}) \;=\; p_0\,\exp\left[-\frac{\Psi(\mathbf{x})-\Psi_0}{c_s^2}\right] \;=\; p_0\,e^{-z(\ell)/H_p} \quad , \tag{2.50}$$

provided the loop is small compared to $R_\odot = 696$ Mm, so the gravitational field may be approximated as planar $\Psi = gz$.

An even more stringent assumption on the smallness of the loop, that its apex height

$$\max({\rm z}) \;\ll\; H_p \;=\; 60\,{\rm Mm}\left(\frac{T}{10^6\,{\rm K}}\right) \quad , \tag{2.51}$$

permits us to approximate the plasma pressure as uniform throughout the loop: $p(\ell)\simeq p_0$. This permits us to use the ideal gas law to replace density,

$$n_e(\ell) \;=\; \frac{p_0}{2k_{\rm B}T(\ell)} \quad , \tag{2.52}$$

[8] This same effect is the basis for magnetic confinement in nuclear fusion reactors. In those experiments a strong magnetic field is produced which never crosses the walls of the solid confinement vessel. The plasma can be heated to $\sim 10^8$ K, with little effect on the walls.

for a fully-ionized hydrogen plasma.[9] With all of these assumptions eq. (2.46) becomes an ODE for a single unknown,

$$\frac{d}{d\ell}\left(\kappa_0\, T^{5/2}\frac{dT}{d\ell}\right) \;-\; \frac{p_0^2}{4k_{\rm B}^2}\,\frac{\Lambda(T)}{T^2} \;+\; h \;=\; 0 \quad , \tag{2.53}$$

with parameters h and p_0. The shape of the loop no longer affects the solution owing to our assumption that it was smaller than the gravitational scale height. This equation was proposed and solved independently by Rosner, Tucker and Vaianna (1978) and by Craig et al. (1978), and its solutions are referred to as RTV loops after the former. If one proposes a simple form of radiative losses, $\Lambda(T)$, the equation can be solved by straightforward means — this is the traditional approach. Owing the the importance of the results these steps are followed explicitly below.

2C.2 The RTV relations

We integrate eq. (2.53) once after first multiplying the entire equation by $T^{5/2}T'(\ell)$

$$\frac{d}{d\ell}\left[\tfrac{1}{2}\kappa_0\, T^5\left(\frac{dT}{d\ell}\right)^2 \;-\; \frac{p_0^2}{4k_{\rm B}^2}\int^T \Lambda(T')\sqrt{T'}\,dT' \;+\; \tfrac{2}{7}\,hT^{7/2}\right] \;=\; 0 \quad . \tag{2.54}$$

The factor in square brackets is constant — a constant of integration — and can be evaluated at one of the boundaries. The ends of the loop, $\ell = 0$ and $\ell = L$, are footpoints where the coronal structure is anchored to cooler layers such as the chromosphere or photosphere. We assume these layers are so cool that they may approximated as $T(0) = T(L) = 0$ without seriously compromising the solution. Evaluating the constant at a footpoint we find the lower limit of the integral is $T = 0$ and obtain a first-order equation

$$\tfrac{1}{2}\kappa_0\, T^5\left(\frac{dT}{d\ell}\right)^2 \;=\; \frac{p_0^2}{4k_{\rm B}^2}\int_0^T \Lambda(T')\sqrt{T'}\,dT' \;-\; \tfrac{2}{7}\,hT^{7/2} \quad , \tag{2.55}$$

clearly satisfying the footpoint boundary conditions.

Equations (2.55) must hold along the entire loop. At some point between the two footpoints $T'(\ell)$ will vanish and the temperature will reach a maximum $T_{\rm max}$. At this point the left hand side of (2.55) will vanish and we find a relation between heating, pressure and maximum temperature

$$h \;=\; \frac{7p_0^2}{8k_{\rm B}^2}\;T_{\rm max}^{-7/2}\int_0^{T_{\rm max}} \Lambda(T')\sqrt{T'}\,dT' \;=\; \frac{7p_0^2}{8k_{\rm B}^2}\,G_1(T_{\rm max}) \quad . \tag{2.56}$$

The function $G_1(T)$ involves an integral over the radiative loss function $\Lambda(T)$, of which fig. 1.1 shows several versions from the Astrophysical literature. Figure 2.2 shows the relationship (bottom panel) for the published forms of $\Lambda(T)$.

[9] Each species is at the same temperature and has equal number density, $n_e = n_i$, to maintain charge neutrality. Each species thus has the same pressures: $p_e = n_e k_{\rm B} T = p_i$. The total pressure p_0 is the sum of these partial pressures and therefore twice p_e. Accounting more accurately for the abundances of minority elements and their ionizations would decrease, slightly the factor of 2 in the denominator of eq. (2.52).

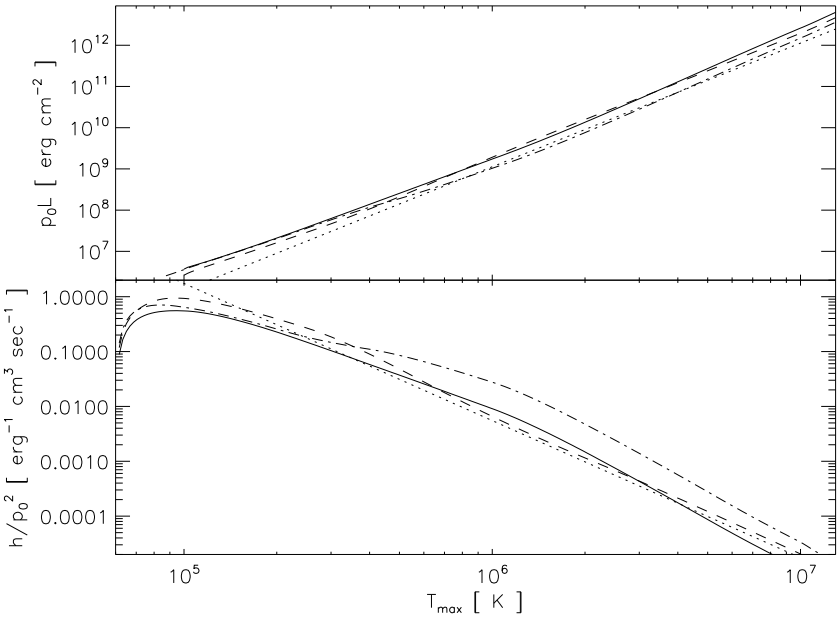

Figure 2.2: Bottom and top panels show relations (2.55) and (2.61), respectively, for the different versions of the radiative loss function shown in fig. 1.1. The line styles correspond to those from fig. 1.1: dashed, Rosner et al. (1978); solid, Cook et al. (1989); dash-dot, Martens et al. (2000); and dotted is the single power-law approximation $\Lambda(T) = 1.2 \times 10^{-19}\, T^{-1/2}$.

A particularly simple approximation,[10] $\Lambda(T) \sim T^{-1/2}$, makes the integral tractable, $G_1(T) \sim T^{-5/2}$, and has been widely used in the literature; we hereafter refer to it as the single power-law approximation. Using it provides the closed-form relation

$$h \;=\; \frac{7p_0^2}{8k_{\rm B}^2}\,\frac{\Lambda(T_{\rm max})}{T_{\rm max}^2} \;\sim\; p_0^2\,T_{\rm max}^{-5/2} \quad , \quad \Lambda(T) \;\sim\; T^{-1/2} \quad . \tag{2.57}$$

Expression (2.57) can be used to write the local radiative losses at any point in the loop

$$n_e^2\Lambda(T) \;=\; \frac{p_0^2}{4k_{\rm B}^2T^2}\Lambda(T) \;=\; \frac{2}{7}\,\frac{T_{\rm max}^2}{T^2}\,\frac{\Lambda(T)}{\Lambda(T_{\rm max})}\,h \;=\; \frac{2}{7}\left(\frac{T_{\rm max}}{T}\right)^{5/2} h \quad . \tag{2.58}$$

as a fraction of heat input. At the point of peak temperature the radiative loss is less than the heating by a factor 2/7. This fraction increases with decreasing T because n_e increases, according to eq. (2.52), leading to more radiation. At the point where

$$T \;=\; T_\times \;\equiv\; \left(\frac{2}{7}\right)^{2/5} T_{\rm max} \;=\; 0.61\,T_{\rm max} \quad , \tag{2.59}$$

radiative losses exactly balance the heating. This point divides the central portion of the loop, where heating exceeds losses, from the outer, denser regions where radiative losses exceed heating. Thermal conduction carries excess heat from the hot central region ($T > T_\times$), to the cooler ($T < T_\times$), denser portions where it is radiated.

Expression (2.56) can be used to eliminate the constant h from eq. (2.55) and obtain the first-order ODE

$$\frac{dT}{d\ell} \;=\; \pm\,\frac{p_0}{\sqrt{2\kappa_0}k_{\rm B}}\,T^{-3/4}\sqrt{G_1(T) \;-\; G_1(T_{\rm max})} \quad , \tag{2.60}$$

for the temperature distribution. The upper (lower) sign corresponds to points before (after) the temperature peak. Symmetry between the two sides of the peak (equal magnitude, opposite sign) means the peak must occur at the exact mid-point of the loop $\ell = L/2$.[11] The actual shape of the loop was rendered irrelevant by the assumption of large gravitational scale height. Thus the mid-point ($\ell = L/2$) is not necessarily the same as the apex where $z(\ell)$ is maximum. The temperature is maximum at the mid-point rather than the apex because it is farthest from the cold boundaries following the path taken by the heat: the field line.

The full length of the loop, L, is a fixed parameter of the problem, along with the heating rate h. Equation (2.60) may be integrated from $\ell = 0$ to $\ell = L/2$, and the and the temperature jump set to $T_{\rm max}$. This yields a new constraint equation

$$p_0 L \;=\; \sqrt{8\kappa_0}k_{\rm B}\int_0^{T_{\rm max}} \frac{T^{3/4}\,dT}{\sqrt{G_1(T) - G_1(T_{\rm max})}} \;=\; \sqrt{8\kappa_0}k_{\rm B}\,G_2(T_{\rm max}) \quad . \tag{2.61}$$

[10] See the dotted curve in fig. 1.1.

[11] Many authors prefer to use L to denote the *half-length* of the loop. This strikes me as needlessly complicated, so I have gone against trend and use L to denote the loop's *full* length — from one footpoint to the other.

This relation is plotted in the top panel of fig. 2.2 for G_2 computed using different versions of the radiative loss function. In the single power-law case $\Lambda(T) \sim T^{-1/2}$ the integral can be evaluated

$$G_2(T_{\rm max}) \;=\; \frac{T_{\rm max}^{11/4}}{\sqrt{\Lambda(T_{\rm max})}} \int_0^1 \frac{s^2 ds}{\sqrt{1-s^{5/2}}} \;=\; 0.716\, \frac{T_{\rm max}^{11/4}}{\sqrt{\Lambda(T_{\rm max})}} \;\sim\; T_{\rm max}^3 \quad . \tag{2.62}$$

The result from this approximation, that $p_0 L \sim T_{\rm max}^3$, was first derived by Rosner, Tucker and Vaiana (1978) and is known generally as the *RTV scaling law*. Figure 2.2 suggests that it is a very good approximation in spite of any other deficiencies the single power-law approximation may have. It is generally true, according to eq. (2.61), that the product, $p_0 L$, is a function — indeed, a very steep function — of $T_{\rm max}$.

This completes the well-known series of steps deriving properties of equilibrium coronal loops. A loop of full length L, and uniform but unspecified cross section, is subject to a uniform volumetric heating h. This heating probably enters the loop is mechanical form through its cold footpoints as a flux $F = hL$. This energy is dissipated within the loop, raising its mid-point to a temperature $T_{\rm max}$. Expressions (2.56) and (2.61) can be combined to yield an explicit relationship,

$$hL^2 \;=\; FL \;=\; 7\kappa_0\, G_1(T_{\rm max})\, G_2^2(T_{\rm max}) \quad , \tag{2.63}$$

between heating, length and peak temperature. The function on the right is monotonic and may be inverted to find the peak temperature resulting from a given heating in a loop of a given length. The relationship is particularly illuminating in the single power-law case

$$hL^2 \;=\; FL \;=\; 3.59\,\kappa_0\, T_{\rm max}^{7/2} \quad , \quad \Lambda(T) \;\sim\; T^{-1/2} \quad . \tag{2.64}$$

Thermal conduction cools the temperature maximum at a volumetric rate whose magnitude matches the local heating rate,

$$\frac{d}{d\ell}\left(\kappa_\parallel \frac{dT}{d\ell}\right) \;\sim\; \frac{\kappa_0 T_{\rm max}^{7/2}}{(L/2)^2} \;=\; \frac{4}{3.59}\, h \quad . \tag{2.65}$$

While conduction cannot rid the loop of energy, it must transport it away from the mid-point to cooler regions where it can be radiated away. This must occur at the same rate it is being added, and thus the rate of conductive cooling must match both other rates, lest there be imbalance somewhere.

An inversion of eq. (2.64) reveals that peak temperature scales as relatively weak powers of heat flux and loop length: $T_{\rm max} \sim F^{2/7} L^{2/7}$. This fact offers a nice explanation of why static coronal loops appear within a relatively narrow range of temperatures even as they span a much wider range of lengths. It may then be combined with eqs. (2.61) and (2.62) to find

$$p_0 \;\sim\; F^{6/7}\, L^{-1/7} \quad , \quad \Lambda(T) \;\sim\; T^{-1/2} \quad .$$

It seems that increasing heat input does more to raise the density of a loop than its temperature.

2C.3 The structure of loop atmospheres

So far we have derived relationships between parameters of the loop but have not found the distribution of fluid properties which we call the equilibrium. Adopting hereafter the single power-law approximation, $\Lambda(T) = \Lambda_0 T^{-1/2}$, the left side ($\ell < L/2$) of the loop is governed by the top sign of eq. (2.60),

$$\frac{dT}{d\ell} \;=\; \frac{p_0}{\sqrt{2\kappa_0}k_{\rm B}}\,T^{-3/4}\sqrt{\Lambda_0}\sqrt{T^{-5/2} \;-\; T_{\rm max}^{-5/2}} \quad , \tag{2.66}$$

The temperature distribution, $T(\ell)$, can be found implicitly from the indefinite integral

$$\begin{aligned}\ell \;&=\; \frac{\sqrt{2\kappa_0}k_{\rm B}}{p_0\sqrt{\Lambda_0}}\int_0^T \frac{T^2\,dT}{\sqrt{1-(T/T_{\rm max})^{5/2}}} \\ &=\; \frac{\sqrt{2\kappa_0}k_{\rm B}T_{\rm max}^3}{p_0\sqrt{\Lambda_0}}\int_0^{T/T_{\rm max}} \frac{s^2\,ds}{\sqrt{1-s^{5/2}}} \quad .\end{aligned} \tag{2.67}$$

The indefinite integral can be expressed using the *incomplete beta function* to yield an implicit distribution

$$\ell \;=\; \frac{\sqrt{2\kappa_0}k_{\rm B}T_{\rm max}^3}{p_0\sqrt{\Lambda_0}}\,\tfrac{5}{2}\,B\left[\,(T/T_{\rm max})^{5/2};\,\tfrac{6}{5},\,\tfrac{1}{2}\,\right] \;=\; \frac{L}{2}\,\frac{B\left[\,(T/T_{\rm max})^{5/2};\,\tfrac{6}{5};\,\tfrac{1}{2}\,\right]}{B\left(\,1;\,\tfrac{6}{5},\,\tfrac{1}{2}\,\right)} \quad , \tag{2.68}$$

plotted in fig. 2.3. The final expression uses the RTV scaling law, eqs. (2.61) and (2.62), to replace $T_{\rm max}^3/p_0$ in terms of L.

The exact solution, eq. (2.68), can appear daunting in its use of an uncommon special function. Expanding the integral in eq. (2.67) shows that $\ell \sim T^3$ at small temperatures. This suggests a heuristic form for the profile

$$T(\ell) \;\sim\; T_{\rm max}\,\sqrt[3]{4(\ell/L)(1-\ell/L)} \quad , \tag{2.69}$$

to replace the incomplete beta functions. It is so close the actual solution that when it is overplotted, as a broken line in fig. 2.3, it can barely be seen.

The loop atmosphere, plotted in fig. 2.3 and approximated by expression (2.69), has properties reflecting the physics which creates it. The temperature has a broad, shallow central peak owing to the very high conduction where T is large. The heating-dominated central region accounts for $\sim 90\%$ of the loop. The temperature drops extremely rapidly at each end, producing a layer of large density where radiation dominates. These thin layers show at least some of the physics which leads to a thin *transition region* between the corona and the lower atmosphere: namely the very rapid drop in thermal conduction with diminishing temperature. Of course it is precisely in these layers that our neglect of gravity (owing to large scale height) becomes suspect.

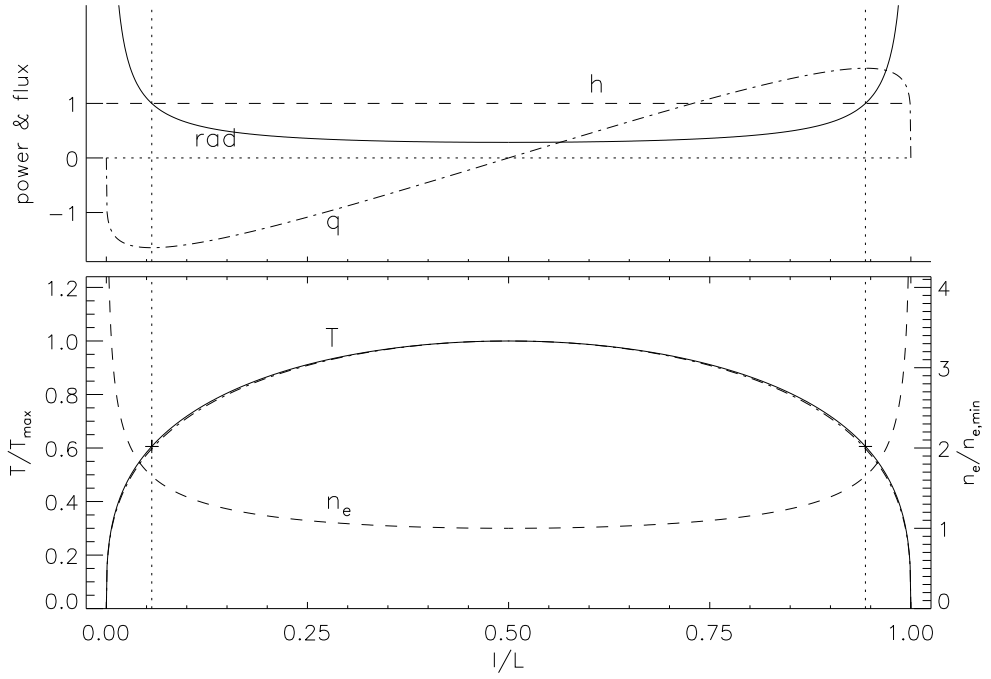

Figure 2.3: The temperature (solid) and density (dashed) profiles in a coronal loop using single power-law approximation. These are plotted in the bottom panel against normalized axis on the left and right respectively. +s mark the locations $T = T_\times$ which separate the central region from the radiatively dominated end regions; vertical dotted lines show this division. A broken line showing the approximate relation from eq. (2.69) is barely visible under the solid. The top panel shows the heating, h (dashed), and radiative losses, (solid), from eq. (2.58). The raditiive loss is plotted as a positive quantity so that it crosses h at the vertical dotted lines from $T = T_\times$. The heat flux, $q = -\kappa\, dT/d\ell$, is plotted (after arbitrary re-scaling) as a dash-dot line. This peaks at the vertical dotted line.

Chapter 3

Steady Flows & Parker's Wind

We next consider equilibrium solutions to the fluid equations (i.e. $\partial/\partial t = 0$) which are *not* static ($\mathbf{u} \neq 0$). This means there is a steady flow described by the velocity field $\mathbf{u}(\mathbf{x})$. A given fluid element moves along a curve, $\mathbf{r}(t)$, called a *streamline* defined by the equation

$$\frac{d\mathbf{r}}{dt} \;=\; \mathbf{u}[\,\mathbf{r}(t)\,] \quad . \tag{3.1}$$

Streamlines are curves which are everywhere parallel to the flow field, and are a very common, and natural way of visualizing a steady flow. The streamline, defined by eq. (3.1), is a special case of the trajectory, when the velocity field in eq. (1.5) has no time dependence. This means a fluid element moves along the streamline, the advective derivative, $D/Dt = \mathbf{u} \cdot \nabla$, the directional derivative along the streamline.

In this case the momentum equation (M), becomes

$$(\mathbf{u} \cdot \nabla)\mathbf{u} \;=\; \frac{-\nabla p}{\rho} + \mathbf{g} \;=\; \frac{-\nabla p}{\rho} - \nabla\Psi \quad . \tag{3.2}$$

where Ψ is the gravitational potential ($\mathbf{g} = -\nabla\Psi$). The force of viscosity has been neglected this time under the assumption of extremely high Reynold's number; a characteristic of most flows of astrophysical interest. Taking the dot product with $\mathbf{u}$ gives

$$\mathbf{u} \cdot \nabla(\tfrac{1}{2}u^2 + \Psi) \;=\; -\frac{1}{\rho}\mathbf{u} \cdot \nabla p \;=\; -\frac{1}{\rho}\frac{Dp}{Dt} \quad , \tag{3.3}$$

where we have used steady state, $\partial p/\partial t = 0$, to write the final expression.

We would like to write the right hand side as a perfect time derivative in order to match the left. To do so we use information about the evolution of the fluid's internal energy (i.e. its pressure). As in the hydrostatic case we will here be able analyze two different, and mutually inconsistent, situations in parallel. In both cases we will neglect all heating except possibly thermal conductivity. If the thermal conductivity is also *very small* we will neglect that too and take the dynamics to be *adiabatic*, $\dot{Q} = 0$. As an alternative to neglecting conductivity we will also consider the opposite situation where thermal conductivity κ is *very large* which will lead to an isothermal flow: $T = T_0$ is uniform.

For adiabatic evolution $p \propto \rho^\gamma$ and the specific entropy is conserved, $Ds/Dt = 0$. The specific energy changes according to the first law of thermodynamics, $d\varepsilon = -p\,dV + T\,ds$,

where $V = 1/\rho$ is specific volume. The specific *enthalpy* of the fluid is $w = \varepsilon + pV$ and therefore changes according to

$$dw \;=\; d\varepsilon + p\,dV + V\,dp \;=\; V\,dp + T\,ds \quad .$$

Replacing the differentials by advective time derivatives gives the enthalpy evolution

$$\frac{Dw}{Dt} \;=\; \frac{1}{\rho}\frac{Dp}{Dt} \;+\; T\frac{Ds}{Dt} \;=\; \frac{1}{\rho}\frac{Dp}{Dt} \quad , \tag{3.4}$$

since $Ds/Dt = 0$ under adiabatic evolution.

In the other scenario, *isothermal* evolution, the temperature is constant and so is the isothermal sound $c_{\rm s} = \sqrt{k_{\rm B}T/m} = \sqrt{p/\rho}$. In that case

$$\frac{1}{\rho}\frac{Dp}{Dt} \;=\; \frac{D}{Dt}\Big(c_s^2 \ln\rho\Big) \quad . \tag{3.5}$$

We can consolidate the two cases using the same kind of enthalpy function introduced for hydrostatic atmospheres in eq. (refeq:w2)

$$w \;=\; \begin{cases} \dfrac{\gamma}{\gamma-1}\dfrac{p}{\rho} = \dfrac{c_s^2}{\gamma-1} = \dfrac{c_{s0}^2}{\gamma-1}(\rho/\rho_0)^{\gamma-1} & , \quad \text{adiabatic} \\[2ex] c_s^2 \ln(\rho/\rho_0) & , \quad \text{isothermal} \end{cases} \tag{3.6}$$

3A Bernoulli's Law

From the foregoing we see that for any definition of enthalpy w

$$\frac{1}{\rho}\frac{Dp}{Dt} \;=\; \frac{Dw}{Dt} \;=\; \mathbf{u}\cdot\nabla w \quad . \tag{3.7}$$

Using this in (3.3) leads to a conservation law

$$\mathbf{u}\cdot\nabla(\tfrac{1}{2}u^2 + w + \Psi) \;=\; 0 \quad , \tag{3.8}$$

whose solution is known as Bernoulli's Law

$$\tfrac{1}{2}u^2 + w + \Psi \;=\; \text{constant along flow.} \;=\; \mathcal{B} \tag{3.9}$$

This is a kind of energy conservation (per unit mass) for a fluid element in a steady flow. The first and last terms clearly represent kinetic and gravitational potential energy of the fluid element. The role of internal energy (per mass) is here played by the enthalpy w in order to account for work done on the fluid element through compression.

Along a single stream line of the steady flow, parameterized by length ℓ, Bernoulli's law holds that

$$\frac{\partial\mathcal{B}}{\partial\ell} \;=\; u\frac{\partial u}{\partial\ell} + \frac{\partial w}{\partial\ell} + \frac{\partial\Psi}{\partial\ell} \;=\; 0 \;=\; u\frac{\partial u}{\partial\ell} + \frac{dw}{d\rho}\frac{\partial\rho}{\partial\ell} + \frac{\partial\Psi}{\partial\ell} \quad , \tag{3.10}$$

considering w to be a function of density alone. Differentiating each of the density-only expressions of the enthalpy function in eq. (3.6) gives

$$\frac{dw}{d\rho} \;=\; \frac{c_s^2}{\rho} \quad , \tag{3.11}$$

although in the adiabatic case the sound speed depends on density. This allows us to rewrite eq. (3.10)

$$u^2 \frac{\partial \ln|u|}{\partial \ell} + c_s^2 \frac{\partial \ln \rho}{\partial \ell} \;=\; -\frac{\partial \Psi}{\partial \ell} \quad . \tag{3.12}$$

We consider steady fluid flow whose structure is prescribed, but whose speed is not — that is what we are trying to find. Perhaps the fluid is flowing though some kind of duct or it is flowing along a magnetic flux tube — a magnetic duct. We consider a piece of the flow called a stream tube[1] thin enough that all properties are approximately constant over its cross sectional area, denoted, $A(\ell)$. The mass flow though this stream tube,

$$\dot{m} \;=\; \rho u A \quad , \tag{3.13}$$

must also be a constant along the flow: $\partial \dot{m}/\partial \ell = 0$. Our analysis of steady flows will essentially hinge on using two conserved quantities: $\mathcal{B}$ and $\dot{m}$. The derivative of $\ln \dot{m} = \ln(\rho u A)$ along the flow will vanish, leading to

$$\frac{\partial \ln \rho}{\partial \ell} \;=\; -\frac{\partial \ln|u|}{\partial \ell} - \frac{\partial \ln A}{\partial \ell} \tag{3.14}$$

Using this to eliminate the density derivative from expression (1.44) gives

$$(u^2 - c_s^2)\frac{\partial \ln|u|}{\partial \ell} \;=\; c_s^2 \frac{\partial \ln A}{\partial \ell} - \frac{\partial \Psi}{\partial \ell} \tag{3.15}$$

3B Laval's nozzle

Let us consider first the case of steady flow though a horizontal duct. Gravity plays no role in this case ($\partial\Psi/\partial\ell = 0$) so eq. (3.15) takes the simpler form

$$(u^2 - c_s^2)\frac{\partial \ln|u|}{\partial \ell} \;=\; c_s^2 \frac{\partial \ln A}{\partial \ell} \quad . \tag{3.16}$$

This simple relation shows how flow speed varies in response to variations in the cross sectional area of the duct. The remarkable fact revealed by this relation is that subsonic ($u^2 < c_s^2$) and super-sonic ($u^2 > c_s^2$) flows respond to a constriction ($\partial \ln A/\partial\ell < 0$) in completely opposite fashions. Examining the signs of the factors reveals that a subsonic flow *speeds up* at a constriction ($\partial \ln|u|/\partial\ell > 0$), while a super-sonic flow encountering the same constriction will *slow down* ($\partial \ln|u|/\partial\ell < 0$). Of course, a flaring duct ($\partial \ln A/\partial\ell > 0$) will cause the flows to do the opposite in each case: slow-down subsonically and speed up super-sonically.

[1] A bundle of stream lines

In either flow the product ρu must *increase* to compensate for a decrease in area A, according to eq. (3.13). Bernoulli's law, eq. (3.9), requires that ρ and u change in opposite senses. The differential of $\mathcal{B}$ at a fixed position is

$$\delta\mathcal{B} \;=\; 0 \;=\; u\;\delta u + \frac{dw}{d\rho}\delta\rho \;=\; u^2\left(\frac{\delta u}{u}\right) + c_s^2\left(\frac{\delta\rho}{\rho}\right) \quad , \tag{3.17}$$

using $dw/d\rho = c_s^2/\rho$; δu and $\delta\rho$ must have opposite signs. The one which undergoes the larger relative change must be the positive one in order to increase the product ρu. Which one this is, ρ or u, depends on whether the flow is subsonic or super-sonic.

In subsonic flows velocity changes dominate density changes, in a relative sense:

$$\left|\frac{\delta u}{u}\right| \;=\; \frac{c_s^2}{u^2}\left|\frac{\delta\rho}{\rho}\right| \;>\; \left|\frac{\delta\rho}{\rho}\right| \quad , \quad \text{when} \;\; u < c_s \quad , \tag{3.18}$$

based on eq. (3.17). This means that u must *increase* in order to keep the same mass flowing through the narrowing duct. (This conforms to our intuition since most of our day-to-day experience is with subsonic flows.) This increased speed demands a decrease in density, according to Bernoulli. According to eq. (3.18), however, the relative density decrease will be smaller by a factor $u^2/c_s^2 < 1$ than the increase in speed,[2] so the product ρu will increase.

In supersonic flow, on the other hand, density changes dominate velocity changes. The flow constriction drives up the density and thereby the enthalpy. This naturally creates a *decrease* in speed, since the flow is going against the pressure gradient. Nevertheless, the relative change in speed is smaller by a factor $c_s^2/u^2 < 1$ than the change in density, so the product ρu once again increases.

A similarly remarkable feature of eq. (3.16) is that at the exact point where the flow passes from sub-sonic to super-sonic (i.e. where $u^2 = c_s^2$) one of two things must be true. Either the flow speed must be discontinuous (this is called *a shock*) or the cross sectional area must be an extremum: $\partial \ln A/\partial\ell = 0$. It turns out that the flow speed always *decreases* at a standing shock, so the only way to *accelerate* a flow to supersonic speeds is to drive it through the throat of a nozzle – a minimum in area $A(\ell)$ — as shown in fig. 3.1. The sub-sonic flow will speed up as the area decreases ahead of the throat. It will pass the sonic point at the narrowest point of the nozzle (its throat) and then continue to speed up as the supersonic flow expands through the flaring opening.

The requirement that $u^2 = c_s^2$ at the throat of the nozzle constitutes a relation between the otherwise free constants $\dot{m}$ and $\mathcal{B}$. In a common circumstance, Bernoulli's constant is set by the state of the fluid in which the flow originates, perhaps a vessel of pressurized gas. If its origin is very subsonic we may approximate $\mathcal{B} \simeq w_0$, the enthalpy in the pressure vessel. The mass flux in the duct, $\dot{m}$, is then fixed by the area at the throat of the nozzle, A_t.

In the isothermal case, $w_0 = 0$, since $\rho = \rho_0$ at the origin. Matching this constant at the throat, where $u_t = c_s$, requires that

$$\tfrac{1}{2}u_t^2 + w_t \;=\; \tfrac{1}{2}c_s^2 + c_s^2\ln(\rho_t/\rho_0) \;=\; w_0 \;=\; 0 \quad . \tag{3.19}$$

[2]This scaling confirms the fact that extremely subsonic flow, $u^2 \ll c_s^2$, is incompressible. The relative change in density is so small that the density remains virtually constant.

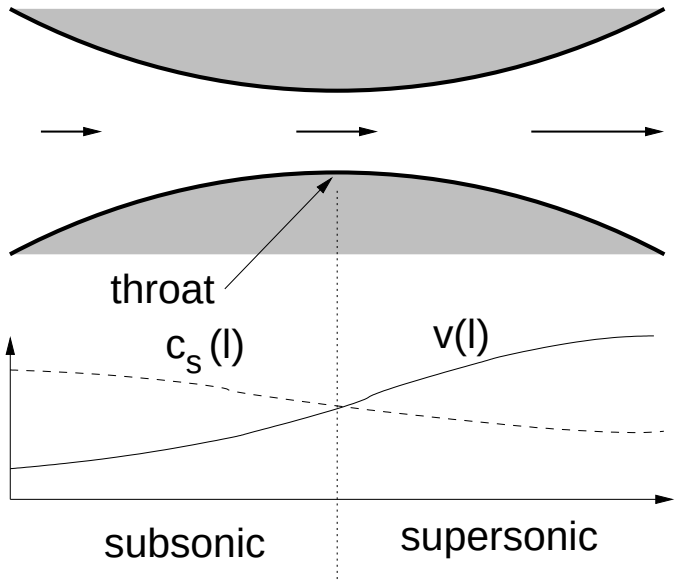

Figure 3.1: Transonic flow through a nozzle. The curves in the lower panel show the flow accelerating as it moves rightward (solid curve) and the sound speed decreasing as the enthalpy drops (dashed curve). The latter is for adiabatic flow; in isothermal flow the sound speed would be a horizontal line.

The mass density at the throat is therefore $\rho_t = e^{-1/2}\rho_0$, and the mass flux is

$$\dot{m} \;=\; \rho_t u_t A_t \;=\; e^{-1/2}\rho_0 c_s A_t \;=\; e^{-1/2}\sqrt{\rho_0 p_0}\, A_t \quad . \tag{3.20}$$

In the adiabatic case, where $w = c_s^2/(\gamma-1)$, Bernoulli's principle dictates that

$$w_0 \;=\; \tfrac{1}{2}c_{s,t}^2 + w_t \;=\; \tfrac{1}{2}(\gamma+1)w_t \quad . \tag{3.21}$$

Using the adiabatic version of eq. (3.6) lets us derive the fluid properties at the throat

$$\rho_t \;=\; \left(\frac{2}{\gamma+1}\right)^{1/(\gamma-1)} \rho_0 \tag{3.22}$$

$$c_{s,t} \;=\; \left(\frac{2}{\gamma+1}\right)^{1/2} c_{s0} \tag{3.23}$$

and from them the total mass flux

$$\dot{m} \;=\; \rho_t c_{s,t} A_t \;=\; K\sqrt{\rho_0 p_0}\, A_t \tag{3.24}$$

where

$$K \;=\; \sqrt{\gamma}\left(\frac{2}{\gamma+1}\right)^{(\gamma+1)/2(\gamma-1)} \;=\; \sqrt{\gamma}\left(1-\frac{\gamma-1}{\gamma+1}\right)^{(\gamma+1)/2(\gamma-1)} \quad . \tag{3.25}$$

The right-most form can be used to see that $K \to e^{-1/2}$ as $\gamma \to 1$, recovering the isothermal result as a special case. The left expression can be used to find $K = \sqrt{2}$ in the other special

case, $\gamma \to \infty$, which corresponds to the paradoxical situation of a transonic incompressible flow. In spite of its seeming self-contradiction, using the value in eq. (3.24) gives

$$\dot{m} = \sqrt{2p_0\rho_0}A_t = u_\infty \rho_0 A_t \quad ,$$

where $u \to u_\infty$ when $w \sim p \to 0$. This is a standard result for the flow of water from a hole at a depth d where $p_0 = \rho_0 g d$.

The mass flux everywhere in the duct is entirely determined by a product of the width of the nozzle's throat, A_t, and the factor

$$\sqrt{\rho_0 p_0} \;=\; \rho_0\, v_{\rm th,0} \quad , \tag{3.26}$$

where $v_{\rm th} = \sqrt{p/\rho}$ is the thermal speed in both adiabatic and isothermal cases. The coefficient of proportionality, K, depends only on γ; over its entire range of possible values $1 \le \gamma < \infty$, it varies by only a factor of 2.33.

3C Parker's wind

The foregoing showed that accelerating a flow to supersonic speed from subsonic origin requires a constriction in the flow. The solar wind is supersonic at Earth and appears to originate from subsonic flows in the lower corona. How is it accelerated to supersonic speeds by flows which seem to monotonically expand, $\partial \ln|A|/\partial\ell > 0$? There is no constriction or nozzle on the way to the Earth. The resolution becomes clear after rewriting (3.15) in the form

$$(u^2 - c_s^2)\frac{\partial \ln|u|}{\partial \ell} \;=\; c_s^2\left(\frac{\partial \ln A}{\partial \ell} - \frac{g}{c_s^2}\frac{\partial r}{\partial \ell}\right) \quad , \tag{3.27}$$

using the local gravitational acceleration $g = \partial\Psi/\partial r$. Even for a purely expanding flow the right hand side can be negative causing the subsonic flow to accelerate. Provided the local hydrostatic scale height, $H_p = c_s^2/g$, is smaller than the expansion scale-length, $(\partial \ln A/\partial\ell)^{-1}$, and the flow tube is vertical enough (i.e. $\partial r/\partial\ell$ close enough to unity), the right-hand-side (rhs) of (3.27) will be negative. In that case gravity can play the same role as a constriction, causing subsonic flow to accelerate even as it moves upward (outward). The flow will cross the sonic point where the expansion scale height precisely matches the inclined gravitational scale height, making the rhs vanish. For radial outward flow $\ell = r$ and $A(r) \sim r^2$, so the sonic point occurs where

$$\frac{2}{r} \;=\; \frac{g}{c_s^2} \;=\; \frac{g_0}{c_s^2}\left(\frac{R_\odot}{r}\right)^2 \quad . \tag{3.28}$$

At this point the local sound speed is one-half the escape speed at that radius

$$c_s = \tfrac{1}{2}\sqrt{\frac{2GM}{r}} \quad . \tag{3.29}$$

3D Radial Isothermal Wind: Graphical Analysis

Over the portion of the atmosphere where the flow remains very subsonic, $u \ll c_s$, we may ignore the velocity contribution to Bernoulli's law (eq. [3.9]) to get

$$w + \Psi \;=\; w_0 + \Psi_0 \quad . \tag{3.30}$$

The result is equivalent to requiring hydrostatic balance in an atmosphere where our dynamical conservation law, either isothermal or adiabatic evolution, has been applied spatially. The solutions are thus the ones from the earlier sections.

In the adiabatic case, accounting for the actual flow speed in Bernoulli's equation, $u > 0$ in eq. (3.9), will make the enthalpy, and thus the temperature, lower than its hydrostatic value. The temperature will then fall to zero even below the point it would vanish in a hydrostatic atmosphere. No adiabatic wind will be possible in that case, so expression (2.20) gives the surface condition for atmospheres which cannot produce an *adiabatic* wind. It is widely believed that the Sun's coronal temperature falls well inside this range, $T_0 < 4.6$ MK taking $\gamma = 5/3$, so we must conclude that the Sun's wind is not adiabatic — at least not all the way from the subsonic low corona. It seems that thermal conduction, which transports heat (i.e. $T\,Ds/Dt$) along field lines, is essential for driving the solar wind. We will assume that it is so effective (κ *is* very large for a million-degree plasma) that the atmosphere is *isothermal*.

There is an elegant graphical method of analyzing the solar wind solutions. Motivated by the discussion above we consider an isothermal flow in a spherical gravitational potential. In this case Bernoulli's law becomes

$$\tfrac{1}{2}u^2 + c_s^2 \ln(\rho/\rho_0) - \frac{GM}{r} \;=\; \mathcal{B} \;=\; -\frac{GM}{R_\odot} \;=\; -\tfrac{1}{2}v_{\rm esc,o}^2 \quad . \tag{3.31}$$

We use mass conservation for radial, outward flow to replace density by

$$\rho \;=\; \frac{\dot{M}}{4\pi r^2 u} \quad , \tag{3.32}$$

where $\dot{M}$ denotes the mass flux over the entire sphere. Placing this into expression (3.31) gives the relation

$$\begin{aligned} \frac{1}{2}\left(\frac{u}{c_s}\right)^2 - \ln\left(\frac{u}{c_s}\right) - 2\ln\left(\frac{r}{R_\odot}\right) + \frac{v_{\rm esc,o}^2}{2v_{\rm th}^2}\left[1 - \frac{R_\odot}{r}\right] \;&=\; F(r,u) \\ &=\; \ln\left(\frac{4\pi R_\odot^2 \rho_0 c_s}{\dot{M}}\right) \quad . \end{aligned} \tag{3.33}$$

If the mass flow, $\dot{M}$, were a free parameter we could find a solution to the flow problem, $u(r)$, from any contour of the function $F(r,u)$.

Contours of the function $F(r,u)$, shown in fig. 3.2, exhibit a single saddle point. This might have been predicted since the two-argument function, $F(r,u)$, is a simple sum of one-argument functions; a concave-up function of u plus a concave-down function of r.

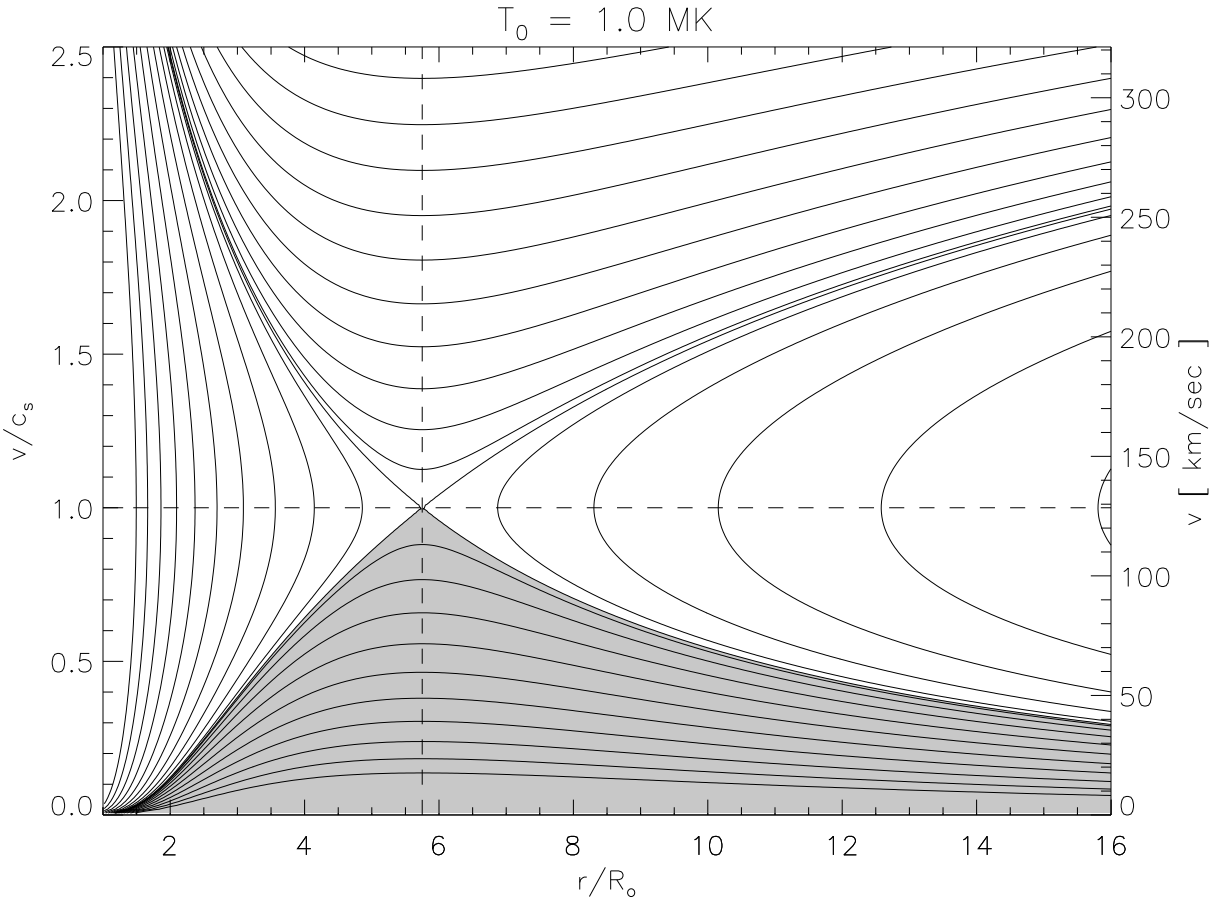

Figure 3.2: Contours of the function $F(r, u)$ for a radial isothermal wind at $T_0 = 1$ MK ($T_{\rm esc,o}/T_0 = 23$). The horizontal dashed line is the sound speed, $u = c_s$ and the vertical dashed line is the sonic radius, $r_x = (T_{\rm esc,o}/T_0)R_\odot/4 = 5.75R_\odot$. The grey shaded region contains all purely subsonic flows.

The saddle point occurs where $\partial F/\partial u = 0$ and $\partial F/\partial r = 0$ simultaneously. Setting to zero the first of these,

$$\frac{\partial F}{\partial u} = \frac{u}{c_s^2} - \frac{1}{u} \quad , \tag{3.34}$$

shows that the saddle occurs at the sonic point, $u = c_s$. Indeed, we see from the contours that at a given radius curves above the sonic line slope in the direction opposite to those below the sonic line. This is a manifestation of the opposing behaviors of subsonic and supersonic flows encountering a constriction. The curves become vertical at the exact sonic line because it is impossible for the flow to adjust the mass flux, ρu, to accommodate any variation. Thus no curves which cross the sonic line, except those that cross at the saddle, represent valid flows. Those curves remaining always beneath the sonic line, the shaded region in fig. 3.2, are known as the *solar breeze* solutions.

Setting to zero the other derivative of our two-argument function

$$\frac{\partial F}{\partial r} = -\frac{2}{r} + \frac{v_{\rm esc,o}^2}{2v_{\rm th}^2}\frac{R_\odot}{r^2} \tag{3.35}$$

shows that the sonic point occurs at radius

$$r_x \;=\; \frac{v_{\rm esc,o}^2}{4v_{\rm th}^2}\,R_\odot \;=\; \frac{T_{\rm esc,o}}{4T_0}\,R_\odot \quad , \tag{3.36}$$

consistent with eq. (3.28).

The purely subsonic flows (solar breeze solution in the shaded gray area) accelerate out to radius r_x and decelerate thereafter. Indeed, at very large radii these curves become very subsonic,

$$u \;\sim\; c_s\frac{R_\odot^2}{r^2} \quad , \quad r \gg r_x \quad . \tag{3.37}$$

Under this circumstance the density approaches the finite limit given by that of a hydrostatic atmosphere, namely eq. (2.15). This can be a very small fraction of the surface density,

$$\rho_\infty \;=\; \rho_0\, e^{-11.5} \;\simeq\; 10^{-5}\rho_0 \quad ,$$

in the case of a $T_0 = 1$ MK corona. Nevertheless, it says something very puzzling about the conditions near the solar surface: the density and/or temperature at the solar surface is fixed by the density of interplanetary (or interstellar) space. If this seems bizarre, it should! The conditions under which subsonic flows will occur require an equilibrium between the Sun and its surroundings that nobody believes to exist. We assume that the Sun is a structure in and of itself, it is not in equilibrium with its environment (at least not yet). It is for this reason that solar breeze solutions are generally rejected and only the transonic solution, *The Solar Wind* is accepted.

The transonic curve passes through the saddle point at $(r, u) = (r_x, c_s)$, exactly where the vertical and horizontal dashed curves cross one another. This occurs at contour level

$$F(r_x, c_s) \;=\; F_x \;=\; \frac{v_{\rm esc,o}^2}{2v_{\rm th}^2} \;-\; 2\ln\left(\frac{T_{\rm esc,o}}{4T_0}\right) \;-\; \frac{3}{2} \quad . \tag{3.38}$$

The separatrix curve at this level is the only transonic flow line, accelerating from subsonic over $r < r_x$ and the continuing to accelerate supersonically over $r > r_x$. Falling on this contour level requires that the mass flow be fixed at

$$\dot{M} \;=\; 4\pi R_\odot^2 c_s \rho_0 \, e^{-F_x} \;=\; 4\pi R_\odot^2 c_s \rho_0 \, e^{3/2} \left(\frac{T_{\rm esc,o}}{4T_0}\right)^2 e^{-T_{\rm esc,o}/2T_0} \quad . \tag{3.39}$$

In this way the mass flux is fixed by the requirement that the sonic point occur at the throat of the effective nozzle.

Isothermal evolution was the more plausible option for describing the acceleration of a solar wind to supersonic speeds. Its enthalpy function, eq. (3.6), is, however, unbounded from below — it becomes increasingly negative as density drops. As a result the velocity grows without bound at very large radii. The $r \to \infty$ limit of eq. (3.33), ignoring the $\ln u$ term,

$$u \;\sim\; 2c_s \sqrt{\ln(r/R_\odot)} \quad . \tag{3.40}$$

explains the logarithmic shape of the separatrix in fig. 3.2. Using this in eq. (3.32) gives a density

$$\rho \;\simeq\; \frac{\rho_0 R_\odot^2 e^{3/2}}{2r^2\sqrt{\ln(r/R_\odot)}} \left(\frac{T_{\rm esc,o}}{4T_0}\right)^2 e^{-T_{\rm esc,o}/2T_0} \;=\; \frac{\rho_\infty \, r_x^2 \, e^{3/2}}{2r^2\sqrt{\ln(r/R_\odot)}} \quad , \tag{3.41}$$

after using expression (2.15) for the asymptotic hydrostatic density ρ_∞. The final form shows clearly that the density of the transonic flow falls well below that of the subsonic solutions, for which $\rho \to \rho_\infty$. This ability to match the relative vacuum in distant space was the reason for accepting the transonic flow rather than the subsonic solar breeze solutions.

Perpetual acceleration is not, however, observed in the actual solar wind. So what happens instead? Recall that our justification for adopting isothermal evolution relied on the efficiency of thermal conduction at transporting heat. In the subsonic low corona temperature variations are smoothed away by thermal conduction very rapidly, and therefore very effectively. As the wind gathers speed, however, it eventually reaches a speed where thermal conduction becomes less effective against the temperature drops produced by expansion: $-(\gamma-1)T(\nabla\cdot\mathbf{u})$. The super-sonic solar wind is probably more appropriately described as adiabatic than isothermal. Finally, there is believed to be additional heat ($T\,Ds/Dt$) being deposited into the fluid along the way — for example by the dissipation of magnetoaccoustic waves. The isothermal or adiabatic winds form an essential framework in which to think about the solar wind, but the real solar wind is probably different from either.

Chapter 4

Linearized fluid equations

4A Nonlinear equations of motion — general & fluid

The equations for an ideal fluid (without viscosity or thermal conduction), subject only to gravity are

$$\begin{aligned} \partial_t \rho \ + \ \nabla \cdot (\mathbf{u}\rho) &= 0 && (4.1)\\ \rho\Big[\, \partial_t \mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u}) \,\Big] &= -\nabla p + \rho\, \mathbf{g} && (4.2)\\ \partial_t p + \mathbf{u}\cdot\nabla p &= -\gamma p \nabla\cdot\mathbf{u} \ . && (4.3) \end{aligned}$$

These represent equations for the time-evolution of the five fields defining the state of the fluid $\rho(\mathbf{x},t)$, $u_x(\mathbf{x},t)$, $u_y(\mathbf{x},t)$, $u_z(\mathbf{x},t)$, and $p(\mathbf{x},t)$. They are non-linear (see below) and therefore difficult/impossible to solve explicitly. Many other equations in Physics fall into this category and the procedure we are about to apply, linearization, is routinely applied to all of them. In every case we can abstractly describe the system in terms of a column vector of unknown functions of space and time. For a fluid the vector is a column of the five fields, $\underline{\mathrm{U}}(\mathbf{x},t) = [\rho, u_x, u_y, u_z, p]^T$. We will employ this abstract notation below in order to illustrate linearization in its essential form. We will also perform the linearization systematically to illustrate the general methodology.

The fluid equations can be recast to explicitly give the motion of its state vector

$$\partial_t \underline{\mathrm{U}} \ = \ \partial_t \begin{bmatrix} \rho \\ u_x \\ u_y \\ u_z \\ p \end{bmatrix} \ = \ \begin{bmatrix} \partial_t\rho &=& -\mathbf{u}\cdot\nabla\rho \ - \ \rho\nabla\cdot\mathbf{u} \\ \partial_t u_x &=& -\mathbf{u}\cdot\nabla u_x - \rho^{-1}\partial_x p + g_x \\ \partial_t u_y &=& -\mathbf{u}\cdot\nabla u_y - \rho^{-1}\partial_y p + g_y \\ \partial_t u_z &=& -\mathbf{u}\cdot\nabla u_z - \rho^{-1}\partial_z p + g_z \\ \partial_t p &=& -\mathbf{u}\cdot\nabla p - \gamma p\nabla\cdot\mathbf{u} \end{bmatrix} \ = \ \underline{F}\{\underline{\mathrm{U}};\mathbf{x}\} \ . \qquad (4.4)$$

The entire set of equations is represented by the vector-valued operator, operating on a vector argument $\underline{F}\{\underline{\mathrm{U}}\}$. The operator also depends explicitly on $\mathbf{x}$ through the gravitational field $\mathbf{g}(\mathbf{x})$, appearing in the equations, but we assume here it does not depend explicitly on time. $\underline{F}$ is an operator since it involves derivatives as well as values of its argument, $\underline{\mathrm{U}}$. Here it involves only first derivatives, but in other cases it may involve higher-order derivatives. The most significant aspect of $\underline{F}$ is that it depends *non-linearly* on its argument:

$$\underline{F}\{\alpha\underline{\mathrm{U}}\} \ \neq \ \alpha\underline{F}\{\underline{\mathrm{U}}\} \quad , \quad \underline{F}\{\underline{\mathrm{U}}_a + \underline{\mathrm{U}}_b\} \ \neq \ \underline{F}\{\underline{\mathrm{U}}_a\} + \underline{F}\{\underline{\mathrm{U}}_b\} \ .$$

This renders useless all methods involving superposition, the basis of most approaches to solving differential equations. It is for this reason that we must resort to *linearization*.

4A.1 Finding the equilibrium solution

The first step in linearization is to find an *equilibrium solution* to the equation. This is a solution to the full equations which does not depend on time. In hydrodynamics the equilibrium will consist of $\rho_0(\mathbf{x})$, $\mathbf{u}_0(\mathbf{x})$ and $p_0(\mathbf{x})$, which we denote by the column vector $\underline{\mathrm{U}}_0(\mathbf{x})$. To be an equilibrium it must satisfy the equations

$$\begin{bmatrix} -\mathbf{u}_0\cdot\nabla\rho_0 \;-\; \rho_0\nabla\cdot\mathbf{u}_0 \\ -\mathbf{u}_0\cdot\nabla u_{0,x} - \rho_0^{-1}\partial_x p_0 + g_x \\ -\mathbf{u}_0\cdot\nabla u_{0,y} - \rho_0^{-1}\partial_y p_0 + g_y \\ -\mathbf{u}_0\cdot\nabla u_{0,z} - \rho_0^{-1}\partial_z p_0 + g_z \\ -\mathbf{u}_0\cdot\nabla p_0 - \gamma p_0\nabla\cdot\mathbf{u}_0 \end{bmatrix} \;=\; \underline{F}\{\underline{\mathrm{U}}_0\} \;=\; 0 \quad . \tag{4.5}$$

Finding the equilibrium can be the most challenging aspect of the problem since it requires solving a set of coupled non-linear partial differential equations. In hydrodynamics, as in most other non-linear systems, there are many possible equilibrium solutions — typically an infinite number. There are static equilibria, where $\mathbf{u}_0 = 0$, and equilibria with steady flow: $\mathbf{u}_0(\mathbf{x}) \neq 0$. Static equilibria include isothermal atmospheres ($p_0 = K\rho_0$) and polytropic atmospheres ($p_0 = K'\rho_0^\Gamma$), among many others. Before we can proceed we must select one such solution to eq. (4.5) and call it $\underline{\mathrm{U}}_0$. We must always bear in mind that the linearization process we are about to undertake is a linearization about one particular equilibrium we chose; linearizing about a different equilibria will produce different spectra of normal modes.

4A.2 Linearizing the equations

Once the equilibrium solution has been selected, from among many possibilities, we propose a small perturbation to that solution. That means introducing perturbations to each and every field

$$\rho(\mathbf{x},t) \;=\; \rho_0(\mathbf{x}) + \rho_1(\mathbf{x},t) + \cdots \tag{4.6}$$
$$\mathbf{u}(\mathbf{x},t) \;=\; \mathbf{u}_0(\mathbf{x}) + \mathbf{u}_1(\mathbf{x},t) + \cdots \tag{4.7}$$
$$p(\mathbf{x},t) \;=\; p_0(\mathbf{x}) + p_1(\mathbf{x},t) + \cdots \quad , \tag{4.8}$$

where the subscript 1 signifies a small perturbation to the equilibrium denoted by subscript 0. Our assumption of small perturbation means $\rho_1 \ll \rho_0$ and $p_1 \ll p_0$. Were we to choose to linearize about a static equilibrium, $\mathbf{u}_0 = 0$, it would be less clear what $\mathbf{u}_1$ must be small compared to — we return to that issue below. In the abstract formalism our linearization amounts to a small perturbation to the equilibrium column vector

$$\underline{\mathrm{U}}(\mathbf{x},t) \;=\; \underline{\mathrm{U}}_0(\mathbf{x}) \;+\; \underline{\mathrm{U}}_1(\mathbf{x},t) + \cdots \quad . \tag{4.9}$$

We place expansions (4.6)–(4.8) into the governing equations and retain terms only up to first order in small parameters. For example the continuity equation (row 1 of eq. [4.4])

expands into several terms

$$\begin{aligned}
\partial_t\rho_0 + \partial_t\rho_1 &= \underbrace{-\mathbf{u}_0\cdot\nabla\rho_0 - \mathbf{u}_1\cdot\nabla\rho_0 - \mathbf{u}_0\cdot\nabla\rho_1 - \mathbf{u}_1\cdot\nabla\rho_1}_{-\mathbf{u}\cdot\nabla\rho} \\
&\quad \underbrace{-\rho_0\nabla\cdot\mathbf{u}_0 - \rho_1\nabla\cdot\mathbf{u}_0 - \rho_0\nabla\cdot\mathbf{u}_1 - \rho_1\nabla\cdot\mathbf{u}_1}_{-\rho\nabla\cdot\mathbf{u}} \\
&= \underbrace{-\mathbf{u}_0\cdot\nabla\rho_0 - \rho_0\nabla\cdot\mathbf{u}_0}_{\mathcal{O}(1)}\ \underbrace{-\mathbf{u}_1\cdot\nabla\rho_0 - \mathbf{u}_0\cdot\nabla\rho_1 - \rho_1\nabla\cdot\mathbf{u}_0 - \rho_0\nabla\cdot\mathbf{u}_1}_{\mathcal{O}(\epsilon)} \\
&\quad \underbrace{-\mathbf{u}_1\cdot\nabla\rho_1 - \rho_1\nabla\cdot\mathbf{u}_1}_{\mathcal{O}(\epsilon^2)}
\end{aligned}$$

where the terms in the final expression have been grouped according to magnitude. The largest of these, the $\mathcal{O}(1)$ terms, are the top row of equilibrium equation eq. (4.5), and therefore vanish. This, along with the requirement that $\partial_t\rho_0 = 0$, were the requirement for equilibrium; thus the first term on the left also vanishes.

Finally we invoke the assumed smallness of our perturbation to neglect the terms $\mathcal{O}(\epsilon^2)$ in comparison to the $\mathcal{O}(\epsilon)$ terms. Following this we are left with the linearized continuity equation

$$\partial_t\rho_1 \;=\; -\,\mathbf{u}_1\cdot\nabla\rho_0 - \mathbf{u}_0\cdot\nabla\rho_1 - \rho_1\nabla\cdot\mathbf{u}_0 - \rho_0\nabla\cdot\mathbf{u}_1 \quad . \tag{4.10}$$

The equilibrium fields $\rho_0(\mathbf{x})$ and $\mathbf{u}_0(\mathbf{x})$ should be considered known in this equation. The equation is therefore linear in the unknowns $\rho_1(\mathbf{x},t)$ and $\mathbf{u}_1(\mathbf{x},t)$.

While a general equilibrium solution may include a steady flow, $\mathbf{u}_0(\mathbf{x})$, we will hereafter simplify the algebra by considering the special case of *static* equilibrium solutions: $\mathbf{u}_0(\mathbf{x}) = 0$. The more general case follows the exact same steps, but with more terms in every expression. A static equilibrium must satisfy

$$\begin{bmatrix} 0 \\ -\rho_0^{-1}\partial_x p_0 + g_x \\ -\rho_0^{-1}\partial_y p_0 + g_y \\ -\rho_0^{-1}\partial_z p_0 + g_z \\ 0 \end{bmatrix} \;=\; \underline{F}\{\underline{\mathrm{U}}_0\} \;=\; 0 \quad , \tag{4.11}$$

which is equivalent to the hydrostatic equations $\nabla p_0 = \rho_0\mathbf{g}$. There are an infinite number of possible solutions to this equations (e.g. polytropic atmosphere of any polytropic index); we will become more specific later on.

Setting $\mathbf{u}_0 = 0$ in (4.10) yields a linear equation with half as many terms on the right-hand-side (the desired simplification)

$$\partial_t\rho_1 \;=\; -\,\mathbf{u}_1\cdot\nabla\rho_0 - \rho_0\nabla\cdot\mathbf{u}_1 \quad . \tag{4.12}$$

Performing the same steps on all rows of eq. (4.4) yields

$$\partial_t\underline{\mathrm{U}}_1 \;=\; \partial_t\begin{bmatrix} \rho_1 \\ u_{1,x} \\ u_{1,y} \\ u_{1,z} \\ p_1 \end{bmatrix} \;=\; \begin{bmatrix} \partial_t\rho_1 &=& -\mathbf{u}_1\cdot\nabla\rho_0 - \rho_0\nabla\cdot\mathbf{u}_1 \\ \partial_t u_{1,x} &=& -\rho_0^{-1}\partial_x p_1 + \rho_0^{-2}\rho_1\partial_x p_0 \\ \partial_t u_{1,y} &=& -\rho_0^{-1}\partial_y p_1 + \rho_0^{-2}\rho_1\partial_y p_0 \\ \partial_t u_{1,z} &=& -\rho_0^{-1}\partial_x p_1 + \rho_0^{-2}\rho_1\partial_x p_0 \\ \partial_t p_1 &=& -\mathbf{u}_1\cdot\nabla p_0 - \gamma p_0\nabla\cdot\mathbf{u}_1 \end{bmatrix} \quad .$$

This is a set of 5 equations linear in the five unknown perturbations. They may be formally cast into the form

$$\partial_t \underline{\mathrm{U}}_1 \;=\; \begin{bmatrix} 0 & -\partial_x\rho_0 - \rho_0\partial_x & -\partial_y\rho_0 - \rho_0\partial_y & -\partial_z\rho_0 - \rho_0\partial_z & 0 \\ \rho_0^{-2}\partial_x p_0 & 0 & 0 & 0 & -\rho_0^{-1}\partial_x \\ \rho_0^{-2}\partial_y p_0 & 0 & 0 & 0 & -\rho_0^{-1}\partial_y \\ \rho_0^{-2}\partial_z p_0 & 0 & 0 & 0 & -\rho_0^{-1}\partial_z \\ 0 & -\partial_x p_0 - \gamma p_0\partial_x & -\partial_y p_0 - \gamma p_0\partial_y & -\partial_z p_0 - \gamma p_0\partial_z & 0 \end{bmatrix} \cdot \begin{bmatrix} \rho_1 \\ u_{1,x} \\ u_{1,y} \\ u_{1,z} \\ p_1 \end{bmatrix} \;=\; \hat{\mathsf{F}}\cdot\underline{\mathrm{U}}_1$$

where $\hat{\mathsf{F}}$ is a matrix operator acting on the column vector of unknowns $\underline{\mathrm{U}}_1(\mathbf{x},t)$. Unlike the original vector operator, $\underline{F}\{\underline{\mathrm{U}}\}$, $\hat{\mathsf{F}}$ is a *linear operator*

$$\hat{\mathsf{F}}\cdot(\alpha\underline{\mathrm{U}}) \;=\; \alpha\hat{\mathsf{F}}\cdot\underline{\mathrm{U}} \quad , \quad \hat{\mathsf{F}}\cdot(\underline{\mathrm{U}}_a + \underline{\mathrm{U}}_b) \;=\; \hat{\mathsf{F}}\cdot\underline{\mathrm{U}}_a + \hat{\mathsf{F}}\cdot\underline{\mathrm{U}}_b \quad .$$

This is similar to operators used in quantum mechanics — except that it is a matrix of such operators.

This abstract form makes it clear that the resulting system is linear and homogeneous in the unknowns. Linearity is no accident since we have discarded all terms higher than first order. Homogeneity comes from our linearizing about a solution to the equilibrium equation. The whole process of linearization is a kind of Taylor expansion about an equilibrium

$$\partial_t(\underline{\mathrm{U}}_0 + \underline{\mathrm{U}}_1) \;=\; \underline{F}\{\underline{\mathrm{U}}_0 + \underline{\mathrm{U}}_1\} \;=\; \underbrace{\underline{F}\{\underline{\mathrm{U}}_0\}}_{=0} \;+\; \hat{\mathsf{F}}\cdot\underline{\mathrm{U}}_1 \;+\; \mathcal{O}(\epsilon^2) \quad . \tag{4.13}$$

The operator $\hat{\mathsf{F}}$ is akin to the derivative of the differential equations evaluated at the equilibrium solution. The linearization procedure then tells us about the dynamics "in the neighborhood" of the equilibrium. This does not tell us everything about the equations, but it tells us a great deal more than the equilibrium alone.

The original non-linear problem has now been reduced to a set of coupled linear PDEs formally written as

$$\partial_t\underline{\mathrm{U}} \;=\; \hat{\mathsf{F}}\cdot\underline{\mathrm{U}} \quad . \tag{4.14}$$

This is a linear equation, and our complete catalog of powerful solution techniques can be brought to bear on it. When concreteness is called for we will consider solving this as an initial value problem, with some initial condition $\underline{\mathrm{U}}_1(\mathbf{x},0)$.

4B Normal modes

The only functions of time or space appearing in the operator $\hat{\mathsf{F}}$ are the equilibrium $\rho_0(\mathbf{x})$ and $p_0(\mathbf{x})$. Since the equilibrium is, by definition, time-independent, these function do not actually depend on time $\partial_t\rho_0 = 0$ and $\partial_t p_0 = 0$. One consequence of this can be written using the traditional commutator

$$[\,\partial_t,\,\hat{\mathsf{F}}\,] \;=\; 0 \quad , \tag{4.15}$$

(this is actually a matrix of commuation relations). Since the operators $\hat{\mathsf{F}}$ and ∂_t commute, we can show that if $\underline{\mathrm{U}}_1(\mathbf{x},t)$ satisfies eq. (4.14) then $\partial_t\underline{\mathrm{U}}_1$ does as well. Taking the time derivative of both sides, and using the commutation of the operators gives

$$\partial_t(\partial_t\underline{\mathrm{U}}) \;=\; \partial_t(\hat{\mathsf{F}}\cdot\underline{\mathrm{U}}) \;=\; \hat{\mathsf{F}}\cdot(\partial_t\underline{\mathrm{U}}) \quad . \tag{4.16}$$

The left-most and right-most terms are the same as eq. (4.14) with $\underline{\mathrm{U}} \to (\partial_t \underline{\mathrm{U}})$. These are either two independent solutions or they are the same solution up to a multiplicative constant λ: $\partial_t \underline{\mathrm{U}}_1 = \lambda \underline{\mathrm{U}}_1$. We can restrict ourselves to the latter case and, for reasons to become clear, redefine the constant $\lambda = -i\omega$.

According to the above arguments, time-independence of $\hat{\mathsf{F}}$ means that there is some possibly complex constant ω for which

$$\partial_t \underline{\mathrm{U}}_1 \;=\; -\,i\omega \underline{\mathrm{U}}_1 \quad , \tag{4.17}$$

for some solution $\underline{\mathrm{U}}_1(\mathbf{x}, t)$ to eq. (4.14). This equation can be immediately solved to yield

$$\underline{\mathrm{U}}_1(\mathbf{x}, t) \;=\; \underline{\tilde{\mathrm{U}}}(\mathbf{x}) e^{-i\omega t} \quad , \tag{4.18}$$

where $\underline{\tilde{\mathrm{U}}}(\mathbf{x})$ is a (possibly complex) normal mode. This is not to say that every solution of (4.14) is oscillatory, but rather that it is *always possible* to find oscillatory solutions to eq. (4.14) provided $[\partial_t, \hat{\mathsf{F}}] = 0$. Linearity then allows us to superpose oscillatory solutions to construct general solutions. We have thus re-posed the time-dependent eq. (4.14) as the problem of finding the complete set of normal modes satisfying the equation

$$\hat{\mathsf{F}} \cdot \underline{\tilde{\mathrm{U}}} \;=\; -\,i\omega \underline{\tilde{\mathrm{U}}} \quad , \tag{4.19}$$

for normal modes $\underline{\tilde{\mathrm{U}}}(\mathbf{x})$ and eigenfrequencies ω. The complete set of eigenfrequencies are known as the *spectrum* of the system.

Expression (4.18) is often proposed as a solution by *ansatz*. We see above that it is actually a consequence of the time-independence of the operator $\hat{\mathsf{F}}$, which follows in turn from the time-independence of the equilibrium. Time-independent equilibria are the natural things to linearize about, so normal modes are the natural solutions to seek.

It is worth restating the solution we are now seeking

$$\rho(\mathbf{x}, t) \;=\; \rho_0(\mathbf{x}) + \tilde{\rho}(\mathbf{x}) e^{-i\omega t} + \cdots \tag{4.20}$$

$$\mathbf{u}(\mathbf{x}, t) \;=\; \tilde{\mathbf{u}}(\mathbf{x}) e^{-i\omega t} + \cdots \tag{4.21}$$

$$p(\mathbf{x}, t) \;=\; p_0(\mathbf{x}) + \tilde{p}(\mathbf{x}) e^{-i\omega t} + \cdots \quad . \tag{4.22}$$

where the normal mode functions, $\tilde{\rho}(\mathbf{x})$, $\tilde{\mathbf{u}}(\mathbf{x})$ and $\tilde{p}(\mathbf{x})$ are all assumed to be small, but we have dropped the subscript 1 for simplicity. Note that perturbations to density, velocity and pressure all oscillate at the same frequency ω; this is the definition of a normal mode.

4B.1 Complex modes and negative frequencies

Before finding the normal modes we consider the meaning of complex solutions to a problem which ultimately involves only real quantities. The operator $\hat{\mathsf{F}}$ is a matrix of operators involving only real quantities. Taking the complex conjugate[1] of all rows in eq. (4.19)

$$\hat{\mathsf{F}} \cdot \underline{\tilde{\mathrm{U}}}^* \;=\; i\omega^* \underline{\tilde{\mathrm{U}}}^* \quad , \tag{4.23}$$

since $\hat{\mathsf{F}}^* = \hat{\mathsf{F}}$. This means that if $\underline{\tilde{\mathrm{U}}}(\mathbf{x})$ is a normal mode with eigenfrequency ω then $\underline{\tilde{\mathrm{U}}}^*(\mathbf{x})$ is a normal mode with eigenfrequency $-\omega^*$. This means the spectrum contains both ω and $-\omega^*$.

[1] We are not constructing an adjoint, so we take the complex conjugate of each element in the matrix.

One option is for the eigenfrequency to be purely imaginary,[2] $\omega = i\sigma_m$ in which case ω and $-\omega^*$ are the same eigenfrequency and thus $\underline{\tilde{\mathrm{U}}}(\mathbf{x})$ and $\underline{\tilde{\mathrm{U}}}^*(\mathbf{x})$ are the same eigenmode: $\underline{\tilde{\mathrm{U}}}_m(\mathbf{x})$ must be purely real.[3] In the alternative an eigenfrequency ω_n with a non-vanishing positive real part demands the existence of a different eigenfrequency, $-\omega_n^*$ with a negative real part; both have the same imaginary part. These are two distinct eigenfrequencies which must have independent normal modes, $\underline{\tilde{\mathrm{U}}}_n(\mathbf{x})$ and $\underline{\tilde{\mathrm{U}}}_n^*(\mathbf{x})$. This means that the real and imaginary parts, $\underline{\mathrm{U}}_n^r(\mathbf{x})$ and $\underline{\mathrm{U}}_n^i(\mathbf{x})$, are themselves independent functions. The normal mode cannot, therefore, be purely real.

4B.2 Superposition of normal modes

A general solution of eq. (4.14) consists of a superposition of normal modes

$$\underline{\mathrm{U}}_1(\mathbf{x},t) \;=\; \sum_m A_m\,\underline{\tilde{\mathrm{U}}}_m(\mathbf{x})e^{\sigma_m t} \;+\; \sum_n \left[B_n\,\underline{\tilde{\mathrm{U}}}_n(\mathbf{x})e^{-i\omega_n t} + B_n^*\,\underline{\tilde{\mathrm{U}}}_n^*(\mathbf{x})e^{i\omega_n^* t}\right] \quad . \tag{4.24}$$

where A_m and B_n are constant coefficients; A_m are real and B_n are complex. It can easily be verified that this superposition of complex factors results in a sum which is always real. This is essential since the elements of $\underline{\mathrm{U}}_1(\mathbf{x},t)$, namely $\rho_1(\mathbf{x},t)$, $u_{1,x}(\mathbf{x},t)$, etc., are always real. In this respect negative frequencies (or frequencies with negative real parts) are essential to producing a real solution to eq. (4.14).

Consider solving eq. (4.14) as an initial value problem. Evaluating the superposition at $t = 0$ gives

$$\begin{aligned}\underline{\mathrm{U}}_1(\mathbf{x},0) \;&=\; \sum_m A_m\,\underline{\tilde{\mathrm{U}}}_m(\mathbf{x}) \;+\; \sum_n \left[B_n\,\underline{\tilde{\mathrm{U}}}_n(\mathbf{x}) + B_n^*\,\underline{\tilde{\mathrm{U}}}_n^*(\mathbf{x})\right] \\ &=\; \sum_m A_m\,\underline{\tilde{\mathrm{U}}}_m(\mathbf{x}) \;+\; 2\sum_n \left[B_n^r\,\underline{\tilde{\mathrm{U}}}_n^r(\mathbf{x}) - B_n^i\,\underline{\tilde{\mathrm{U}}}_n^i(\mathbf{x})\right] \quad ,\end{aligned} \tag{4.25}$$

where $B_n = B_r + iB_n^i$ and $\tilde{U}_n(\mathbf{x}) = \tilde{U}_n^r(\mathbf{x}) + i\tilde{U}_n^i(\mathbf{x})$. The real constants, A_m, B_n^r and B_n^i are set by matching the initial condition. Matching a general vector of functions will require the *complete set* of normal modes.[4] This illustrates, at least schematically, how the general initial value problem can be solved from the entire spectrum of normal modes.

4B.3 Instability

It has been noted that the eigenfrequencies may be imaginary or have an imaginary part: either $\sigma_m \neq 0$ or $\mathrm{Im}(\omega_n) \neq 0$. If a particular mode has a frequency whose imaginary part is positive then its contribution to the superposition, eq. (4.24), will grow exponentially in time. This is not unphysical and actually happens routinely; that particular mode is called *unstable*. Provided the unstable mode appears in the superposition, eq. (4.24), with a non-vanishing coefficient A_m or B_n — no matter how small — its steady growth will make it the dominant term after some time. The exponential growth will then make the whole expansion

[2] This option includes the possibility $\omega = \sigma_m = 0$.

[3] Actually, it might have any complex phase, but it is always possible, and usually easiest, to choose the phase that makes it purely real.

[4] We forego the formal process of demonstrating that this is possible through completeness of the normal modes.

large enough to violate the conditions for linearization. The only alternative would be for the initial condition to be *orthogonal* to all unstable modes so that their coefficients were exactly zero. This would require such incomprehensibly delicate balancing we dismiss it as an irrelevant possibility. We therefore conclude that: *if an equilibrium contains even a single unstable normal mode then the equilibrium itself is unstable.*

4C Homogeneous equilibria and plane waves

We found above that we could anticipate the form of solution to eq. (4.14) by knowing that the equilibrium did not depend explicitly on t. Using the fact $[\partial_t, \hat{\mathsf{F}}] = 0$ we were able to propose solutions oscillating in time as $e^{-i\omega t}$. We can make similar progress when one or more spatial coordinates are absent from the equilibrium. For example, if the equilibrium does not depend explicitly on x then $[\partial_x, \hat{\mathsf{F}}] = 0$ and there will be normal modes $\underline{\tilde{\mathrm{U}}}(x, y, z) \propto e^{ik_x x}$ where k_x is some possibly complex constant.

Extending this reasoning makes it clear that the simplest equilibrium to linearize about is a *homogeneous* equilibrium: one that does not depend on any spatial variables. We will hereafter restrict ourselves to this equilibrium: *a homogeneous equilibrium* in which ρ_0 and p_0 are both uniform ($\nabla\rho_0 = \nabla p_0 = 0$). We choose this equilibrium for the same reason it is usually chosen: it makes the analysis easy because we already know the eigenmodes: plane waves.

In order for the homogeneous state to count as a solution to the equilibrium equation, eq. (4.5), gravity must be neglected: $\mathbf{g} \simeq 0$. This is the appropriate approximation for cases of very large pressure scale heights. Including gravity, on the other hand, will lead to a *stratified atmosphere* where ρ_0 and p_0 will depend on radius or height. This is a far more complex problem not solvable with plane waves.

For any homogeneous equilibrium the operator $\hat{\mathsf{F}}$ will commute with every spatial derivative and the normal mode can be written as a plane wave

$$\underline{\mathrm{U}}(\mathbf{x}) \;=\; \underline{\hat{\mathrm{U}}}\, e^{i\mathbf{k}\cdot\mathbf{x}} \quad , \tag{4.26}$$

where $\underline{\hat{\mathrm{U}}}$ is a column vector of complex elements. This means that we are seeking solutions

$$\rho(\mathbf{x}, t) \;=\; \rho_0 \;+\; \hat{\rho}\, e^{i\mathbf{k}\cdot\mathbf{x} - i\omega t} + \cdots \tag{4.27}$$

$$\mathbf{u}(\mathbf{x}, t) \;=\; \hat{\mathbf{u}}\, e^{i\mathbf{k}\cdot\mathbf{x} - i\omega t} + \cdots \tag{4.28}$$

$$p(\mathbf{x}, t) \;=\; p_0 \;+\; \hat{p}\, e^{i\mathbf{k}\cdot\mathbf{x} - i\omega t} + \cdots \quad . \tag{4.29}$$

The factors $\hat{\rho}$, $\hat{u}_x$, ..., $\hat{p}$, are constants, possibly complex and small. The perturbation to the homogeneous equilibrium is a plane wave with wave number $\mathbf{k}$ and (possibly complex) frequency ω.

Placing the plane wave solution into the normal mode equation, eq. (4.19), turns the set of eigenfunction equations into an algebraic system

$$-i\omega\underline{\hat{\mathrm{U}}} \;=\; \begin{bmatrix} 0 & -ik_x\rho_0 & -ik_y\rho_0 & -ik_z\rho_0 & 0 \\ 0 & 0 & 0 & 0 & -ik_x\rho_0^{-1} \\ 0 & 0 & 0 & 0 & -ik_y\rho_0^{-1} \\ 0 & 0 & 0 & 0 & -ik_z\rho_0^{-1} \\ 0 & -ik_x\gamma p_0 & -ik_y\gamma p_0 & -ik_z\gamma p_0 & 0 \end{bmatrix} \cdot \underline{\hat{\mathrm{U}}} \;=\; \underline{\underline{F}} \cdot \underline{\hat{\mathrm{U}}} \quad . \tag{4.30}$$

The matrix of operators, $\hat{\mathsf{F}}$ has been replaced by a 5×5 matrix $\underline{\underline{F}}(\mathbf{k})$ depending on the wave vector defining the plane wave. Equation (4.30) is a simple eigenvalue equation. It will have five eigenfrequencies, $\omega_n(\mathbf{k})$, and corresponding eigenvectors $\hat{\underline{\mathsf{U}}}^{(n)}(\mathbf{k})$, both also depending on the wave vector.

4C.1 Why wave vectors are real and frequencies complex

The absence of spatial coordinates from the equilibrium allowed us to immediately pose the plane-wave solution $e^{i\mathbf{k}\cdot\mathbf{x}-i\omega t}$. In this expression the wave numbers, k_x, k_y and k_z appear to play identical roles to the frequency ω. It would thus seem that if one of them might be complex (ω) then all might be. This analogy breaks down, however, in most practical applications.

Consider first the prospect of solving the linearized problem in an unbounded domain, $x \in (-\infty, \infty)$, as an initial value problem. We solve for normal modes (plane waves) and then superpose them to match the initial condition. The normal modes must remain finite over the unbounded domain in order to be useful in our superposition. To achieve this we require that k_x, k_y and k_z be purely real. Indeed, the expansion in normal modes is nothing more than a Fourier transform (albeit in three dimensions) and all wave numbers are real in a Fourier transform.

We evaluate this expansion at times $t \geq 0$, but recognize that, in the case of instability, the solution may grow without bound. Thus there is no requirement that ω be real. Here we see that application to an initial value problem, $t \in (0, \infty)$, in an unbounded domain, $x \in (-\infty, \infty)$, breaks the analogy: k_x will be real while ω can be complex. In effect, we have performed a *Laplace* transform in time, but using $-i\omega$ in place of the traditional s variable. It is permissible for s to be complex, and so it is in our application.

There is an alternative problem where the system is driven from one boundary, $x = 0$, over all time: $t \in (-\infty, \infty)$. In this case ω will be real and k_x can be complex. If the other directions are unbounded then k_y and k_z are still real.

4D The hydrodynamic waves

Equation (4.30) is of a familiar type, an eigenvalue equation, and thus we know how to proceed from here, at least in principle. It is not so easy in practice because it is not simple to find eigenvalues and eigenvectors of a 5×5 matrix in general. Happily for us, it is possible to identify, by inspection, three different eigenvectors satisfying eq. (4.30). These are all eigenvectors with eigenfrequency $\omega = 0$, meaning they are *null vectors* of the matrix $\underline{\underline{F}}$. The first

$$\hat{\underline{\mathsf{U}}}^{(1)} = \begin{bmatrix} \hat{\rho} \\ \hat{u}_x \\ \hat{u}_y \\ \hat{u}_z \\ \hat{p} \end{bmatrix} = \begin{bmatrix} \rho_0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{4.31}$$

is known as the *entropy mode*. The other two, called *shear modes*, have velocities perpendicular to the wave vector. Define two unit vectors $\hat{\mathbf{m}}$ and $\hat{\mathbf{n}}$ which form, along with the

direction of $\mathbf{k}$, a right-handed triad. The two column vectors

$$\underline{\hat{\mathrm{U}}}^{(2)} = \begin{bmatrix} 0 \\ \hat{m}_x \\ \hat{m}_y \\ \hat{m}_z \\ 0 \end{bmatrix} \quad , \quad \underline{\hat{\mathrm{U}}}^{(3)} = \begin{bmatrix} 0 \\ \hat{n}_x \\ \hat{n}_y \\ \hat{n}_z \\ 0 \end{bmatrix} \quad , \tag{4.32}$$

can be seen to be eigenvectors of $\underline{\underline{F}}$ with $\omega = 0$ after using the fact that $\hat{\mathbf{m}} \cdot \mathbf{k} = \hat{\mathbf{n}} \cdot \mathbf{k} = 0$. Note that in these two modes the velocity vector is parallel to either $\hat{\mathbf{m}}$ or $\hat{\mathbf{n}}$ and is therefore perpendicular to $\mathbf{k}$. Shear modes are therefore transverse modes; the two modes represent two polarizations.

This leaves two other eigenvectors which are not so evident. These will turn out to have non-zero real eigenfrequencies which, following discussion above, must have equal magnitude and opposite sign. Since the shear modes are transverse modes we propose a longitudinal mode with velocity vector parallel to $\mathbf{k}$. We can simplify the matrices by taking the wave vector along $\hat{\mathbf{x}}$: $\mathbf{k} = k\hat{\mathbf{x}}$. With this choice the third and fourth rows and columns of $\underline{\underline{F}}$ vanish. Omitting them leaves the 3×3 reduced system

$$\omega \begin{bmatrix} \hat{\rho} \\ \hat{u}_x \\ \hat{p} \end{bmatrix} = \begin{bmatrix} 0 & k\rho_0 & 0 \\ 0 & 0 & k\rho_0^{-1} \\ 0 & \gamma k p_0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \hat{\rho} \\ \hat{u}_x \\ \hat{p} \end{bmatrix} \quad . \tag{4.33}$$

When we choose $\mathbf{k}$ along $\hat{\mathbf{x}}$ the perpendiculars are $\hat{\mathbf{m}} = \hat{\mathbf{y}}$ and $\hat{\mathbf{n}} = \hat{\mathbf{z}}$. The two shear modes have eigenvectors with non-vanishing $\hat{u}_y$ and $\hat{u}_z$ respectively. These modes are thus omitted when rows three and four are dropped from $\underline{\hat{\mathrm{U}}}$ to make the reduced 3×3 system.

The eigenfrequencies of this reduced system will satisfy the characteristic equation

$$\det \begin{bmatrix} -\omega & k\rho_0 & 0 \\ 0 & -\omega & k\rho_0^{-1} \\ 0 & \gamma k p_0 & -\omega \end{bmatrix} = -\omega(\omega^2 - \gamma k^2 p_0/\rho_0) = 0 \quad . \tag{4.34}$$

This is a cubic with three roots. The first,

$$\omega = 0 \quad \Longrightarrow \quad \begin{bmatrix} \hat{\rho} \\ \hat{u}_x \\ \hat{p} \end{bmatrix} = \begin{bmatrix} \rho_0 \\ 0 \\ 0 \end{bmatrix} \tag{4.35}$$

is the entropy mode, which was not eliminated in the reduction. The other two

$$\omega = \pm\sqrt{\frac{\gamma p_0}{\rho_0}}k = \pm\, c_s k \quad \Longrightarrow \quad \begin{bmatrix} \hat{\rho}_1 \\ \hat{u}_{1,x} \\ \hat{p}_1 \end{bmatrix} = \begin{bmatrix} \rho_0 \\ \pm c_s \\ \gamma p_0 \end{bmatrix} \tag{4.36}$$

are *acoustic modes* or sound waves and $c_s = \sqrt{\gamma p_0/\rho_0}$ is the speed of sound. As predicted there is one acoustic mode with positive frequency and one with negative frequency.

Returning to the full 5-dimensional system gives the general acoustic eigenvectors

$$\underline{\hat{\mathrm{U}}}^{(4)} \;=\; \begin{bmatrix} \rho_0 \\ \hat{k}_x c_s \\ \hat{k}_y c_s \\ \hat{k}_z c_s \\ \gamma p_0 \end{bmatrix} \quad , \quad \underline{\hat{\mathrm{U}}}^{(5)} \;=\; \begin{bmatrix} \rho_0 \\ -\hat{k}_x c_s \\ -\hat{k}_y c_s \\ -\hat{k}_z c_s \\ \gamma p_0 \end{bmatrix} \quad , \tag{4.37}$$

where $\hat{\mathbf{k}} = \mathbf{k}/|\mathbf{k}|$ is the unit vector along the wave vector (no longer $\hat{\mathbf{x}}$). Using $\underline{\hat{\mathrm{U}}}^{(4,5)}$ in eq. (4.30) gives the eigenfrequencies

$$\omega_4(\mathbf{k}) \;=\; c_s|\mathbf{k}| \quad , \quad \omega_5(\mathbf{k}) \;=\; -\,c_s|\mathbf{k}| \;\; . \tag{4.38}$$

Clearly mode 4 is the positive frequency branch and mode 5 provides the negative frequencies needed to yield real results. In a plane wave these frequencies produce traveling waves

$$\underline{\mathrm{U}}_1(\mathbf{x},t) \;=\; \underline{\hat{\mathrm{U}}}^{(4,5)}(\mathbf{k})\, e^{i\mathbf{k}\cdot\mathbf{x} - i\omega_{4,5}t} \;=\; \underline{\hat{\mathrm{U}}}^{(4,5)}(\mathbf{k})\, \exp\Big[\, ik(\hat{\mathbf{k}}\cdot\mathbf{x} \mp c_s t)\Big] \quad , \tag{4.39}$$

where the upper (lower) sign correspond to mode 4 (5). Phase surfaces for mode 4 move at the sound speed in direction $+\hat{\mathbf{k}}$, while mode 5 moves along $-\hat{\mathbf{k}}$. A single traveling wave is produced by pairing $\underline{\hat{\mathrm{U}}}^{(4)}(\mathbf{k})e^{i\mathbf{k}\cdot\mathbf{x}}$ with $\underline{\hat{\mathrm{U}}}^{(5)}(-\mathbf{k})e^{-i\mathbf{k}\cdot\mathbf{x}}$. The latter can be seen to be the complex conjugate of the former so the superposition will produce a real wave traveling in the $+\hat{\mathbf{k}}$ direction.

These two eigenvectors supplement eqs. (4.31) and (4.32) to form a complete set of five eigenvectors. For each wave vector $\mathbf{k}$ the five eigenvectors, $\underline{\hat{\mathrm{U}}}^{(n)}$, form five independent vectors spanning the 5-dimensional space; they are not, however, orthogonal. While sound wave are the most well-known fluid wave, it would be impossible to match an initial condition — an arbitrary 5-dimensional vector — using those two modes alone.

The vectors $\hat{U}^{(n)}$ are not normalized. In fact, the rows of $\underline{\mathrm{U}}_1$ have different units, so the idea of normalization is not even meaningful. The expressions of $\hat{U}^{(1)}$, $\hat{U}^{(4)}$ and $\hat{U}^{(5)}$ all use rows with the same units as $\underline{\mathrm{U}}_1$, so coefficients in a superposition will be dimensionless. It is evident that tese dimensionless coefficients will need to be small compared to unity in order for $\rho_1 \ll \rho_0$ and $p_1 \ll p_0$. This will also make the velocity perturbation $|\mathbf{u}_1| \ll c_s$. While there is no equilibrium flow, it seems that the velocity perturbation must be very sub-sonic to qualify as a small perturbation.

4D.1 The waves

The analysis above has revealed the three types of waves present in a homogeneous fluid: entropy waves, shear waves and acoustic waves. The first two of these have zero frequency and are often ignored. They are, however, an essential part of the general linearization of the fluid equations. The system has five fields and therefore must have five distinct normal modes. Had we dropped the zero-frequency modes we would come up three modes short. As a result we would not be able to match a general initial conditions.

The first thing to recognize is that modes with zero frequency are legitimate eigenmodes. They do not oscillate because there is no "restoring force". This occurs when there

are multiple equilibrium solutions in the vicinity of the one selected for linearization. A perturbation which creates a new, different equilibrium will not result in any subsequent evolution: the system will simply remain in its new equilibrium state.

The zero-frequency modes in fluids arise from the multiplicity of equilibrium solutions. Dropping gravity the equation for a static ($\mathbf{u}_0 = 0$) equilibrium becomes $\nabla p_0 = 0$. This is satisfied by a homogeneous pressure, but the density is completely free: any $\rho_0(\mathbf{x})$ is an equilibrium. We selected the homogeneous density, $\nabla\rho_0 = 0$ in order to obtain plane wave solutions. If we now make an initial perturbation to the density but not the pressure or velocity we change to a different, inhomogeneous equilibrium solution. Therefore, no change will occur in response to this perturbation: the perturbation will persist.

The case is similar for the shear modes. We opted for static equilibria, $\mathbf{u}_0 = 0$, in order to simplify analysis. It is possible, however, to find equilibria with flows. The shear mode solution suggests the form of such an equilibrium

$$\mathbf{u}_0(x, y, z) \;=\; A\hat{\mathbf{x}}\,\cos(k_y y) \quad , \tag{4.40}$$

where A and k_y are constants. For uniform pressure and density, $\nabla p_0 = \nabla\rho_0 = 0$, this flow satisfies the equilibrium eq. (4.5). This is a shear flow and there is no fluid force opposing it — except viscosity which we have chosen to neglect. Linear shear modes represent shear flows of small amplitude, $A \ll c_s$. Since these are equilibria then there is no force restoring the system back toward a static equilibrium.

Shear modes are the set of *transverse* waves meaning they are all incompressible:

$$\nabla\cdot\mathbf{u}_1 \;\to\; i\mathbf{k}\cdot\hat{\mathbf{u}} \;=\; 0 \quad . \tag{4.41}$$

The fact that these flows represent a zero-frequency space underlies the critical distinction made between *compressible* and *incompressible* hydrodynamics. Acoustic modes represent the high-frequency responses of a fluid. If one is interested only in the low-frequency (slow, subsonic) response of the fluid one may restrict one's attention to *incompressible* hydrodynamics. The linear dynamics of incompressible flows are trivial (they all have zero frequency) so any worthwhile treatment will be *non-linear*. Incompressible hydrodynamics is therefore completely non-linear and therefore more mathematically challenging than compressible hydrodynamics, in general.

4D.2 Left eigenvectors

Since the matrix in eq. (4.30) is not symmetric each eigenfrequency, ω_n will have a right eigenvector, $\underline{\hat{\mathrm{U}}}^{(n)}$, and a possibly different *left eigenvector* $\underline{\hat{\mathrm{W}}}^{(n)}$ satisfying

$$\underline{\hat{\mathrm{W}}}^{(n)}\cdot\underline{\underline{F}} \;=\; -i\omega_n\,\underline{\hat{\mathrm{W}}}^{(n)} \quad . \tag{4.42}$$

The left eigenvector of matrix is the right eigenvector of its transpose, so the distinction is only needed when the matrix is not symmetric — does not equal its own transpose. Left and right eigenvectors with different eigenvalues are orthogonal to one another

$$\underline{\hat{\mathrm{W}}}^{(m)}\cdot\underline{\hat{\mathrm{U}}}^{(n)} \;=\; 0 \quad , \quad \text{if } \omega_m \neq \omega_n \quad .$$

It is possible to combine degenerate eigenvectors of a symmetric matrix to be mutually orthogonal. For the non-symmetric case it is possible to combine degenerate left eigenvectors to make a set orthogonal to all other right eigenvectors; i.e. to make $\underline{\hat{\mathrm{W}}}^{(m)}$ orthogonal to all $\underline{\hat{\mathrm{U}}}^{(n)}$ for which $m \neq n$.

The taking the transpose of $i\underline{\underline{F}}$ leads to the equation

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ k_x\rho_0 & 0 & 0 & 0 & \gamma k_x p_0 \\ k_y\rho_0 & 0 & 0 & 0 & \gamma k_y p_0 \\ k_z\rho_0 & 0 & 0 & 0 & \gamma k_z p_0 \\ 0 & k_x\rho_0^{-1} & k_y\rho_0^{-1} & k_z\rho_0^{-1} & 0 \end{bmatrix} \cdot \underline{\hat{\mathrm{W}}}^{(n)} \;=\; \omega_n \, \underline{\hat{\mathrm{W}}}^{(n)} \quad , \tag{4.43}$$

for left eigenvector $\underline{\hat{\mathrm{W}}}^{(n)}$. The null eigenvalues, $\omega_1 = \omega_2 = \omega_3 = 0$, have left eigenvectors

$$\underline{\hat{\mathrm{W}}}^{(1)} \;=\; \begin{bmatrix} \gamma\rho_0^{-1} \\ 0 \\ 0 \\ 0 \\ -p_0^{-1} \end{bmatrix} \quad , \quad \underline{\hat{\mathrm{W}}}^{(2)} \;=\; \begin{bmatrix} 0 \\ \hat{m}_x \\ \hat{m}_y \\ \hat{m}_z \\ 0 \end{bmatrix} \quad , \quad \underline{\hat{\mathrm{W}}}^{(3)} \;=\; \begin{bmatrix} 0 \\ \hat{n}_x \\ \hat{n}_y \\ \hat{n}_z \\ 0 \end{bmatrix} \quad , \tag{4.44}$$

defined to be orthogonal to other right eigenvectors; here $\hat{\mathbf{m}}$ and $\hat{\mathbf{n}}$ are mutually orthogonal unit vectors which are also orthogonal to $\hat{\mathbf{k}}$. Note that $\underline{\hat{\mathrm{W}}}^{(1)}$ is orthogonal to right eigenvectors, $\underline{\hat{\mathrm{U}}}^{(4)}$ and $\underline{\hat{\mathrm{U}}}^{(5)}$, defined in eq. (4.37). It is also clearly orthogonal to $\underline{\hat{\mathrm{U}}}^{(2)}$ and $\underline{\hat{\mathrm{U}}}^{(3)}$. The acoustic modes, $\omega_{4,5} = \pm c_s|\mathbf{k}|$ has left eigenvectors

$$\underline{\hat{\mathrm{W}}}^{(4)} \;=\; \begin{bmatrix} 0 \\ \hat{k}_x\rho_0 c_s \\ \hat{k}_y\rho_0 c_s \\ \hat{k}_z\rho_0 c_s \\ 1 \end{bmatrix} \quad , \quad \underline{\hat{\mathrm{W}}}^{(5)} \;=\; \begin{bmatrix} 0 \\ \hat{k}_x\rho_0 c_s \\ \hat{k}_y\rho_0 c_s \\ \hat{k}_z\rho_0 c_s \\ -1 \end{bmatrix} \quad . \tag{4.45}$$

Left eigenvectors are the easiest way to determine the coefficients in the superposition of normal modes defining a full solution,

$$\underline{\hat{\mathrm{U}}}(\mathbf{k},t) \;=\; \sum_n \hat{A}_n(\mathbf{k})\, \underline{\hat{\mathrm{U}}}^{(n)}(\mathbf{k})\, e^{-i\omega_n(\mathbf{k})\,t} \quad , \tag{4.46}$$

where

$$\underline{\mathrm{U}}(\mathbf{x},t) \;=\; \int \underline{\hat{\mathrm{U}}}(\mathbf{k},t)\, e^{i\mathbf{k}\cdot\mathbf{x}}\, d^3x \quad . \tag{4.47}$$

The coefficients, $\hat{A}_n(\mathbf{k})$ are found from the initial conditions, as in eq. (4.25). Multiplying on the left by $\underline{\hat{\mathrm{W}}}^{(m)}$, and using its orthogonality to all $\underline{\mathrm{U}}^{(n)}$ where $n \neq m$, yields

$$\hat{A}_m(\mathbf{k}) \;=\; \frac{\underline{\hat{\mathrm{W}}}^{(m)} \cdot \underline{\hat{\mathrm{U}}}(\mathbf{k},0)}{\underline{\hat{\mathrm{W}}}^{(m)} \cdot \underline{\hat{\mathrm{U}}}^{(m)}} \quad . \tag{4.48}$$

Thus mode m is present when the initial perturbation is not orthogonal to the left eign-evector $\underline{\hat{\mathrm{W}}}^{(m)}$.

Consider the dot product of an initial perturbation with $\underline{\hat{\mathrm{W}}}^{(1)}$

$$\underline{\hat{\mathrm{W}}}^{(1)} \cdot \underline{\hat{\mathrm{U}}} \;=\; \left[\gamma\rho_0^{-1},\, 0,\, 0,\, 0,\, -p_0^{-1}\right] \cdot \begin{bmatrix} \hat{\rho} \\ \hat{u}_x \\ \hat{u}_y \\ \hat{u}_z \\ \hat{p} \end{bmatrix} \;=\; \gamma\frac{\hat{\rho}}{\rho_0} - \frac{\hat{p}}{p_0} \quad . \tag{4.49}$$

This is the proportional to the fist-order perturbation of the entropy. In an expansion of expression (1.12)

$$\begin{aligned} s \;&=\; \frac{k_\mathrm{B}}{m(\gamma-1)} \ln\left(\frac{p}{\rho^\gamma}\right) \;=\; \frac{k_\mathrm{B}}{m(\gamma-1)} \left[\frac{p_0(1+p_1/p_0)}{\rho_0^\gamma(1+\rho_1/\rho_0)^\gamma}\right] \\ &=\; \frac{k_\mathrm{B}}{m(\gamma-1)} \left[\ln\left(\frac{p_0}{\rho_0^\gamma}\right) + \frac{p_1}{p_0} - \gamma\frac{\rho_1}{\rho_0}\right] \quad , \end{aligned} \tag{4.50}$$

the first term is the equilibrium entropy $s_0 \;=\; k_\mathrm{B}\ln(p_0/\rho_0^\gamma)/m(\gamma-1)$ and the remainder is the perturbation

$$s_1 \;=\; \frac{k_\mathrm{B}}{m(\gamma-1)} \left[\frac{p_1}{p_0} - \gamma\frac{\rho_1}{\rho_0}\right] \;=\; -\frac{k_\mathrm{B}}{m(\gamma-1)} \underline{\hat{\mathrm{W}}}^{(1)} \cdot \underline{\mathrm{U}} \quad . \tag{4.51}$$

Thus the perturbation will have some component of $\underline{\hat{\mathrm{U}}}^{(1)}$, if and only if, the initial perturbation changes the entropy. It is for this reason that it is dubbed the *entropy mode*.

The governing fluid equations, lacking viscosity or thermal conduction, are adiabatic equations. The continuity and energy equations can be combined to yield an equation for the evolution of specific entropy

$$\frac{Ds}{Dt} \;=\; 0 \quad \Longrightarrow \quad \partial_t s \;=\; -\,\mathbf{u}\cdot\nabla s \quad . \tag{4.52}$$

Linearizing the rightmost expression gives

$$\partial_t s_1 \;=\; -\,\mathbf{u}_1\cdot\nabla s_0 - \mathbf{u}_0\cdot\nabla s_1 \quad . \tag{4.53}$$

For the case of a static ($\mathbf{u}_0 = 0$) homogeneous equilibrium ($\nabla s_0 = 0$) entropy evolution becomes $\partial_t s_1 = 0$. This clearly calls for a solution at zero-frequency with $s_1 \neq 0$ or an adiabatic change ($s_1 = 0$) at arbitrary frequency. The first option is the entropy mode — a zero frequency mode. It represents a change to the fluid's entropy with no change to its velocity or pressure.

The right eigenvectors of the acoustic modes, $\underline{\hat{\mathrm{U}}}^{(4)}$ and $\underline{\hat{\mathrm{U}}}^{(5)}$, are orthogonal the the left eigenvector of the entropy mode, $\underline{\hat{\mathrm{W}}}^{(1)}$. This means that the acoustic modes are *adiabatic* — they do not change the entropy. In fact the wave speed c_s is sometimes called the *adiabatic sound speed* and defined as

$$c_s^2 \;=\; \frac{\partial p}{\partial\rho}\Big)_s \quad , \tag{4.54}$$

where the derivative is at constant entropy. Using the adiabatic gas law $p = K\rho^\gamma$ in this derivative yields our sound speed.

4D.3 Dissipation: viscosity & conductivity

We omitted viscosity and conductivity in the analysis above. Here we briefly consider the effect they would have on linear waves. Viscosity of either the full tensorial form, (1.35), or the simplified form, (1.34), adds the same term to the first order momentum equation

$$\partial_t \mathbf{u}_1 \;=\; -\frac{1}{\rho_0}\nabla p_1 \;+\; \frac{\mu}{\rho_0}\nabla^2\mathbf{u}_1 \;=\; -\frac{1}{\rho_0}\nabla p_1 \;+\; \nu\,\nabla^2\mathbf{u}_1 \tag{4.55}$$

where $\nu = \mu/\rho_0$ is the kinematic viscosity of the equilibrium. Viscosity also contributes viscous heating to the energy equation, eq. (1.47), but since this is quadratic in $\mathbf{u}_1$ it does not enter the linearized equation.

Introducing the plane wave solution leads to a modified momentum equation

$$-i\omega\hat{\mathbf{u}}_1 \;=\; -\,i\rho_0^{-1}\mathbf{k}\,\hat{p}_1 \;-\; \nu k^2\hat{\mathbf{u}}_1 \quad . \tag{4.56}$$

The components of $\hat{\mathbf{u}}_1$ perpendicular to $\mathbf{k}$ enter only this equation. These are the shear modes which were previously zero frequency modes. Viscosity modifies their dispersion relation to

$$\omega_{2,3}(\mathbf{k}) \;=\; -\,i\,\nu\,k^2 \quad . \tag{4.57}$$

These negative imaginary frequencies lead to pure exponential damping: $e^{-i\omega t} = e^{-\nu k^2 t}$, characteristic of diffusion. Viscosity therefore does not provide a restoring force, but does lead to the decay of velocity shear. The kinematic viscosity of air $\nu \simeq 0.15\,\mathrm{cm}^2/\mathrm{sec}$ shows that shear across scales of a meter will decay on hour-long time scales. Most fluid dynamics on that scale occurs on much faster times-scales (think of the air around you as you walk) so the effect of viscosity is negligible. This accords with our earlier scaling analysis, and justifies our neglect in the foregoing.

The remaining component of the velocity couples to both pressure and density. The 3×3 system, i.e. modification of eq. (4.33), becomes

$$\omega\begin{bmatrix}\hat{\rho}\\ \hat{u}\\ \hat{p}\end{bmatrix} \;=\; \begin{bmatrix}0 & k\rho_0 & 0\\ 0 & -i\,\nu k^2 & k\rho_0^{-1}\\ 0 & \gamma k p_0 & 0\end{bmatrix}\cdot\begin{bmatrix}\hat{\rho}\\ \hat{u}\\ \hat{p}\end{bmatrix} \quad . \tag{4.58}$$

The characteristic equation for this system is

$$\omega^2(\omega + i\nu k^2) \;-\; \omega\,c_s^2 k^2 \;=\; \omega\,[\,\omega^2 + i\nu k^2\omega - c_s^2 k^2\,] \;=\; 0 \quad . \tag{4.59}$$

One zero frequency mode, the entropy mode, remains as such. Viscosity does nothing to affect the persistence of entropy perturbations.

The acoustic modes now have frequencies

$$\omega_{4,5}(\mathbf{k}) \;=\; \mp\,k\,\sqrt{c_s^2 - \nu^2k^2/4} \;-\; \tfrac{1}{2}i\nu k^2 \quad . \tag{4.60}$$

Viscosity evidently defines the characteristic length scale, $\ell_\nu = \nu/c_s$, which in air is of order 40 nm — truly small.[5] For wavelengths shorter than this, in particular $k\ell_\nu > 2$ (i.e.

[5] We see in later sections that ℓ_ν is roughly the mean free path of a particle. We also see that diffusive transport terms, like viscosity, are of questionable validity at scales this small or smaller. We will continue to discuss this questionable regime here, if only for mathematical completeness.

$\nu k > 2c_s$), acoustic frequencies are are purely imaginary. Indeed they lie on the negative imaginary axis so the behavior is decay on times shorter than $\ell_\nu/c_s \sim 10^{-10}$ sec — very fast. For longer wavelengths both modes have the same negative imaginary part, and therefore decay at half the rate of shear modes: $e^{-\nu k^2 t/2}$. Viscosity therefore damps acoustic waves.

There is a clear analogy here to damped harmonic motion, and the corresponding quality factor is

$$Q \;=\; \frac{\mathrm{Re}(\omega)}{2\,\mathrm{Im}(\omega)} \;\simeq\; \frac{1}{2\nu k/c_s} \;=\; \frac{1}{2k\ell_\nu} \quad ,$$

characterizing the number of oscillations during one decay time. Waves within the acoustic range, $f \sim 10^2$–10^4 Hz, will have quality factors $Q \sim 10^6$–10^8 — they damp very little due to viscosity.

Thermal conductivity contributes only to the energy equation. Its contribution to first order is

$$\begin{aligned}
\partial_t p_1 &= -\gamma p_0\,\nabla\cdot\mathbf{u}_1 \;+\; (\gamma-1)\kappa\,\nabla^2 T_1 \\
&= -\gamma p_0\,\nabla\cdot\mathbf{u}_1 \;+\; \frac{(\gamma-1)m\kappa}{k_{\rm B}\rho_0}\left(\nabla^2 p_1 \;-\; \frac{p_0}{\rho_0}\nabla^2\rho_1\right) \\
&= -\gamma p_0\,\nabla\cdot\mathbf{u}_1 \;+\; \tilde\kappa\left(\nabla^2 p_1 \;-\; \frac{c_s^2}{\gamma}\nabla^2\rho_1\right) \quad ,
\end{aligned}$$

where $\tilde\kappa$ is the thermal diffusion coefficient. Since it does not enter the momentum equation thermal conductivity does not affect the shear modes: they still have zero frequency. The 3×3 system for longitudinal velocity disturbances, i.e. (4.33), becomes

$$\omega\begin{bmatrix}\hat\rho\\ \hat u\\ \hat p\end{bmatrix} \;=\; \begin{bmatrix} 0 & k\rho_0 & 0 \\ 0 & 0 & k\rho_0^{-1} \\ i(c_s^2/\gamma)\,\tilde\kappa k^2 & \gamma k p_0 & -i\,\tilde\kappa k^2\end{bmatrix}\cdot\begin{bmatrix}\hat\rho\\ \hat u\\ \hat p\end{bmatrix} \quad , \tag{4.61}$$

The characteristic equation for this

$$\omega^2(\omega + i\,\tilde\kappa k^2) - \omega\,c_s^2 k^2 - i(c_s^2/\gamma)\,\tilde\kappa k^4 = \omega^3 + i\tilde\kappa k^2\,\omega^2 - c_s^2 k^2\,\omega - i(c_s^2/\gamma)\,\tilde\kappa k^4 = 0 \quad , \tag{4.62}$$

is cubic and therefore not readily solved.

In spite of its non-trivial order (cubic) we can establish that eq. (4.62) has at least one purely imaginary root, $\omega_3 = -i\,\sigma$, where the real value σ satisfies the cubic

$$\sigma^3 - (\tilde\kappa k^2)\sigma^2 + (c_s^2 k^2)\sigma - (c_s^2/\gamma)\,\tilde\kappa k^4 \;=\; 0 \quad . \tag{4.63}$$

This has no real solution for $\sigma < 0$, since every term would be negative. Since $\sigma > 0$, the frequency lies on the negative imaginary axis and therefore corresponds to damping $e^{-\sigma t}$. Substituting $\sigma = 0$ and $\sigma = \tilde\kappa k^2$ into the left of eq. (4.63) yields values of opposite sign. This proves that at least one solution must lie within the range

$$0 \;<\; \sigma \;\le\; \tilde\kappa k^2 \quad ; \tag{4.64}$$

the equality holds only under the condition $\gamma = 1$. For wave lengths longer and shorter than the characteristic scale, $\ell_\kappa = \tilde\kappa/c_s$, eq. (4.63) has the solution

$$\sigma \;\simeq\; \frac{1}{\gamma}\,\tilde\kappa k^2 \qquad , \qquad k\ell_\kappa \;\ll\; 1 \tag{4.65}$$

$$\sigma \;\simeq\; \tilde\kappa k^2 \;-\; \frac{\gamma-1}{\gamma}\,\frac{c_s^2}{\tilde\kappa} \qquad , \qquad k\ell_\kappa \;\gg\; 1 \quad , \tag{4.66}$$

clearly falling within the range (4.64). Both limits represent diffusive damping, $\sigma \sim \tilde{\kappa} k^2$.

This damped mode has structure

$$\hat{u}_1 \;=\; -\frac{i\sigma}{k}\frac{\hat{\rho}_1}{\rho_0} \quad , \quad \hat{p}_1 \;=\; -\,p_0\,\frac{\gamma\,\sigma^2}{k^2 c_s^2}\,\frac{\hat{\rho}_1}{\rho_0} \quad . \tag{4.67}$$

For long wavelengths $\sigma \sim \tilde{\kappa} k^2 \ll c_s k$ so the perturbation is primarily to the density, $\hat{p}_1/p_0 \ll \hat{\rho}_1/\rho_0$. This is evidently the entropy mode and it is damped by thermal conductivity. Perturbations to the temperature, and thus the density, decay by thermal diffusion. This leads to opposing pressure perturbations, albeit much smaller, which in turn drive flows.

The other two solutions to (4.62) can be found in terms of this first one[6]

$$\omega_{4,5} \;=\; \pm\sqrt{c_s^2 k^2 - (\tilde{\kappa}k^2 - \sigma)(\tilde{\kappa}k^2 + 3\sigma)/4} \;-\; \frac{i}{2}\,(\tilde{\kappa}k^2 - \sigma) \quad . \tag{4.68}$$

These are evidently the sound waves. Using eq. (4.64) shows that the imaginary part of both frequencies are negative: they are damped. It seems that sound is damped by conductivity as well as by viscosity. Here the damping occurs because the wave creates temperature perturbation through its compression and rarefaction. Thermal conduction diffuses these during the course of an oscillation, thereby removing a bit of the restoring force (the phase speed is slightly decreased) and causing irreversible loss: damping.

In the long wavelength limit, $k\ell_\kappa \ll 1$, the damping rate of sound waves scales with the diffusion rate $\sim \tilde{\kappa} k^2$. If both viscosity and thermal conductivity are present then both contribute to a net damping rate, $\sim (\nu + \tilde{\kappa})k^2$. The dimensionless ratio $\mathrm{Pr} = \nu/\tilde{\kappa}$, called the Prandtl number, determines which, if either, is dominant. Air has $\mathrm{Pr} \sim 0.7$, so both effects contribute comparably. A high-temperature plasma has $\mathrm{Pr} \sim 10^{-2}$, so conductivity is far more effective than viscosity at damping acoustic waves.

One final point of interest comes from the short wavelength limit, i.e. eq. (4.66), of eq. (4.68)

$$\omega_{4,5} \;\simeq\; \pm\,\frac{c_s}{\sqrt{\gamma}}\,k \;-\; i\,\frac{\gamma - 1}{2\gamma}\,\frac{c_s^2}{\tilde{\kappa}} \quad . \tag{4.69}$$

The real part can be identified with propagation at the *isothermal* sound speed, $\sqrt{p_0/\rho_0}$. Evidently the action of thermal diffusion removes temperature perturbation faster than the sound wave can create them. Instead of diffusive damping, these short-wavelength waves damp at a fixed rate, $\sim c_s^2/\tilde{\kappa} = c_s/\ell_\kappa$, related to the collision frequency between particles. Since the viscous damping continues to scale as $\sim \nu k^2$, this becomes the dominant effect at small scales.

4E Example

To illustrate the excitation of normal modes we consider an initial perturbation to temperature which does not affect density or velocity. This can be interpreted as a detonation enhancing the temperature without affecting (initially) density or velocity. (It will have to be a minor detonation to make only a relatively small change to the equilibrium.) To

[6]Subtract the defining cubic of the known root, x_1, from that of the unknown, x_2, then factor $x_2 - x_1$ from the resulting polynomial. The result will be quadratic in x_2.

simplify the math we consider a one-dimensional enhancement, depending only on x, characterized by a function $f(x)$. The initial condition will be

$$p_1(x,y,z,0) \;=\; \epsilon p_0 f(x) \;=\; \epsilon p_0 \int_{-\infty}^{\infty} \hat{f}(k_x)\, e^{ik_x x} dk_x \quad , \tag{4.70}$$

where $\epsilon \ll 1$ assures a small perturbation and $\hat{f}(k_x)$ is the Fourier transform of $f(x)$. The density and velocity are unperturbed so $\rho_1(x,y,z,0) = u_{1,x}(x,y,z,0) = 0$.

Due to the shape of the perturbation only waves with wave vectors $\mathbf{k} = k_x\hat{\mathbf{x}}$ are excited. The directions perpendicular to this, $\hat{\mathbf{m}} = \hat{\mathbf{y}}$ and $\hat{\mathbf{n}} = \hat{\mathbf{z}}$, are the polarizations of the shear modes. Since these are not excited by the initial perturbation they are not present. The full solution thus consists of the superposition

$$\underline{\mathrm{U}}_1(x,t) \;=\; \int_{-\infty}^{\infty} e^{ik_x x} \left[A(k_x)\hat{\underline{\mathrm{U}}}^{(1)} e^{-\omega_1 t} + B(k_x)\hat{\underline{\mathrm{U}}}^{(4)} e^{-i\omega_4 t} + C(k_x)\hat{\underline{\mathrm{U}}}^{(5)} e^{-i\omega_5 t} \right] dk_x \tag{4.71}$$

where A, B and C are amplitudes of the entropy mode, forward and reverse acoustic modes respectively. The amplitudes are set by the initial conditions of all fields. Omitting rows 3 and 4, corresponding to the shear modes, this is

$$\begin{aligned} \underline{\mathrm{U}}_1(x,0) \;=\; \begin{bmatrix} \rho_1(x,0) \\ u_{1,x}(x,0) \\ p_1(x,0) \end{bmatrix} \;&=\; \int_{-\infty}^{\infty} e^{ik_x x} \begin{bmatrix} 0 \\ 0 \\ \epsilon p_0 \end{bmatrix} \hat{f}(k_x)\, dk_x \\ &=\; \int_{-\infty}^{\infty} e^{ik_x x} \begin{bmatrix} \rho_0 & \rho_0 & \rho_0 \\ 0 & s\,c_s & -s\,c_s \\ 0 & \gamma p_0 & \gamma p_0 \end{bmatrix} \cdot \begin{bmatrix} A(k_x) \\ B(k_x) \\ C(k_x) \end{bmatrix} dk_x \end{aligned}$$

where $s = \hat{k}_x = \mathrm{sgn}(k_x)$. The columns of the matrix combines rows 1, 2 and 5 of column vectors $\hat{\underline{\mathrm{U}}}^{(1)}$, $\hat{\underline{\mathrm{U}}}^{(4)}$ and $\hat{\underline{\mathrm{U}}}^{(5)}$. The amplitudes must therefore satisfy the equations

$$\begin{bmatrix} 0 \\ 0 \\ \epsilon p_0 \end{bmatrix} \hat{f}(k_x) \;=\; \begin{bmatrix} \rho_0 & \rho_0 & \rho_0 \\ 0 & s\,c_s & -s\,c_s \\ 0 & \gamma p_0 & \gamma p_0 \end{bmatrix} \cdot \begin{bmatrix} A(k_x) \\ B(k_x) \\ C(k_x) \end{bmatrix} \tag{4.72}$$

for each k_x. The solution is

$$\begin{bmatrix} A(k_x) \\ B(k_x) \\ C(k_x) \end{bmatrix} \;=\; \frac{\epsilon}{2\gamma} \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix} \hat{f}(k_x) \quad . \tag{4.73}$$

Returning the amplitudes to eq. (4.71) the evolving perturbation we can perform the inverse Fourier transform. For the pressure this gives us

$$p_1(x,t) \;=\; \int_{-\infty}^{\infty} e^{ik_x x} \left[A(k_x)\hat{U}_p^{(1)} e^{-i\omega_1 t} + B(k_x)\hat{U}_p^{(4)}\, e^{-i\omega_4 t} + C(k_x)\hat{U}_p^{(5)}\, e^{-i\omega_5 t} \right] dk_x$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} e^{ik_x x}\left[\, A(k_x)\times 0\, e^{-i\omega_1 t} + B(k_x)\gamma p_0\, e^{-i\omega_4 t} + C(k_x)\gamma p_0\, e^{-i\omega_5 t}\,\right] dk_x \\
&= \int_{-\infty}^{\infty} e^{ik_x x}\left[\frac{\epsilon}{2} p_0 \hat{f}(k_x)\, e^{-ic_s|k_x|t} + \frac{\epsilon}{2} p_0 \hat{f}(k_x)\, e^{+ic_s|k_x|t}\right] dk_x \\
&= \frac{\epsilon}{2} p_0 \int_{-\infty}^{\infty} e^{ik_x x}[2\cos(c_s|k_x|t)]\hat{f}(k_x)\, dk_x \;=\; \frac{\epsilon}{2} p_0 \int_{-\infty}^{\infty} e^{ik_x x}[2\cos(c_s k_x t)]\hat{f}(k_x)\, dk_x \\
&= \frac{\epsilon}{2} p_0 \int_{-\infty}^{\infty} \left[e^{ik_x(x-c_s t)} + e^{ik_x(x+c_s t)}\right] \hat{f}(k_x)\, dk_x \;=\; \frac{\epsilon}{2} p_0\,[\, f(x-c_s t) \;+\; f(x+c_s t)\,] \quad .
\end{aligned}
$$

It is easy to verify that this expression matches eq. (4.70) at $t = 0$. At later times the pressure perturbation breaks into two equal parts and propagates at c_s in each direction.

Performing similar inverse transforms on the other rows of the vector gives the full set of perturbation

$$\rho_1(x,t) = \frac{\epsilon}{2\gamma}\rho_0\,[\, f(x-c_s t) \;+\; f(x+c_s t) \;-\; 2f(x)\,] \tag{4.74}$$

$$u_{1,x}(x,t) = \frac{\epsilon}{2\gamma} c_s\,[\, f(x-c_s t) \;-\; f(x+c_s t)\,] \tag{4.75}$$

$$p_1(x,t) = \frac{\epsilon}{2} p_0\,[\, f(x-c_s t) \;+\; f(x+c_s t)\,] \tag{4.76}$$

This shows that all normal modes are dispersionless so the perturbation propagates without changing shape. The initial pressure pulse propagates in equal parts in each direction appearing at $x > 0$ at time $t = x/c_s$ raising the pressure half the initial amount.

4F Waves in inhomogeneous equilibria: the Eikenol approximation

4F.1 Why plane waves work

Plane waves will be the normal modes for any system of equations provided they are linearized about a homogeneous equilibrium — one independent of any spatial coordinate. The reasoning behind this general conclusion goes as follows. When acting on the function $e^{i\mathbf{k}\cdot\mathbf{x}}$ the spatial derivative ∂_j will bring down a factor ik_j, but otherwise leave the function unchanged. The linear operator $\hat{\mathsf{F}}(\nabla)$ incorporates various spatial derivatives, i.e. components of the ∇ operator, with coefficients which are constant (independent of $\mathbf{x}$) provided the equilibrium is homogeneous. When acting on the plane wave the derivatives become $\nabla \to i\mathbf{k}$ so the matrix of operators becomes a simple matrix $\underline{\underline{F}} = \hat{\mathsf{F}}(i\mathbf{k})$. The eigenvalues of this matrix are $-i\omega$, where ω is the eignefrequency of the mode. The plane wave is thus an eigenfunction of $\hat{\mathsf{F}}(\nabla)$ simply because it is an eigenfunction of the derivative operator ∇.

4F.2 When plane waves do not work

The above reasoning fails to apply when the equilibrium depends on any spatial coordinate. It fails, for example, when the hydrostatic atmosphere is stratified by gravity. The linearized system can still be solved using normal modes, $\underline{\tilde{\mathrm{U}}}(\mathbf{x})e^{-i\omega t}$, but the modes must satisfy a PDE

$$\hat{\mathsf{F}}(\nabla;\mathbf{x})\cdot\underline{\tilde{\mathrm{U}}} \;=\; -i\omega\underline{\tilde{\mathrm{U}}} \quad , \tag{4.77}$$

more complicated than that of of the homegeneous equilibrium.

Equations of the Sturm-Liouville type are examples of this more general kind of eigenfunction equation. Application to a gravitationally stratified atmosphere leads to an equation of this form. Normal mode solutions exist with various frequencies, depending on the boundary conditions. These modes have oscillatory structure (think of the Bessel functions), and those with shorter length-scale have higher eigenfrequency. They are not, generally speaking, plane waves but as the length-scale of oscillation becomes shorter, and the frequency higher, they become increasingly similar to plane waves. This high-frequency-short-wavelength limit is the object of the approximate solution known as the *Eikenol method* or WKB method. It is useful in solving any inhomogeneous linear system provided it depends on spatial coordinates smoothly.

4F.3 What to do without plane waves — the Eikenol method

The basic idea of the Eikenol approximation is to propose a normal mode solution of the form

$$\underline{\mathrm{U}}_1(\mathbf{x},t) \;=\; \underline{\tilde{\mathrm{U}}}(\mathbf{x})e^{-i\omega t} \;=\; \underline{\tilde{\mathrm{A}}}(\mathbf{x})\,\exp[i\phi(\mathbf{x})-i\omega t] \quad , \tag{4.78}$$

where $\underline{\tilde{\mathrm{A}}}(\mathbf{x})$ and $\phi(\mathbf{x})$, the amplitude and phase, are what must now be found.[7] It hardly qualifies as progress to replace one vector of unknowns, $\underline{\tilde{\mathrm{U}}}(\mathbf{x})$, with another $\underline{\tilde{\mathrm{A}}}(\mathbf{x})$ plus an additional unknown function $\phi(\mathbf{x})$. What makes this worthwhile is that we assume that the amplitude, $\underline{\tilde{\mathrm{A}}}(\mathbf{x})$, varies relatively smoothly in space — on length scale comparable to variations in the equilibrium. The complex phase of the solution, $\phi-\omega t$, is assumed to change rapidly in space and time so that the solution oscillates on short time-scales and length-scales. The upshot of this reasoning is that spatial derivatives of the two unknown fields are not comparable

$$|\nabla\phi| \;\gg\; |\nabla\ln\tilde{A}_j| \quad . \tag{4.79}$$

More than that, we assume that $\nabla\phi$ varies on scales far longer than ϕ itself varies:

$$|\partial_i\partial_j\phi| \;\ll\; |\nabla\phi|^2 \quad . \tag{4.80}$$

This is a sophisticated way of saying we assume the normal mode looks *locally* like a plane wave, but with gradually varying properties like amplitude and wave vector (see Weinberg, 1962; Born & Wolf, 1980, for more thorough descriptions).

The approximations above mean that every spatial derivative ∂_j will act only on the exponential factor in expression (4.78) bringing down a factor $i\partial_j\phi$ and leaving the expression

[7] Once again we plan to find several such solutions and superpose them into a real function.

otherwise unchanged. Subsequent derivatives will bring down similar factors. Differentiation of the amplitude or factors from previous differentiation are negligible owing to inequalities (4.79) and (4.80). This means the normal mode equation becomes

$$\hat{\mathsf{F}}(\nabla;\mathbf{x})\cdot\underline{\tilde{\mathrm{U}}} \;\simeq\; \hat{\mathsf{F}}(i\nabla\phi,\mathbf{x})\cdot\underline{\tilde{\mathrm{A}}}\,e^{i\phi-\omega t} \;=\; -i\omega\,\underline{\tilde{\mathrm{A}}}\,e^{i\phi-\omega t} \quad . \tag{4.81}$$

Just as with the plane wave, the operator $\hat{\mathsf{F}}$ is now a matrix multiplying a column vector. The only difference is that the wave vector, $\mathbf{k}$, has been replaced by the vector $\nabla\phi$ which will vary in space. To make the analogy clearer still it is customary to define a local wave vector

$$\mathbf{k}(\mathbf{x}) \;=\; \nabla\phi \quad . \tag{4.82}$$

In a genuine plane wave $\phi = \mathbf{k}\cdot\mathbf{x}$ and $\mathbf{k}$ is a constant.

Thanks to the Eikenol approximation the complicated operator $\hat{\mathsf{F}}(\nabla)$ has now been replaced with a simple $N\times N$ matrix, $\underline{\underline{F}}$; in hydrodynamics $N=5$. This means the eigenfunction equations has become, in the Eikeonal approximation, a simple eigenvector equation

$$\underline{\underline{F}}(\nabla\phi,\mathbf{x})\cdot\underline{\tilde{\mathrm{A}}} \;=\; -i\omega\underline{\tilde{\mathrm{A}}} \quad . \tag{4.83}$$

This is exactly the same as equation eq. (4.30) for plane waves, except that now the matrix depends on position, through the equilibrium, and on $\nabla\phi$ instead of $\mathbf{k}$. There is a non-trivial solution when ω satisfies the determinantal equation

$$\det\Big[\,\underline{\underline{F}}(\nabla\phi,\mathbf{x})+i\omega\mathsf{I}\,\Big] \;=\; \prod_{n=1}^{N}[\,\omega-\omega_n(\nabla\phi,\mathbf{x})\,] \;=\; 0 \quad . \tag{4.84}$$

The determinant is an N^{th}-order polynomial in ω whose coefficients depend on $\nabla\phi$ and $\mathbf{x}$ through the matrix $\underline{\underline{F}}$. We designate by ω_n the n^{th} of N roots of this polynomial (including possible repetition in the case of degeneracy). Their dependence on $\mathbf{x}$ and $\nabla\phi$, through the coefficients, can be quite complicated. Nevertheless we can formally express the solution to eq. (4.83) as

$$\omega \;=\; \omega_n(\nabla\phi,\mathbf{x}) \quad , \quad \underline{\tilde{\mathrm{A}}}(\mathbf{x}) \;=\; \underline{\hat{\mathrm{U}}}^{(n)}(\nabla\phi,\mathbf{x}) \tag{4.85}$$

where n indexes the different eigenvectors of the matrix. These will be identical to the eigenvalues and eigenvectors for the homogeneous equilibrium whose properties match those found at position $\mathbf{x}$.

Equation (4.85) tells us that the eikonal solution is intimately related to the plane wave solution. Each plane wave mode, n, has a dispersion relation $\omega_n(\mathbf{k})$. The present situation differs from that of the homogeneous problem, used to derive $\omega_n(\mathbf{k})$, in a few subtle, but important respects. Firstly, in the present case we take ω to be fixed and invert the dispersion relation to find $\nabla\phi$. The inversion can be very complicated since ω_n is the root of a high-order polynomial. Even if we know this complicated form, however, the relation provides only a single constraint on the three components of the vector $\nabla\phi$. Consider, for example, a hydrostatic equilibrium, $\nabla p_0 = \rho_0\mathbf{g}$ characterized by sound speed $c_s(\mathbf{x})$. The eikenol solution for the forward acoustic mode ($n=4$) satisfies the relation

$$\omega \;=\; \omega_4(\nabla\phi,\mathbf{x}) \;=\; c_s(\mathbf{x})|\nabla\phi| \quad . \tag{4.86}$$

Inverting this relation gives the magnitude of the gradient but not its direction: $|\nabla\phi| = \omega/c_s(\mathbf{x})$. This is insufficient to find the entire function $\phi(\mathbf{x})$.

Before we address this issue we must admit to another subtle difference between the eikenol approach and the homogeneous problem used to derive the dispersion relation $\omega_n(\mathbf{k})$. Assuming a homogeneous equilibrium of infinite extent led us to plane waves whose wave vectors $\mathbf{k}$ had purely real components. Any imaginary components would have led to unbounded growth of the eigenfunction in some direction, and we rejected that. The dispersion relation, $\omega_n(\mathbf{k})$, on the other hand, is one eigenvalue of a non-symmetric $N \times N$ matrix with possibly complex coefficients. It is therefore possible for $\omega_n(\mathbf{k})$ to be complex, even though $\mathbf{k}$ is purely real. In the eikenol problem we set ω to some real value and attempt to invert the dispersion relationship to find $\mathbf{k}$. If $\omega_n(\mathbf{k})$ takes on complex values for real $\mathbf{k}$, then it will take on real values for *complex* wave vectors $\mathbf{k}$ — that is wave vectors with complex components.[8] Thus the eikenol problem will lead to complex wave vectors and real frequencies if and when the homogeneous problem has complex frequencies for real wave vectors.

We will note this possibility but assume, for the time being, that $\omega_n(\mathbf{k})$ is real when $\mathbf{k}$ is real. Thus we have only undamped waves. This allows us to return to the problem of finding all components of $\mathbf{k}$ from a single equation, namely eq. (4.85). To make progress we must take advantage of additional information about $\mathbf{k}(\mathbf{x})$: it is the gradient of a scalar. This means that $\nabla \times \mathbf{k} = 0$ at all points. A different version of the same information is contained in the expression

$$\frac{\partial k_\ell}{\partial x_j} \;=\; \frac{\partial k_j}{\partial x_\ell} \;=\; \frac{\partial^2 \phi}{\partial x_j \partial x_\ell} \quad . \tag{4.87}$$

To take advantage of this we must look at the spatial variation of the constraint provided by eq. (4.85). Differentiating the relation with respect to x_j, and accounting for the spatial dependence of $\mathbf{k}(\mathbf{x}) = \nabla\phi$, gives

$$0 \;=\; \sum_\ell \frac{\partial \omega_n}{\partial k_\ell}\frac{\partial k_\ell}{\partial x_j} \;+\; \frac{\partial \omega_n}{\partial x_j} \quad , \tag{4.88}$$

since the oscillation frequency ω is a constant. Using relation (4.87) allows us to re-write this in the suggestive form

$$\left(\frac{\partial \omega_n}{\partial \mathbf{k}} \cdot \nabla\right)\mathbf{k} \;=\; -\nabla\omega_n \quad . \tag{4.89}$$

Equation (4.89) is a first-order differential equation for the wave vector $\mathbf{k} = \nabla\phi$. The function $\omega_n(\mathbf{k}, \mathbf{x})$ can be a complicated function of its argument, but it is a function we know; thus we also know derivatives like $\nabla\omega_n$ and $\partial\omega_n/\partial\mathbf{k}$, as functions of $\mathbf{x}$ and $\mathbf{k}$. Beginning with $\mathbf{k}(\mathbf{x}_0)$ at one particular point, $\mathbf{x}_0$, eq. (4.89) shows us how to find $\mathbf{k}$ at a neighboring point. That point will, however, lie in the direction of $\pm\partial\omega_n/\partial\mathbf{k}$, since (4.89) involves a single directional derivative. From that new point we can continue filling in the wave vector everywhere along some curve which is everywhere tangent to $\partial\omega_n/\partial\mathbf{k}$. This curve is called the *characteristic* of the PDE.

[8] To fully appreciate this complimentarily one should note that determinant in (4.84) will be a polynomial, and therefore analytic, in both $\mathbf{k}$ and ω. Its root, $\omega_n(\mathbf{k})$, will therefore be analytic in each component of $\mathbf{k}$. This function can be evaluated for complex values of those components even though we assumed real values in its derivation.

Equation (4.89) is, in fact, a first-order equation of the hyperbolic variety and therefore has one characteristic. This means it can be re-cast as an *ordinary* differential equation along its characteristic curve; this is a formal re-statement of the reasoning presented above. We parameterize the curve with a time-like coordinate τ and refer to it as a *ray path*. The tangent of the ray path, $d\mathbf{x}/d\tau$, must be parallel to the direction $\partial\omega_n/\partial\mathbf{k}$ along which eq. (4.89) contains information. We can redefine the parameter τ in order to make those two equal so that

$$\frac{d}{d\tau} \;=\; \frac{d\mathbf{x}}{d\tau}\cdot\nabla \;=\; \frac{\partial\omega_n}{\partial\mathbf{k}}\cdot\nabla \quad .$$

This is the directional derivative on the left hand side of eq. (4.89). Thus the equation and the definition of the ray can be combined into a pair of ODEs

$$\frac{d\mathbf{x}}{d\tau} \;=\; \frac{\partial\omega_n}{\partial\mathbf{k}} \tag{4.90}$$

$$\frac{d\mathbf{k}}{d\tau} \;=\; -\frac{\partial\omega_n}{\partial\mathbf{x}} \quad . \tag{4.91}$$

known as Hamilton's equations.[9]

Variation of $\mathbf{x}$ and $\mathbf{k}$ along the ray path lead to possible frequency variation along the path. Using the chain rule we find

$$\frac{d\omega_n}{d\tau} \;=\; \sum_\ell \frac{\partial\omega_n}{\partial x_\ell}\frac{dx_\ell}{d\tau} \;+\; \sum_\ell \frac{\partial\omega_n}{\partial k_\ell}\frac{dk_\ell}{d\tau} \;=\; 0 \quad , \tag{4.92}$$

once eqs. (4.90) and (4.91) are used. This result is re-assuring since we began by assuming that $\omega_n[\mathbf{k}(\mathbf{x}),\mathbf{x}] = \omega$, a constant, at all points in space.

Hamilton's equations will, in principle, permit us to solve for the complete phase function $\phi(\mathbf{x})$. We must begin with its value and its normal derivative over some surface. The known function can be differentiated within the surface to complete the gradient $\mathbf{k} = \nabla\phi$. Thus value and normal derivative of ϕ specifies $\mathbf{k}$ at every point on the surface. From this we use Hamilton's equations to find $\mathbf{k}(\mathbf{x})$ along each ray path piercing the surface. This will extend the gradients to a volume intersected by all the rays. We then integrate the gradients along paths to find the values of $\phi(\mathbf{x})$ throughout that same volume. Our use of eq. (4.87) guarantees that the value found will not depend on the path. All that remains is to find the amplitude vector $\underline{\tilde{\mathrm{A}}}(\mathbf{x})$, but we will not do that here.

This makes clear the necessary and sufficient conditions for solution: data must be specified on a surface pierced no more than once by any ray. The solution will be found throughout the volume filled by these rays, and left unknown at any point not intersected by a ray from the data surface. If all N of the wave modes had dispersion relations $\omega_n(\mathbf{k})$, which were real for real vectors $\mathbf{k}$, then our original system has N characteristics along which information propagates. The condition of real eigenfrequencies thus leads us to the conclusion that the original system is an N^{th} order *hyperbolic* equation. One or more complex eigenfreqencies, for real $\mathbf{k}$, would then indicate that the governing equation was not entirely hyperbolic. This information is found in the eikenol limit of the equations.

[9] If this reminds you of the canonical equations of motion from a Hamiltonian, this is no accident — it is the same Dr. Hamiliton.

The true value of the above approach lies in the insight it provides into the propagation of information through the linearized system. It shows that the in the limit of very short wave lengths, solutions to linear systems locally resemble plane wave solutions. The spatial variation is described by a wave vector, $\mathbf{k}$, which will vary gradually in space with equilibrium properties. The possible waves at any point are the same as those possible in a homogeneous equilibrium with the same properties.

There are characteristics, or ray paths, in the linear system describing how information moves through it. According to eq. (4.90), the ray is parallel to the *group velocity*

$$\mathbf{v}_g(\mathbf{k},\mathbf{x}) \;=\; \frac{\partial \omega_n}{\partial \mathbf{k}} \quad . \tag{4.93}$$

The group velocity therefore characterizes the propagation of information by linear waves, at least waves of short wavelengths. In many cases this direction differs from that of the wave vector, $\mathbf{k}$.

4F.4 Examples of the eikenol method

Spatially varying sound speed

We illustrate the eikonal method using a fluid with sound speed $c_s(\mathbf{x})$ varying in space. We must assume that variation is smooth, and think of waves whose wave-length is much smaller than the scale of variation in c_s. Forward propagating acoustic modes are described by the dispersion relation

$$\omega_4(\mathbf{k},\mathbf{x}) \;=\; c_s(\mathbf{x})|\mathbf{k}| \quad . \tag{4.94}$$

The group velocity of this mode,

$$\mathbf{v}_g(\mathbf{k},\mathbf{x}) \;=\; \frac{\partial \omega_4}{\partial \mathbf{k}} \;=\; c_s(\mathbf{x})\,\hat{\mathbf{k}} \quad , \tag{4.95}$$

turns out to be in the same direction as the wave vector; the speed of propagation is the sound speed. Dividing eq. (4.90), by $d\ell/d\tau = |\mathbf{v}_g| = c_s$ gives an equation for the ray parameterized by its length, ℓ. The component perpendicular to $\hat{\mathbf{k}}$ gives the curvature of that path

$$\frac{d\hat{\mathbf{k}}}{d\ell} \;=\; \frac{1}{|\mathbf{k}|c_s}\,\frac{d\mathbf{k}}{d\tau}\Big)_{\perp} \;=\; -\,\nabla_\perp \ln(c_s) \quad . \tag{4.96}$$

This shows that the ray path is refracted in the direction of lower sound speed.

To make things more concrete, let's assume the sound speed depends on only one spatial coordinate, x. In this case eq. (4.91) becomes

$$\frac{d\mathbf{k}}{d\tau} \;=\; -\,\frac{\partial \omega_4}{\partial \mathbf{x}} \;=\; -\,c_s'(x)\,|\mathbf{k}|\,\hat{\mathbf{x}} \quad . \tag{4.97}$$

This shows immediately that k_y and k_z are constant along the ray path. They play a role analogous to conserved momenta in a Hamiltonian with cyclic coordinates. Any ray path will be defined by its value of ω and by k_y and k_z, which are set at the initial point of the

ray, $\tau = 0$. Continuing with that analogy, we can use the conservation of $\omega_4(\mathbf{k}, \mathbf{x})$ to solve for k_x, instead of trying to solve eq. (4.97). Since $k^2 = \omega^2/c_s^2(x)$ we can write

$$k_x \;=\; \pm\sqrt{\frac{\omega^2}{c_s^2(x)} \;-\; k_y^2 \;-\; k_z^2} \;=\; \pm\sqrt{\frac{\omega^2}{c_s^2(x)} \;-\; k_\perp^2} \quad , \tag{4.98}$$

where we have introduced $k_\perp = \sqrt{k_y^2 + k_y^2}$, and the choice of sign will depend on the direction of the wave vector.

Equation (4.98) shows immediately that the ray cannot propagate into regions where $c_s(x) > \omega/k_\perp$ — recall we the eikenol solution requires real wave vectors. The allowed region depends on the properties of the wave we are posing, namely ω and $k_\perp$. There can points along the ray path, *classical turning points*, where $k_x = 0$ and the ray must reflect. As it does so k_x will change from one to the other of the signs in eq. (4.98). If the sound speed has a local minimum, then a ray with $\omega/k_\perp > c_s(x_{\rm min})$ can be trapped between two classical turning points, one on either side of $x_{\rm min}$.

We now turn to the three components of eq. (4.90) to find the ray path itself

$$\frac{dx}{d\tau} \;=\; c_s(x)\frac{k_x}{|\mathbf{k}|} \quad , \quad \frac{dy}{d\tau} \;=\; c_s(x)\frac{k_y}{|\mathbf{k}|} \quad , \quad \frac{dz}{d\tau} \;=\; c_s(x)\frac{k_z}{|\mathbf{k}|} \quad . \tag{4.99}$$

We will simplify things by considering the safe $k_z = 0$ so the ray remains in the (x, y) plane. We can simplify things further by eliminating the artificial parameter τ in favor of the coordinate y. We do this by taking the ratio

$$\frac{dx}{dy} \;=\; \frac{dx/d\tau}{dy/d\tau} \;=\; \frac{k_x}{k_y} \;=\; \pm\sqrt{\frac{\omega^2}{k_y^2 c_s^2(x)} \;-\; 1} \quad , \tag{4.100}$$

after using eq. (4.98) to eliminate k_x. This is a first order ODE for $x(y)$, which can at least be reduced to quadrature. The ray will reflect from a classical turning point by having dx/dy change sign there; $dx/dy = 0$ at the turning point itself.

In the case considered above, of a ray trapped near a local minimum in the sound speed, the ray will oscillate back and forth between its two classical turning points, x_1 and x_2, where $c_s(x_{1,2}) = \omega/k_y$. The ray travels

$$\Delta y \;=\; 2\int_{x_1}^{x_2}\left[\frac{\omega^2}{k_y^2 c_s^2(x)} \;-\; 1\right]^{-1/2} dx \quad , \tag{4.101}$$

in a single such "oscillation". This distance must not be confused with the wavelength of the mode, which is $2\pi/k_y$. In fact, to justify the eikenol approach this wave length must be very short compared to Δy.

An equilibrium with flow

For a slightly more complicated example we consider an equilibrium with uniform pressure, p_0, but non-uniform density, $\rho_0(\mathbf{x})$. We will also assume the system is not static, but has

an equilibrium flow $\mathbf{u}_0(\mathbf{x})$. To be in equilibrium we insist that the flow be incompressible, perpendicular to the density gradient, and have no centrifugal force

$$\nabla \cdot \mathbf{u}_0 \;=\; 0 \quad , \qquad \mathbf{u}_0 \cdot \nabla \rho_0 \;=\; 0 \quad , \qquad (\mathbf{u}_0 \cdot \nabla)\mathbf{u}_0 \;=\; 0 \quad . \tag{4.102}$$

Linearizing about this non-uniform equilibrium leads to more complicated equations

$$\partial_t \rho_1 + \underbrace{\mathbf{u}_0 \cdot \nabla \rho_1}_{1.1} \;=\; -\underbrace{\mathbf{u}_1 \cdot \nabla \rho_0}_{1.2} \;-\; \underbrace{\rho_0 (\nabla \cdot \mathbf{u}_1)}_{1.3} \quad , \tag{4.103}$$

$$\partial_t \mathbf{u}_1 + \underbrace{(\mathbf{u}_0 \cdot \nabla)\mathbf{u}_1}_{2.1} \;=\; -\underbrace{(\mathbf{u}_1 \cdot \nabla)\mathbf{u}_0}_{2.2} \;-\; \underbrace{\rho_0^{-1} \nabla p_1}_{2.3} \quad , \tag{4.104}$$

$$\partial_t p_1 + \underbrace{\mathbf{u}_0 \cdot \nabla p_1}_{3.1} \;=\; -\underbrace{\gamma p_0 (\nabla \cdot \mathbf{u}_1)}_{3.2} \quad . \tag{4.105}$$

Terms have been numbered for reference and terms 1.1, 2.1 and 3.1 have been moved to the left of the equal sign for convenience.

The eikenol approximation means those terms where the derivative acts on a first order quantity, namely 1.1, 1.3, 2.1, 2.3, 3.1 and 3.2, will be far larger than terms where the derivative acts on the equilibrium: 1.2 and 2.2. Those terms may thus be dropped and the gradient in the first set of terms will be replaced by $i\nabla\phi = i\mathbf{k}$. The resulting system

$$-i(\omega - \mathbf{u}_0 \cdot \mathbf{k})\hat{\rho}_1 \;=\; -\rho_0 \, i\mathbf{k} \cdot \hat{\mathbf{u}}_1 \quad , \tag{4.106}$$

$$-i(\omega - \mathbf{u}_0 \cdot \mathbf{k})\hat{\mathbf{u}}_1 \;=\; -\frac{i\mathbf{k}}{\rho_0}\, \hat{p}_1 \quad , \tag{4.107}$$

$$-i(\omega - \mathbf{u}_0 \cdot \mathbf{k})\hat{p}_1 \;=\; -\gamma p_0 \, i\mathbf{k} \cdot \hat{\mathbf{u}}_1 \quad , \tag{4.108}$$

is formally identical to eq. (4.30) except that the frequency ω, on the left, has been replaced by the *doppler-shifted* frequency $\omega - \mathbf{u}_0 \cdot \mathbf{k}$.

The system thus has the same eigenvectors, i.e. the same normal modes, but with doppler-shifted eigenfrequencies. The two shear modes and the entropy mode now have non-zero eigenfrequnecies

$$\omega_1(\mathbf{k}, \mathbf{x}) \;=\; \omega_2(\mathbf{k}, \mathbf{x}) \;=\; \omega_3(\mathbf{k}, \mathbf{x}) \;=\; \mathbf{k} \cdot \mathbf{u}_0(\mathbf{x}) \quad ; \tag{4.109}$$

they are no longer zero-frequency modes. Their group velocity

$$\mathbf{v}_g \;=\; \frac{\partial \omega_1}{\partial \mathbf{k}} \;=\; \mathbf{u}_0(\mathbf{x}) \quad , \tag{4.110}$$

is the same as the equilibrium flow velocity. This means that any perturbation to the entropy (i.e. for mode 1) will move with the equilibrium flow. The ray path of the entropy mode is therefore the same as the stream line.

The acoustic modes each have doppler-shifted frequencies

$$\omega_{4,5} \;=\; \mathbf{k} \cdot \mathbf{u}_0(\mathbf{x}) \;\pm\; c_s(\mathbf{x})\, |\mathbf{k}| \quad . \tag{4.111}$$

Their group velocities

$$\mathbf{v}_g \;=\; \frac{\partial \omega_{4,5}}{\partial \mathbf{k}} \;=\; \mathbf{u}_0(\mathbf{x}) \;\pm\; c_s(\mathbf{x})\, \hat{\mathbf{k}} \quad , \tag{4.112}$$

no longer point in the same direction as the wave vector. The group velocity vectors of all possible waves, i.e. all directions $\hat{\mathbf{k}}$, lie on a sphere of radius c_s centered on $\mathbf{u}_0$. If the equilibrium flow is supersonic, $|\mathbf{u}_0| > c_s$, then these vectors all lie within a cone, known as the *Mach cone*, whose axis is along $\mathbf{u}_0$. Those velocity vectors tangent to the sphere define the opening half-angle

$$\alpha_{\rm Mach} \;=\; \sin^{-1}\left(\frac{c_s}{|\mathbf{u}_0|}\right) \;=\; \sin^{-1}(1/M) \quad , \tag{4.113}$$

where $M > 1$ is the Mach number. Thus even a forward wave (ω_4) directed upstream: $\hat{\mathbf{k}} \parallel -\mathbf{u}_0$ sends information *downstream*. No information moves upstream.

For subsonic flows, $|\mathbf{u}_0| < c_s$, on the other hand the sphere of group velocities encloses the origin so $\mathbf{v}_g$ can be made to point in any direction by the appropriate choice of $\hat{\mathbf{k}}$. Upstream-directed waves send information upstream, albeit at speed $c_s - |\mathbf{u}_0|$: slower than the speed of sound.

Chapter 5

Waves in a stratified atmosphere: helioseismology

5A Linearization about a stratified atmosphere

Consider a static, planar atmosphere in a Cartesian coordinate system where z increases *downward*. We make this choice so that $z = 0$ is the *top* of the atmosphere where $p = 0$; because of this choice $\mathbf{g} = g\hat{\mathbf{z}}$. The equation for hydrostatic equilibrium is

$$\nabla p_0 \;=\; \rho_0 \mathbf{g} \quad\Longrightarrow\quad \partial_z p_0 = \rho_0 g \quad . \tag{5.1}$$

This equation alone does not fully specify the solution, $p_0(z)$ or $\rho_0(z)$. It turns out we can do most of the analysis without committing to this significant decision.

Linearizing the fluid equations about this equilibrium and assuming perturbations depend only on x, z

$$\begin{aligned}
\partial_t \rho_1 &= -u_z \partial_z \rho_0 - \rho_0 \nabla\cdot\mathbf{u} && (5.2)\\
\rho_0 \partial_t u_x &= -\partial_x p_1 && (5.3)\\
\rho_0 \partial_t u_y &= -\partial_y p_1 \;=\; 0 && (5.4)\\
\rho_0 \partial_t u_z &= -\partial_z p_1 + g\rho_1 && (5.5)\\
\partial_t p_1 &= -u_z \partial_z p_0 - \gamma p_0 \nabla\cdot\mathbf{u} \;=\; -\rho_0 g u_z - \gamma p_0 \nabla\cdot\mathbf{u} && (5.6)
\end{aligned}$$

We see from (5.4) that u_y will never change. This is one component of the zero-frequency shear mode. We hereafter neglect u_y in our analysis.

5B Normal modes: the waves

Assume normal modes: $\propto e^{-i\omega t}$. Typically we would use a different symbol for the variables, $\tilde{p}$ instead of p_1 — we'll stick with the old variables and remember that they no longer depend on t.

$$\begin{aligned}
i\omega\rho_1 &= u_z \partial_z \rho_0 + \rho_0 \nabla\cdot\mathbf{u} && (5.7)\\
i\omega\rho_0 u_x &= \partial_x p_1 && (5.8)
\end{aligned}$$

$$i\omega\rho_0 u_z = \partial_z p_1 - g\rho_1 \tag{5.9}$$
$$i\omega p_1 = \rho_0 g u_z + \gamma p_0 \nabla\cdot\mathbf{u} \tag{5.10}$$

We need to solve these coupled PDEs to find the eigenfrequencies and modes. Since the coefficients of terms involve $p_0(z)$ and $\rho_0(z)$ we cannot assume plane wave solutions — at least not $\propto e^{ik_z z}$. Instead we proceed by a kind of Guassian row-elimination, eliminating one unknown at a time, constructing PDEs of increasing order, until we face a single high-order PDE for one unknown function and ω.

We begin by eliminating the "least coupled" variables. The horizontal velocity u_x enters other equation only through $\nabla\cdot\mathbf{u}$. Since $\partial_x\rho_0 = 0$ we can differentiate (5.8) to get

$$i\omega\rho_0\partial_x u_x = \partial_x^2 p_1 \ .$$

This means that

$$i\omega\rho_0\nabla\cdot\mathbf{u} = \partial_x^2 p_1 + i\omega\rho_0\partial_z u_z \tag{5.11}$$

Multiplying the energy eq., (5.10), by $-i\omega$ we get

$$\begin{aligned}\omega^2 p_1 &= -i\omega\rho_0 g u_z - \gamma p_0(i\omega\nabla\cdot\mathbf{u}) = -i\omega\rho_0 g u_z - c_s^2(i\omega\rho_0\nabla\cdot\mathbf{u}) \\ &= -(i\omega\rho_0 g)u_z - c_s^2\partial_x^2 p_1 - i\omega c_s^2\rho_0\partial_z u_z\end{aligned} \tag{5.12}$$

This shows an energy equation without u_x.

Next we eliminate ρ_1, which enters into only the vertical momentum equaiton, (5.9). Multiplying the continuity eq., (5.7), by $-i\omega$ we get

$$\begin{aligned}\omega^2\rho_1 &= -i\omega u_z\partial_z\rho_0 - i\omega\rho_0\nabla\cdot\mathbf{u} = -i\omega u_z\partial_z\rho_0 - \partial_x^2 p_1 - i\omega\rho_0\partial_z u_z \\ &= -\partial_x^2 p_1 - i\omega\partial_z(\rho_0 u_z)\end{aligned} \tag{5.13}$$

This can be used to eliminate ρ_1 from the vertical momentum equation. Multiplying eq. (5.7) by ω^2

$$i\omega^3\rho_0 u_z = \omega^2\partial_z p_1 - g\omega^2\rho_1 = \omega^2\partial_z p_1 + g\partial_x^2 p_1 + i\omega g\partial_z(\rho_0 u_z) \tag{5.14}$$

Equations (5.12) and (5.14) are two second-order PDEs for the two remaining unknowns, p_1 and u_z.

The equilibrium depends on z but not on x. This means we *can* assume modes which are plane wave in x: $\propto e^{ik_x x}$. Doing so in (5.12) gives

$$\omega^2 p_1 = -i(\omega\rho_0 g)u_z + c_s^2 k_x^2 p_1 - i\omega c_s^2\rho_0\partial_z u_z \ .$$

since $\partial_x^2 \to -k_x^2$. The can be solved for pressure explicitly:

$$\underbrace{(\omega^2 - c_s^2 k_x^2)}_{R}\, p_1 = -(i\omega\rho_0 g)u_z - i\omega c_s^2\rho_0\partial_z u_z = -i\omega[\, g\rho_0 u_z + c_s^2\rho_0\partial_z u_z\,] \tag{5.15}$$

The explicit form of the pressure is

$$p_1 = -i\omega\left[\frac{c_s^2\rho_0}{R}\partial_z u_z + \frac{g\rho_0}{R}u_z\right] \tag{5.16}$$

$$
\begin{aligned}
\omega^2\rho_0 u_z &= -i\omega\partial_z p_1 - \frac{gk_x^2}{i\omega}p_1 \;+\; g\partial_z(\rho_0 u_z) \\
&= -\omega^2\partial_z\left[\frac{c_s^2\rho_0}{R}\partial_z u_z + \overbrace{\frac{g\rho_0}{R}u_z}^{1}\right] + gk_x^2\left[\overbrace{\frac{c_s^2\rho_0}{R}}^{2}\partial_z u_z + \overbrace{\frac{g\rho_0}{R}}^{3}u_z\right] + \overbrace{g\partial_z(\rho_0 u_z)}^{4} \\
&= -\omega^2\partial_z\left[\frac{c_s^2\rho_0}{R}\partial_z u_z\right] \overbrace{-\underbrace{\omega^2\frac{g\rho_0}{R}}_{1'}\partial_z u_z + \underbrace{\frac{gk_x^2c_s^2\rho_0}{R}}_{2}\partial_z u_z + \underbrace{g\rho_0}_{4'}\partial_z u_z}^{=0} \\
&\quad \underbrace{-\omega^2\partial_z\left[\frac{g\rho_0}{R}\right]}_{1''\to 5}u_z + \underbrace{\frac{g^2k_x^2\rho_0}{R}}_{3}u_z + \underbrace{g(\partial_z\ln\rho_0)}_{4''}\rho_0 u_z \\
&= -\omega^2\partial_z\left[\frac{c_s^2\rho_0}{R}\partial_z u_z\right] - \overbrace{\omega^2\frac{g\rho_0(\partial_z\ln\rho_0)}{R}}^{5'}u_z + \overbrace{\frac{g^2k_x^2\rho_0}{R}}^{3}u_z \\
&\quad + \underbrace{g(\partial_z\ln\rho_0)}_{4''}\rho_0 u_z - \underbrace{\omega^2 g\rho_0\partial_z\left[\frac{1}{R}\right]}_{5''}u_z \\
\omega^2\rho_0 u_z &= -\omega^2\partial_z\left[\frac{c_s^2\rho_0}{R}\partial_z u_z\right] - \overbrace{\frac{gc_s^2k_x^2(\partial_z\ln\rho_0) - g^2k_x^2}{R}}^{5'+3+4''}\rho_0 u_z - \omega^2 g\partial_z\left[\frac{1}{R}\right]\rho_0 u_z
\end{aligned}
$$

SIDEBAR: The step-bystep derivation of eq. (5.18) by eliminating p_1 from eq. (5.17) using eq. (5.16).

The vertical momentum equation, divided by $i\omega$ is

$$
\omega^2\rho_0 u_z \;=\; -i\omega\partial_z p_1 - \frac{gk_x^2}{i\omega}p_1 \;+\; g\partial_z(\rho_0 u_z) \quad . \tag{5.17}
$$

Eliminating p_1 using (5.16) leads to an equation of six terms. This can be reduced, through a lengthy series of steps presented as a SIDEBAR, to a slightly more concise and informative differential equation for u_z alone,

$$
\begin{aligned}
\omega^2\rho_0 u_z &= -\omega^2\partial_z\left[\frac{c_s^2\rho_0}{\omega^2 - k_x^2c_s^2}\partial_z u_z\right] - \frac{g^2k_x^2\xi(z)}{\omega^2 - k_x^2c_s^2}\rho_0 u_z - \partial_z\left[\frac{\omega^2 g}{\omega^2 - c_s^2k_x^2}\right]\rho_0 u_z \qquad (5.18)\\
&= -\omega^2\partial_z\left[\frac{c_s^2\rho_0}{\omega^2 - k_x^2c_s^2}\partial_z u_z\right] - \frac{g^2k_x^2\xi(z)}{\omega^2 - k_x^2c_s^2}\rho_0 u_z - \frac{\omega^2g^2k_x^2[\gamma - 1 - \xi(z)]}{(\omega^2 - c_s^2k_x^2)^2}\rho_0 u_z \quad ,
\end{aligned}
$$

where we have introduced function

$$
\xi(z) \;=\; \frac{c_s^2}{g}\frac{\partial_z\rho_0}{\rho_0} \;-\; 1 \quad , \tag{5.19}
$$

which depends on the structure of the equilibrium atmosphere. This is a general kind of eigenfunction equation. Its solution, subject to boundary conditions on the eigenfunction $u_z(z)$, will be possible only for specific values of ω^2: the eigenfrequencies.

The Entropy profile

The function defined in expression (5.19), $\xi(z)$, is a local property of the atmosphere which turns out to be very important. To understand its meaning and significance let us use the definition of specific entropy

$$p_0(z) \;=\; \rho_0(z)^\gamma \exp\left[\frac{(\gamma-1)}{k_b/m} s_0(z)\right] \tag{5.20}$$

Taking the derivative

$$\frac{dp_0}{dz} \;=\; \gamma\frac{p_0}{\rho_0}\frac{d\rho_0}{dz} \;+\; \frac{(\gamma-1)}{k_b/m} p_0 \frac{ds_0}{dz} \tag{5.21}$$

Setting this equal to $\rho_0 g$, in accordance with the equation of equilibrium, eq. (5.1), we find

$$1 \;=\; \frac{c_s^2}{g}\frac{d\ln\rho_0}{dz} \;+\; \frac{(\gamma-1)T}{g}\frac{ds_0}{dz} \tag{5.22}$$

and thus

$$\xi(z) \;=\; -\frac{(\gamma-1)T}{g}\frac{ds_0}{dz} \;=\; -\frac{k_B T}{mg}\frac{d}{dz}\ln\left(\frac{p_0}{\rho_0^\gamma}\right) \tag{5.23}$$

This shows that $\xi(z)$ is related to the entropy gradient in the hydrostatic atmosphere. Recall that we never committed to the full structure of the hydrostatic equilibrium. Had we chosen an isentropic atmosphere,[1] $ds_0/dz = 0$, we would have found $\xi(z) = 0$.

More general expressions are facilitated by the transform

$$\xi \;=\; \frac{c_s^2}{g}\frac{d\ln\rho_0}{dz} - 1 = \frac{c_s^2}{g}\left(\frac{dp_0/dz}{p_0}\right)\left(\frac{d\ln\rho_0}{d\ln p_0}\right) - 1 \;=\; \gamma\left(\frac{d\ln\rho_0}{d\ln p_0}\right) - 1 \quad . \tag{5.24}$$

For an isothermal atmosphere where $\ln(p_0/\rho_0)$ is a constant we find $d\ln p_0/d\ln\rho_0 = 1$ and $\xi = \gamma - 1$. In a polytrope where $p_0 \sim \rho_0^\Gamma$, $d\ln p_0/d\ln\rho_0 = \Gamma$ and we have

$$\xi \;=\; \frac{\gamma}{\Gamma} \;-\; 1 \quad . \tag{5.25}$$

Polytropic index $\Gamma > \gamma$ gives $\xi < 0$ — entropy increasing with depth — which we find below are interesting.

5C Solving the Sturm-Liouville equation — the eikenol limit

An easy and usually informative way to solve a complicated equation is to use the eikenol approximation. We assume an oscillatory eigenmode

$$u_z(z) \;\propto\; e^{i\phi(z)}$$

[1] There are, unfortunately, some in the field who use the term "adiabatic" here.

where $\phi(z)$ is the unknown phase function. This only helps us because we also assume the eigenfunction oscillates *rapidly* — its length scale of oscillation is much shorter than all other length scales in the problem. This means the derivative of the phase function

$$k_z = \frac{d\phi}{dz} \;=\; k_z(z) \quad , \tag{5.26}$$

is greater than all other derivatives in the problem it is even larger than its own derivative

$$\frac{dk_z}{dz} = \frac{d^2\phi}{dz^2} \ll \left|\frac{d\phi}{dz}\right|^2 \quad . \tag{5.27}$$

Under this assumption we simply replace every $\partial_z \to ik_z$ just as we would have for a plane wave. Doing this to (5.18) gives

$$-\frac{\omega^2 c_s^2 k_z^2}{\omega^2 - c_s^2 k_x^2}\,\rho_0 u_z = -\left\{\,\omega^2 + \frac{g^2 k_x^2 \xi(z)}{\omega^2 - k_x^2 c_s^2} + \omega^2 \frac{g^2 k_x^2 [\gamma - 1 - \xi(z)]}{(\omega^2 - k_x^2 c_s^2)^2}\,\right\}\rho_0 u_z \tag{5.28}$$

Multiplying by $(\omega^2 - k_x^2 g^2)^2$ gives and dividing by $\rho_o u_z$

$$-\omega^2 c_s^2 k_z^2 (\omega^2 - k_x^2 c_s^2) = -\omega^2(\omega^2 - k_x^2 c_s^2)^2 - (\omega^2 - k_x^2 c_s^2) g^2 k_x^2 \xi(z) - \omega^2 g^2 k_x^2 [\gamma - 1 - \xi(z)] \quad ,$$

from which we find

$$\omega^2[\omega^2 - (k_x^2 + k_z^2)c_s^2](\omega^2 - k_x^2 c_s^2) \;=\; -\omega^2(\gamma - 1)g^2 k_x^2 + k_x^4 c_s^2 g^2 \xi(z) \quad . \tag{5.29}$$

Expression (5.29) is a cubic in ω^2 from which we should expect three roots. Before proceeding we note that according the the eikenol approximation $k_x \gg g/c_s^2$. It follows that

$$k_x^2 c_s^2 \;\gg\; g k_x \quad . \tag{5.30}$$

Using this we may drop the term $\propto \omega^2$ on the right, with coefficient $\gamma g^2 k_x^2$, which is much smaller than the corresponding term with coefficient $c_s^4 k_x^2 k^2$, which currently appears on the left. The resulting cubic

$$\omega^2(\omega^2 - k^2 c_s^2)(\omega^2 - k_x^2 c_s^2) \;=\; g^2 k_x^4 c_s^2\, \xi(z) \quad , \tag{5.31}$$

has been written with its small constant term on the right.

It is notable that (5.31) no longer contains any contribution from the final term in the curly brackets of of eq. (5.28); it is negligible in the eikenol limit. That was the term with $(\omega^2 - k_x^2 c_s^2)^2$ in the denominator, which led to a cubic. Had we dropped it at that stage, and performed similar steps, we would end up with a quadratic equation for ω^2

$$\omega^2(\omega^2 - k^2 c_s^2) \;=\; g^2 k_x^4 c_s^2\, \xi(z) \quad . \tag{5.32}$$

Solving this for ω^2 gives two dispersion relations, corresponding to the upper and lower signs,

$$\omega^2 \;=\; \tfrac{1}{2} c_s^2 k^2 \pm \tfrac{1}{2}\sqrt{c_s^4 k^4 - 4 g^2 k_x^2\, \xi} \;=\; \tfrac{1}{2} c_s^2 k^2 \pm \tfrac{1}{2} c_s^2 k^2 \sqrt{1 - \frac{4 g^2 k_x^2}{c_s^4 k^4}\,\xi} \tag{5.33}$$

For each choice of sign there are two eigenfrequencies, one positive and one negative. These would seem to be the four normal modes corresponding to the four degrees of freedom: ρ, u_x, u_z and p. Making further use of eq.. (5.30) we can expand the square-root to get

$$\omega^2 \simeq \tfrac{1}{2}c_s^2k^2 \pm \tfrac{1}{2}c_s^2k^2 \mp \frac{g^2k_x^2}{c_s^2k^2}\,\xi \tag{5.34}$$

The top and bottom signs yield the eigenfrequencies

$$\omega_p^2 \simeq c_s^2k^2 \quad , \quad \omega_g^2 \simeq \frac{g^2}{c_s^2}\,\xi\,\sin^2\theta_k = N^2\sin^2\theta_k \tag{5.35}$$

where

$$N^2 = \frac{g^2}{c_s^2}\,\xi = -\,g\left[\frac{1}{\gamma}\frac{d\ln p_0}{dz} - \frac{d\ln\rho_0}{dz}\right] \tag{5.36}$$

is the square of the Brunt-Väisälä frequency N. These two eigenmodes are called *p-modes* and *g-modes* or *gravity waves* respectively. We can return to the unstratified (i.e. homogeneous) case by taking $g \to 0$ and c_s uniform in space. In that case the p-mode frequencies go over to acoustic modes – they are sound waves. The g-modes become zero-frequency modes, of which the homogeneous system has three. The shear mode polarized in the $\hat{\mathbf{y}}$ direction still exists as such, so the g-mode must account for the other two: the entropy mode and the shear mode with $\hat{\mathbf{x}}$ polarization. While there is no restoring force for either perturbation in an unstratified atmosphere, gravity supplies a restoring force. This why these are called *gravity waves*.

What should we think of the third root of the cubic, eq. (5.31), which was eliminated when we reduced it to a quadratic. Though general cubic equations are very difficult to solve in general, eq. (5.31) is written in a partially factored from $(x-x_1)(x-x_2)(x-x_3) = \epsilon$, where the right hand side, ϵ, is in some sense small. The three roots of such a cubic can be approximated

$$x_1 + \frac{\epsilon}{(x_1-x_2)(x_1-x_3)} \quad , \quad x_2 + \frac{\epsilon}{(x_2-x_1)(x_2-x_3)} \quad , \quad x_3 + \frac{\epsilon}{(x_3-x_1)(x_3-x_2)} \quad .$$

Using $x_1 = 0$, $x_2 = k^2c_s^2$ and $x_3 = k_x^2c_s^2$ yields the three approximate roots of the cubic

$$\omega_g^2 = \frac{g^2k_x^2\xi}{k^2c_s^2} \quad , \quad \omega_p^2 = c_s^2k^2 + \frac{g^2k_x^4\xi}{k^2k_z^2c_s^2} \quad , \quad \omega_s^2 = c_s^2k_x^2 - \frac{g^2k_x^2\xi}{k_z^2c_s^2} \quad . \tag{5.37}$$

The first two roots, ω_g^2 and ω_p^2 are recognizable as the roots of the quadratic, the g-mode and the p-mode; the latter is slightly modified. The third one, which is new, is a frequency very close to that of a horizontally propagating should wave: c_sk_x. This serves as a cut-off frequency for sound waves, and provided $\xi > 0$, ω_s is *below* that cut-off. When we solve the full differential equation we find this mode is confined to the surface.

Ray paths

Dispersion relations eq. (5.35) allow us to find the ray path of each wave, which turn out to differ in illuminating ways. Considering first the p-mode with positive frequency, $\omega = c_sk$, we find its group velocity from eq. (4.90),

$$\mathbf{v}_g = \frac{\partial\omega}{\partial\mathbf{k}} = c_s\frac{\partial|\mathbf{k}|}{\partial\mathbf{k}} = c_s\frac{\mathbf{k}}{k} = c_s\hat{\mathbf{k}} \ , \tag{5.38}$$

to be directed along the wave vector's direction and independent of wavelength. This conforms to our general experience with rays of other kinds, for example light.

The ray does not, however, travel in a straight line since the wave vector changes direction according to eq. (4.91)

$$\frac{d\mathbf{k}}{dt} \;=\; -\nabla\omega \;=\; -k\nabla c_s \;=\; -k\frac{\partial c_s}{\partial z}\,\hat{\mathbf{z}} \quad . \tag{5.39}$$

This shows that k_x remains constant, but that k_z will vary. We expect $\partial c_s/\partial z > 0$ since temperature will generally increase with increasing pressure at increasing depth. A downward propagating wave ($k_z > 0$) will then refract away from the vertical ($|k_z/k_x|$ decreasing) as it travels into deeper layers with higher sound speed.

Instead of solving the equation to determine k_z we can appeal to the fact that ω is constant along the ray and therefore

$$k_z \;=\; \pm\sqrt{k^2 - k_x^2} \;=\; \pm\sqrt{\frac{\omega^2}{c_s^2} - k_x^2} \;=\; \pm\, k_x\sqrt{\frac{\omega^2}{c_s^2 k_x^2} - 1} \quad , \tag{5.40}$$

with upper and lower signs applying to propagating downward and upward directed rays (assuming for concreteness that $k_x > 0$). This shows that the ray will not enter regions where $c_s(z) > \omega/k_x$. The limit of this forbidden region is a turning point where $k_z = 0$, and the ray is directed horizontally. Application of eq. (5.39) reveals that the wave changes from downward to upward propagation at the turning point. At the top surface, $c_s(0) = 0$ and vertical wave number $k_z \to \infty$. This means that ray the propagates vertically and the wavelength goes to zero as wavefronts pile together. Thus the ray is trapped between the surface and the turning point where $c_s(z_{tp}) = \omega/k_x$.

In a polytropic atmosphere, with index Γ, expression (2.18) can be used to write the sound speed

$$c_{\rm s}^2 \;=\; \gamma\frac{k_{\rm B}T}{m} \;=\; \frac{\gamma}{\Gamma}(\Gamma-1)\,g\,z \quad . \tag{5.41}$$

The lower turning point therefore occurs at the depth

$$z_{tp} \;=\; \frac{\Gamma}{\gamma}\frac{\omega^2}{(\Gamma-1)\,g k_x^2} \quad . \tag{5.42}$$

The ray path follows a curve $x(z)$ satisfying the equation

$$\frac{dx}{dz} \;=\; \frac{v_{g,x}}{v_{g,z}} \;=\; \frac{\hat{k}_x}{\hat{k}_z} \;=\; \frac{k_x}{k_z} \;=\; \pm\left(\frac{z_{tp}}{z} \;-\; 1\right)^{-1/2} \quad , \tag{5.43}$$

whose solution

$$x(z) \;=\; x_0 \;\pm\, z_{tp}\,\sin^{-1}\left(\sqrt{\frac{z}{z_{tp}}}\right) \;\mp\; \sqrt{z}\sqrt{z_{tp}-z} \quad , \tag{5.44}$$

is a *cycloid* with cusps along the surface, $z = 0$, as shown in fig. 5.1a. It is at the cusps that k_z passes from $-\infty$ to $+\infty$, representing a reflection form upward to downward. It is evident that the ray is refracted away from hotter depths toward the surface at which it repeatedly reflects. The distance between reflections

$$\Delta x \;=\; \pi\, z_{tp} \;=\; \frac{\Gamma}{\gamma}\frac{\pi\omega^2}{(\Gamma-1)\,g k_x^2} \quad , \tag{5.45}$$

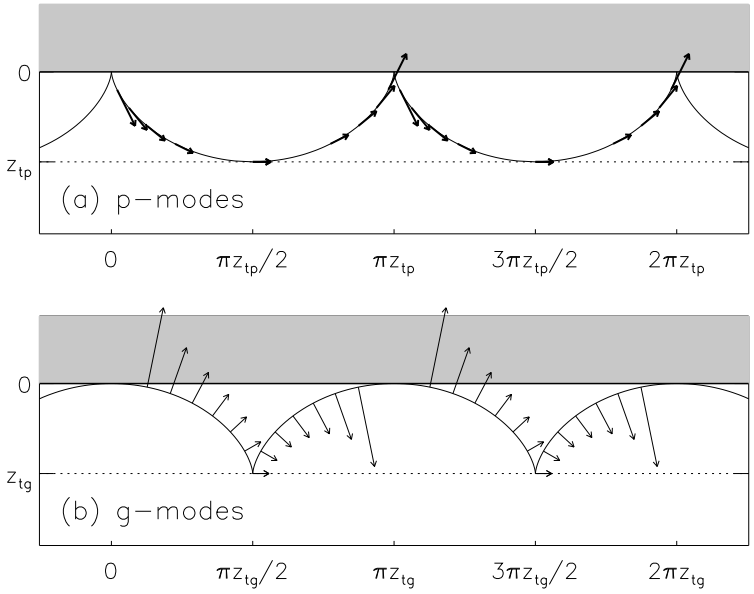

Figure 5.1: The ray path of a p-mode (a) and a g-mode (b) plotted as solid curves. Both rays propagate from left to right. Arrows show the wave vector $\mathbf{k}$ at selected points along each path. The white region is the atmosphere, $z > 0$ and the grey region is space above the atmosphere. The dotted curve shows the turning points, z_{tp} and z_{tg}.

scales inversely with horizontal wavenumber *squared*. Recall that the assumption underlying the eikenol approach, $k_x \Delta x \sim \omega^2/g k_x \gg 1$, means the space between reflections is *many* wavelengths.

The rays paths from g-modes are far more surprising, and more interesting, than those from p-modes. The frequency from eq. (5.35), can be written

$$\omega(\mathbf{k}) \;=\; N \sin\theta_k \;=\; N \frac{\hat{\mathbf{x}}\cdot\mathbf{k}}{|\mathbf{k}|} \quad ,$$

where the Brunt-Väisälä frequency, N is taken to be real for the mode to be oscillatory. The group velocity of this wave

$$\mathbf{v}_g \;=\; \frac{\partial\omega}{\partial\mathbf{k}} \;=\; N\left[\frac{\hat{\mathbf{x}}}{k} \;-\; \frac{(\hat{\mathbf{x}}\cdot\mathbf{k})\mathbf{k}}{k^3}\right] \;=\; \frac{N}{k}\Big(\underline{\underline{\mathrm{I}}} - \hat{\mathbf{k}}\hat{\mathbf{k}}\Big)\cdot\hat{\mathbf{x}} \quad , \tag{5.46}$$

is already more interesting than the p-mode case. First of all, the group velocity (as well as the phase velocity) depends on wave-number k, so these waves are dispersive: longer wavelengths (i.e. smaller ks) travel faster. It can be shown that the magnitude of the group velocity is

$$|\mathbf{v}_g| \;=\; \frac{N}{k}\sqrt{1-(\hat{\mathbf{k}}\cdot\hat{\mathbf{x}})^2} \;=\; \frac{N}{k}|\cos\theta_k| \;=\; v_\varphi|\cot\theta_k| \quad , \tag{5.47}$$

where $v_\varphi = \omega/k = N\sin\theta_k/k$ is the phase speed. So the group speed and phase speeds differ.

Even more surprising is that the energy propagates *at right angles* to the wave vector. This is apparent once it is recognized that $\underline{\underline{\mathrm{I}}} - \hat{\mathbf{k}}\hat{\mathbf{k}}$ is a projection matrix: it projects any vector into the plane perpendicular to $\hat{\mathbf{k}}$. Thus the group velocity is always perpendicular to the wave vector direction $\hat{\mathbf{k}}$: $\hat{\mathbf{k}} \cdot \mathbf{v}_g = 0$. It is equally curious that the group velocity does not change if we change the sign of the wave vector: $\hat{\mathbf{k}} \to -\hat{\mathbf{k}}$. The positive-frequency waves always propagate to the right and the negative-frequency waves ($\omega = -N \sin\theta_k$) to the left. In more familiar cases, like the p-mode in eq. (5.38), the positive frequency wave propagates in the direction of $\hat{\mathbf{k}}$, so its group velocity changes sign with $\hat{\mathbf{k}}$.

The equation for the ray path, derived from (5.46),

$$\frac{dx}{dz} \;=\; \frac{v_{g,x}}{v_{g,z}} \;=\; \frac{1-(\hat{\mathbf{x}}\cdot\hat{\mathbf{k}})^2}{-(\hat{\mathbf{z}}\cdot\hat{\mathbf{k}})(\hat{\mathbf{x}}\cdot\hat{\mathbf{k}})} \;=\; -\frac{\cos\theta_k}{\sin\theta_k} \quad ,$$

depends only on the angle of $\hat{\mathbf{k}}$ relative to vertical. Once again we can use to effect the fact that ω is constant along the ray path, in eq. (5.35),

$$\sin\theta_k \;=\; \frac{\omega}{N} \quad , \quad \cos\theta_k \;=\; \pm\sqrt{1-\frac{\omega^2}{N^2}} \quad ,$$

where $\cos\theta_k$ has the same sign as k_z. The ray path then satisfies

$$\frac{dx}{dz} \;=\; \mp\sqrt{\frac{N^2}{\omega^2} \;-\; 1} \;=\; \mp\sqrt{\frac{g^2\xi}{c_{\rm s}^2\omega^2} \;-\; 1} \quad , \tag{5.48}$$

after using (5.36) for the Brunt-Väisälä frequency. Like the p-mode, the g-modes are confined to a layer along the surface; in this case where $c_{\rm s} \le \sqrt{\xi}g/\omega$.

If we again take a polytropic atmosphere we find the g-mode turning point to be

$$z_{tg} \;=\; \frac{g\Gamma\xi}{\gamma(\Gamma-1)\omega^2} \;=\; \frac{g(\gamma-\Gamma)}{\gamma(\Gamma-1)\omega^2} \quad , \tag{5.49}$$

after using expression (5.25) to replace ξ. The ray equation for a g-mode

$$\frac{dx}{dz} \;=\; \mp\sqrt{\frac{g\Gamma\xi}{\gamma(\Gamma-1)\omega^2\,z} \;-\; 1} \;=\; \mp\sqrt{\frac{z_{tg}}{z} \;-\; 1} \quad , \tag{5.50}$$

differs from eq. (5.43) for a p-mode, but also describes a *cycloid*

$$x(z) \;=\; x_0 \;\mp\; z_{tg}\,\sin^{-1}\left(\sqrt{\frac{z}{z_{tg}}}\right) \;\mp\; \sqrt{z}\sqrt{z_{tg}-z} \quad . \tag{5.51}$$

In this case the cusps of the cycloid lie at the turning points rather than at the surface, as shown in fig. 5.1b. It is there that $\omega = N$ and the wave vector is purely horizontal ($\sin\theta_k = \omega/N = 1$). Remarkably the ray path is vertical there and it reflects as k_z pass through zero.

Because of invariance in x Hamilton's equations once again call for the horizontal wave vector to be constant

$$\frac{dk_x}{dt} \;=\; -\frac{\partial\omega}{\partial x} \;=\; 0 \quad . \tag{5.52}$$

The vertical component

$$k_z \;=\; k_x \frac{\cos\theta_k}{\sin\theta_k} \;=\; \pm\, k_x \sqrt{\frac{N^2}{\omega^2} \;-\; 1} \;=\; \pm\, k_x \sqrt{\frac{z_{tg}}{z} \;-\; 1}\;\;, \tag{5.53}$$

diverges at the surface as it did for the p-modes. It is there that k_z passes from $-\infty$ to $+\infty$ to turn the upward ray back downward. Oddly this reflection occurs as the ray *grazes* the surface horizontally.

WKB method

The WKB method is an application of the eikenol approximation to finding the eigenfrequencies of normal modes. Normal modes are standing waves composed in equal parts of waves propagating in both directions. Both p-modes and g-modes are trapped between the surface ($z = 0$) and some turning point z_t. Standing waves will exist in this cavity, provided they have some integral number of wavelengths. This translates into the condition

$$\int_0^{z_t} k_z\, dz \;=\; \pi(n+\alpha) \;\;, \tag{5.54}$$

where n is a (large) integer giving the mode number: the number of nodes within the cavity. The constant α is related to the behavior of the wave at the reflection points. In the case most commonly encountered k_z goes smoothly to zero and $\alpha = 1/4$; this is true of the lower turning point, $z = z_t$. At the surface, however, $k_z \to \infty$ and a different boundary condition is required, resulting in a different value of α. Rather than delving into this issue we regard α as some fixed but unknown number.

To find normal mode frequencies for p-modes we use expression (5.40) in (5.54),

$$k_x \int_0^{z_{tp}} \left(\frac{z_{tp}}{z} - 1\right)^{-1/2} dz \;=\; k_z z_{tp} \int_0^1 \frac{\sqrt{u}\, du}{\sqrt{1-u}} \;=\; \frac{\pi}{2}\, k_x\, z_{tp} \;=\; \pi(n+\alpha_p) \;\;, \tag{5.55}$$

where we admit the possibility that the upper boundary condition, which would determine α, might depend on the nature of the p-mode. Using eq. (5.42) for z_{tp} yields the spectrum of p-mode frequencies

$$\omega_n^{(p)}(k_x) \;=\; \sqrt{\frac{2\gamma(\Gamma-1)}{\Gamma}}\, \sqrt{g k_x}\, \sqrt{n+\alpha_p}\;. \tag{5.56}$$

The horizontal wave number, k_x, is not fixed by any boundary, so it is completely free. There is therefore a full branch of waves for each n.

For the case of g-modes we eq. use (5.53) in (5.54)

$$k_x \int_0^{z_{tg}} \frac{\cos\theta_k}{\sin\theta_k}\, dz \;=\; k_z z_{tg} \int_0^1 \sqrt{1-u}\, \frac{du}{\sqrt{u}} \;=\; \frac{\pi}{2}\, k_x\, z_{tg} \;=\; \pi(n+\alpha_g) \;\;. \tag{5.57}$$

Using (5.49) yields the normal mode frequencies

$$\omega_n^{(g)}(k_x) \;=\; \sqrt{\frac{(\gamma-\Gamma)}{2\gamma(\Gamma-1)}}\,\sqrt{gk_x}\,\frac{1}{\sqrt{n+\alpha_g}} \quad . \tag{5.58}$$

Comparison between (5.56) and (5.58) reveals that both frequencies depend on k_x in the same way: $\sqrt{gk_x}$. This kind of dispersion relation is common to other waves in stratified media, including surface waves on "deep" water.[2] Indeed, dimensional analysis can be used to show that this form is required when the system has no intrisnic length scale and involves only g. The horizontal phase speed and group velocity from the horizontal form alone

$$v_\varphi(k_x) \;=\; \frac{\omega_n}{k_x} \;\sim\; \left(\frac{g}{|k_x|}\right)^{1/2} \tag{5.59}$$

$$v_g(k_x) \;=\; \frac{\partial\omega_n}{\partial k_x} \;=\; \tfrac{1}{2}v_\varphi \tag{5.60}$$

The speeds depend on wave-length so they are *dispersive* — disturbances lose their shapes as the travel since their various Fourier components lose phase with one another. This is why an impulsive disturbance caused by throwing a pebble into water gives rise to many concentric ripples rather than a single ring. The dispersion relation also results in the wake behind moving ships, which always form a wedge of angle of 39° independent of the ship's speed.

The chief difference between (5.56) and (5.58) is their dependence on mode number n. For a given horizontal wave number k_x p-mode frequencies increase with n while g-mode frequencies decrease with n. This arises from the opposing dependence on ω of turning points. Higher mode numbers need to extend deeper. A deeper turning point requires p-modes of higher frequency and g-modes of lower frequency.

Expressions (5.56) and (5.58) are strictly only valid for very short wave numbers, thus large values of n. In the next sections we consider the waves without this restriction. It turns out that the same forms apply in this case as well.

5D Gravity waves a.k.a. g-modes

The fist thing to notice about g-modes is that they only oscillate when $\xi > 0$. If $\xi < 0$ then $\omega_g^2 < 0$ and the two eigenfrequencies are purely imaginary: $\omega_g = \pm i\sigma$. The one on the positive imaginary axis

$$e^{-i\omega_g t} \;\simeq\; e^{+\sigma t}$$

is purely growing in time: it is an instability called the *convective instability*. This instability occurs when $\xi < 0$, which is when entropy increases with depth.

We have assumed an oscillatory structure of the vertical velocity $u_z(z)$. To relate this to the other fields we substitute the frequency into each of the relationships we used to eliminate them. This is much simpler if we use the fact that $\omega \ll c_s k_x \sim c_s k$. Thus eq. (5.15) becomes

$$-k_x^2 c_s^2\, p_1 \;=\; -\,i\omega g\rho_0 u_z \;-\; i\omega\, c_s^2\rho_0\partial_z u_z \;\simeq\; -\,i\omega g\rho_0 u_z \;+\; \omega c_s^2\rho_0 k_z u_z \quad .$$

[2] The depth of the water must be very large compared to the horizontal wavelength.

After solving this

$$p_1 \;=\; i\omega\frac{g}{c_s^2k_x^2}\rho_0 u_z - \omega\frac{c_s^2k_z}{k_x^2}\rho_0 u_z \quad ,$$

we may also use eq. (5.13) to find the density

$$\omega^2\rho_1 \;=\; k_x^2 p_1 - i\omega\partial_z(\rho_0 u_z) \;\simeq\; k_x^2 p_1 + \omega k_z\rho_0 u_z - i\omega(\partial_z\rho_0)u_z \quad .$$

We have kept the term with $\partial_z\rho_0$ even though it is much smaller than that with k_z. We must do this because when the expression for p_1 is substituted in, its second term cancels that larger term. This leaves the density perturbation

$$\rho_1 \;=\; i\frac{g}{\omega c_s^2}\,\rho_0 u_z - i\frac{\partial_z\rho_0}{\omega}u_z \;=\; -\frac{i}{\omega}\frac{g}{c_s^2}\underbrace{\left[(c_s^2/g)\partial_z\ln\rho_0 - 1\right]}_{\xi}\rho_0 u_z \quad , \tag{5.61}$$

where the profile ξ has re-appeared. The overall factor of i shows that, when ω is real, the density perturbation is 90° out of phase with the velocity perturbation. This is very different from a sound wave where density and velocity are either in phase or 180° out of phase, depending on the sense of propagation.

It is easier to understand this behavior in terms of the vertical displacement δz_1 of the fluid elements. Its advective derivative

$$u_z \;=\; \frac{D}{Dt}\delta z_1 \;\simeq\; \partial_t\delta z_1 \;\rightarrow\; -i\omega\delta z_1 \quad . \tag{5.62}$$

since the $\mathbf{u}_1\cdot\nabla\delta z_1$ term is second order. Using this we find

$$\rho_1 \;=\; -\frac{g}{c_s^2}\,\xi\,\rho_0\,\delta z_1 \tag{5.63}$$

For the oscillatory case, $\xi > 0$, the density perturbation is 180° out of phase with the displacement, δz_1, regardless of which way the wave is propagating. When the displacement is downward, $\delta z_1 > 0$, the density is *lower* than its surroundings, $\rho_1 < 0$. This means the fluid element is *underdense*. An underdense element is buoyant and is thus accelerating upward. This is the nature of the restoring force from gravity. Indeed, had we simply equated the buoyancy force, $g\rho_1$, with an acceleration per volume

$$\rho_0\frac{d^2\delta z_1}{\delta t^2} \;=\; g\rho_1 \;=\; -\rho_0\frac{g^2}{c_s^2}\,\xi\,\delta z_1 \tag{5.64}$$

we find that the fluid element should oscillate with angular frequency given by $N = \sqrt{\xi}g/c_s$. The factor of $\sin\theta_k$ evidently comes from the full fluid treatment.

Just as clearly we see that for the unstable cases, $\xi < 0$, *upward* displacement of a fluid element, $\delta z_1 < 0$, leads to an underdense, and therefore buoyant fluid element, $\rho_1 < 0$. The resulting upward acceleration only makes the element more buoyant still and the process runs away. This is the nature of the instability.

Equation (5.25) above showed that for a polytropic atmosphere $\xi < 0$, and therefore instability occurs, when the polytropic index $\Gamma > \gamma$ the adiabatic index. Chapter 2 provided an example of a polytropic atmosphere when conductive heat flux, F, and conductivity κ

are both uniform. In this case the polytropic index Γ is set by these quantities according to (2.23). Instability therefore occurs when the heat flux exceeds a threshold

$$F \;>\; \left(1-\frac{1}{\gamma}\right)\frac{\kappa m g}{k_{\rm B}} \quad . \tag{5.65}$$

The heat flux F carried by the Earth's atmosphere is set by the amount sunlight warming the ground. When the ground heats up beyond a critical point the atmosphere becomes unstable and spontaneous airflows occur (think shimmering over a hot road, or afternoon clouds building up on a hot day). The spontaneous amplification of flow will lead ultimately to turbulence which will have the effect of mixing the entropy and miking it more uniform. If the fluctuations are averaged over the mean atmospheric profile will look more like an isentropic one, $\Gamma = \gamma$, in spite of the larger heat flux being carried.

5E P-modes

To simplify our analysis of p-modes we consider an isentropic atmosphere, $\xi = 0$, so g-modes have zero frequency. In this case the Sturm-Liouville equation, (5.18) takes the simpler form

$$\frac{d}{dz}\left[\frac{c_s^2\rho_0}{\omega^2-k_x^2c_s^2}\frac{du_z}{dz}\right] \;=\; -\rho_0 u_z \;-\; \frac{g k_x^2 \partial_z c_s^2}{(\omega^2-k_x^2c_s^2)^2}\rho_0 u_z \quad . \tag{5.66}$$

We have eliminated the ω^2 factors common to both sides, assuming thereby $\omega \neq 0$. Those zero-frequency modes are the g-modes and we are no longer looking for them.

An isentropic atmosphere is also simper because it has a linear temperature profile

$$T_0(z) \;=\; \frac{(\gamma-1)g}{\gamma k_b/m}\,z \quad , \tag{5.67}$$

meaning a linear squared sound speed

$$c_s^2 \;=\; \gamma(k_b/m)T_0(z) \;=\; (\gamma-1)gz \;=\; \tfrac{2}{3}gz \quad , \tag{5.68}$$

using $\gamma = 5/3$. The pressure and density profiles are related as $p_0 \sim \rho_0^{5/3}$ and $p_0/\rho_0 \sim z$. It can be checked that both conditions are satisfied when

$$p_0(z) \;\propto\; z^{5/2} \quad , \quad \rho_0(z) \;\propto\; z^{3/2} \quad . \tag{5.69}$$

Using all of these profiles in (5.66) gives

$$\frac{d}{dz}\left[\frac{\frac{2}{3}gz^{5/2}}{\omega^2-\frac{2}{3}k_x^2gz}\frac{du_z}{dz}\right] \;=\; -z^{3/2}u_z \;-\; \frac{\frac{2}{3}k_x^2g^2}{(\omega^2-\frac{2}{3}k_x^2gz)^2}\,z^{3/2}u_z \tag{5.70}$$

since the coefficient of ρ_0 appears on both sides. We define a dimensionless frequency

$$\beta = \tfrac{3}{2}\frac{\omega^2}{g|k_x|} \quad , \tag{5.71}$$

and use in a dimensionless length variable $y = |k_x|z/\beta$ (we should not confuse this with the horizontal coordinate — but that coordinate has never appeared in this analysis), and then a rescaled eigenmode

$$u_z(z) = \psi(|k_x|z/\beta) = \psi(y) \quad .$$

The result is a dimensionless equation for dimensionless variables

$$\frac{d}{dy}\left[\frac{y^{5/2}}{1-y}\frac{d\psi}{dy}\right] + \frac{3}{2}\frac{y^{3/2}}{(1-y)^2}\psi = -\beta^2 y^{3/2}\psi \tag{5.72}$$

Equation (5.72) is a Sturm-Liouville equation with eigenvalue β^2. Like all Sturm-Liouville equations, this has a discrete spectrum of eigenvalues, β_n. The corresponding eigenfunctions $\psi_n(y)$ can be normalized so they are ortho-normal

$$\int \psi_m(y)\,\psi_n(y)\,y^{3/2}\,dy = \delta_{mn} \quad . \tag{5.73}$$

The bounds of integration depend on the domain we are considering, and orthogonality follows only if the boundary conditions applied there are homogeneous. Once they are found, we can convert the eigenvalues directly into normal mode frequencies

$$\omega_n(k_x) = \sqrt{\frac{2\beta_n}{3}}\,(g|k_x|)^{1/2} \tag{5.74}$$

This matches expression (5.56) form the WKB treatment. The unknown eigenvalue evidently goes as $\beta_n \sim 2n$ for large n.

Equation (5.72) can be expanded to the form

$$\psi'' + \left(\frac{5/2}{y} + \frac{1}{1-y}\right)\psi' + \left[\frac{3/2}{y(1-y)} - \beta^2 + \frac{\beta^2}{y}\right]\psi = 0 \quad , \tag{5.75}$$

after multiplying by $(1-y)y^{-5/2}$. This equation is singular at two points: $y = 0$ and $y = \beta$. For very large values of y it becomes

$$\psi'' - \beta^2\psi = 0 \quad , \tag{5.76}$$

whose general solutions are $e^{+\beta y}$ and $e^{-\beta y}$. One boundary condition will consist of requiring the asymptotic solution to match only the decaying solution, $\psi \sim e^{-\beta y}$ as $y \to \infty$. The lowest singular point in eq. (5.75), however, it appear that both independant solutions are regular there. The top of the atmosphere, $y = z = 0$, is a second singular point. The independent solutions there, $y \ll 1$, go as

$$\psi(y) \sim y^{-3/4} J_{\pm 3/2}(2\zeta\sqrt{y}) \sim y^{-3/4\pm 3/4} \quad , \quad \zeta = \sqrt{\beta^2 - 3/2} \quad . \tag{5.77}$$

The solution with the lower sign, $\psi(y) \sim y^{-3/2}$, diverges as $y \to 0$, so if that point is included in the domain we must eliminate it by demanding

$$y^{3/2}\,\psi(y) \sim \rho_0(z)\,u_z(z) \to 0 \quad , \quad y,\, z \to 0 \quad . \tag{5.78}$$

n	1	2	3	4	5	6	7	8
β_n	2.93	5.01	7.04	9.06	11.07	13.08	15.09	17.09

Table 5.1: The first eight eigenvalues for eq. (5.72).

This means $\psi \sim u_z$ approaches a constant value at the top of the atmosphere, but the mass flux, $\rho_0 u_z$, vanishes there. It is also notable, that a solution with some of the singular piece, $\psi \sim y^{-3/2}$, could not be normalized in (5.73).

Applying these boundary conditions and solving the Sturm-Liouville equation leads to eigenvalues listed in table 5.1. It is clear that the values are well approximated as $\beta_n \simeq 2n+1$, agreeing with (5.56) after taking $\alpha_n \simeq 1/2$. A few of the eigenfunctions are ploted in fig. 5.2. The structure of mode n is oscillatory down to a depth

$$z_n \;=\; \frac{\beta_n}{|k_x|} \quad , \tag{5.79}$$

(i.e. $y = 1$) and decays below this. The speed of sound at the depth is

$$c_s^2(z_n) \;=\; \tfrac{2}{3} g z_n \;=\; \tfrac{2}{3}\frac{g\beta_n}{|k_x|} \quad , \tag{5.80}$$

after using (5.68). At this point $c_s^2 k_x^2 = \omega^2$, which matches the turning point for rays under the WKB approximation, i.e. z_{tp} in eq. (5.42).

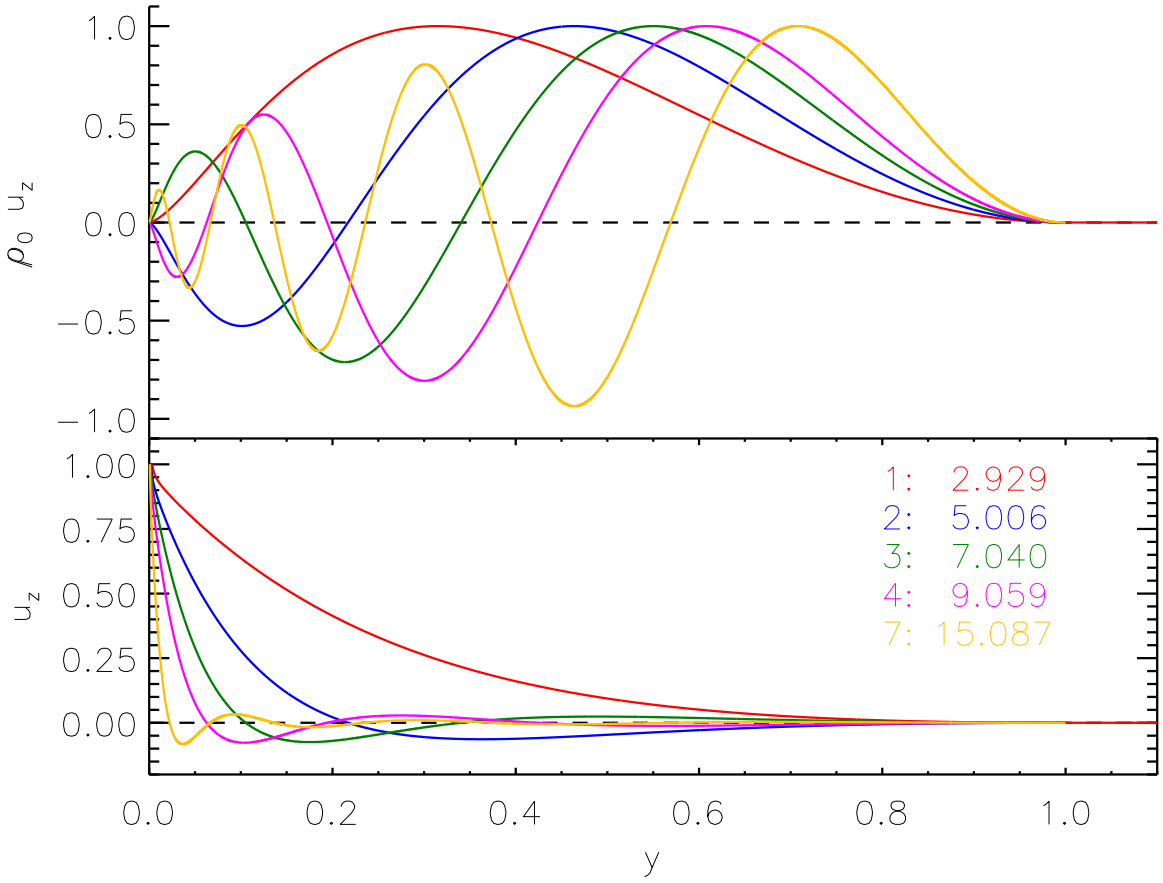

Figure 5.2: The eigenfunctions for $n = 1$ (red), 2, (blue), 3 (green), 4 (magenta) and $n = 7$ (orange). These are plotted as u_z *vs.* y in the bottom panel and $\rho_0 u_z$ *vs.* y in the top panel.

Chapter 6

Turbulent diffusion

6A Turbulent transport for incompressible flows

6A.1 Average and fluctuating quantities

Turbulent diffusion occurs when fluid velocity, $\mathbf{u}(\mathbf{x},t)$, varies rapidly in space and time, seemingly at random. We will, in fact, consider it to be a random variable with specified statistical properties. The averaging operator, $\langle\cdot\rangle$, applies to the random velocity field and to all quantities influenced by it such as temperature. We designate the random quantities with a prime, and averages with an overbar

$$T(\mathbf{x},t) \;=\; \bar{T}(\mathbf{x},t) + T'(\mathbf{x},t) \quad , \quad \langle T\rangle = \bar{T} \quad , \quad \langle T'\rangle \;=\; 0 \quad .$$

For simplicity we consider an incompressible flow which is purely fluctuating, i.e. there is no mean velocity,

$$\nabla\cdot\mathbf{u} \;=\; 0 \quad , \quad \langle\mathbf{u}\rangle = 0 \quad . \tag{6.1}$$

The evolution of temperature is given by eq. (1.15), which becomes, after neglecting thermal conductivity ($\kappa = 0$) and other heating ($\dot{Q} = 0$)

$$\partial_t T \;=\; -\,\mathbf{u}\cdot\nabla T \;=\; -\,\nabla\cdot(\mathbf{u}T) \;=\; -\,\partial_i(u_i T) \quad , \tag{6.2}$$

where summation of repeated indices is implied. Averaging this equation we find

$$\partial_t\bar{T} \;=\; -\,\partial_i(\underbrace{\langle u_i\rangle}_{=0}\bar{T}) - \partial_i\langle u_i T'\rangle \;=\; -\,\partial_i\langle u_i T'\rangle \;=\; -\,\nabla\cdot\mathbf{\Gamma} \tag{6.3}$$

where we have introduced the average heat flux or turbulent heat flux

$$\mathbf{\Gamma} = \langle\mathbf{u}T'\rangle \quad . \tag{6.4}$$

It is essential to note that while both $u_i = u_i'$ and T' vanish upon averaging (by definition) their product does not. This is because the average of a product does not equal the product of the averages. Consider, as an example, a case where leftward ($u_x < 0$) and rightward ($u_x > 0$) motions of equal magnitude occur with equal likelihood in order to make $\langle u_x\rangle = 0$.

Assume as well that regions warmer ($T' > 0$) and cooler ($T' < 0$) than average occur with equal likelihood in order that $\langle T' \rangle = 0$. If, however, the warmer regions occur preferentially in rightward flows and cooler regions in leftward flows then the product $u_x T'$ will tend to be positive — it is positive in each of those cases. This positive tendency means $\Gamma_x = \langle u_x T' \rangle > 0$, implying a net *rightward* flow of heat. This occurs in the absence of a net flow, $\langle u_x \rangle = 0$, and the absence of excess heat $\langle T' \rangle = 0$. It is the correlation between these two fluctuating quantities which gives rise to a mean flux. This raises the question of what might lead to a correlation in the fluctuating quantities.

6A.2 Equation governing evolution of mean quantities

To answer this question we must uncover the relation between flow and temperature. Subtracting eq. (6.3) from eq. (6.2) gives an equation for the evolution of temperature perturbations

$$\partial_t T' \;=\; -u_j \partial_j \bar{T} - \partial_j (u_j T)' \quad . \tag{6.5}$$

where we have introduced a notation for the fluctuating part of a product of fluctuating variables

$$(u_j T)' \;=\; u_j' T' - \langle u_j' T' \rangle \quad . \tag{6.6}$$

This is not necessarily zero but will vary on times scales typically shorter than terms linear in random quantities. We will henceforth ignore it according to a procedure given the name *first-order smoothing*.[1] What remains of eq. (6.5) is linear in T' and is readily solved by integration

$$T'(\mathbf{x}, t) \;=\; -\int^t u_i(\mathbf{x}, s) \partial_j \bar{T}(\mathbf{x}, s)\, ds \quad . \tag{6.7}$$

This formal solution may now be used in eq. (6.3). It will appear inside the average

$$\Gamma_i \;=\; \langle u_i T' \rangle \;=\; -\int^t \Big\langle u_i(\mathbf{x}, t) u_j(\mathbf{x}, s) \Big\rangle\, \partial_j \bar{T}(\mathbf{x}, s)\, ds \tag{6.8}$$

The average in the integrand, $\langle u_i(\mathbf{x}, t) u_j(\mathbf{x}, s) \rangle$, is one statistical property of the randomly fluctuating velocity field — one of many such averages used to characterize turbulence. The particular case is called the *one-point, two-time velocity correlation*. It is possible, in principle, to use the momentum equation to determine properties of this tensor — but nobody has yet managed to do so in practice. Such *a priori* characterization is the aim of turbulence theory, a still-ongoing enterprise. Instead of using first-principles, we will make several assumptions about the turbulence from which we can draw conclusions abut the velocity correlation.

We will first of all make the assumption of *statistically steady turbulence*. This means that while random quantities such as $\mathbf{u}(\mathbf{x}, t)$, vary in time their statistical properties do not. Under this assumption, the average of any quantity of time, $f(t)$, is independent of t. The average of two-time product, $\langle f(t) g(s) \rangle$, is independent of either t or s separately. Instead, the average depends only on the difference, $t - s$. Thus the velocity correlation is

$$\langle u_i(\mathbf{x}, t) u_j(\mathbf{x}, s) \rangle \;=\; V_{ij}(\mathbf{x}, t - s) \quad . \tag{6.9}$$

[1] This step is rather difficult to justify, but is very common. A cynic might conclude the fancy name is used in place of an actual justification.

The one-time correlation

$$\langle u_i(\mathbf{x},t)u_j(\mathbf{x},t)\rangle \;=\; V_{ij}(\mathbf{x},0) \quad , \tag{6.10}$$

characterizes the turbulence at any point in space, $\mathbf{x}$. Half its trace gives the specific kinetic energy of the fluid

$$\varepsilon(\mathbf{x}) \;=\; \tfrac{1}{2}\langle|\mathbf{u}|^2\rangle \;=\; \tfrac{1}{2}V_{ii}(\mathbf{x},0) \;=\; \tfrac{1}{2}[V_{xx}(0)+V_{yy}(0)+V_{zz}(0)] \quad . \tag{6.11}$$

This is the most intuitive measure of how vigorous the turbulence is.

Consider a single term of the correlation tensor, for example

$$V_{xx}(\Delta t) \;=\; \langle u_x(s+\Delta t)u_x(s)\rangle \quad ,$$

where the spatial coordinate $\mathbf{x}$, which would be repeated for each quantity, is omitted for clarity. The one-time case, $\Delta t = 0$ gives the V_{xx} contribution to ε. The velocity fluctuations at one point have a typical time scale, $\tau_{\rm cor}$, on which they are self-correlated. After an interval much longer than this, $\mathbf{u}(t)$ and $\mathbf{u}(t+\Delta t)$ are virtually independent of one another and

$$V_{xx}(t-s) \;=\; \langle u_x(t)u_x(s)\rangle \;\simeq\; \langle u_x(t)\rangle\langle u_x(s)\rangle \;=\; 0 \quad , \quad |t-s| \gg \tau_{\rm cor} \quad . \tag{6.12}$$

We now make the critical assumption that the time-scale for turbulent fluctuation, $\tau_{\rm cor}$, is far smaller than the time scale, $\tau_{\rm avg}$, on which mean variables, such as $\bar{T}$, evolve. To be quantitative we assume sufficient separation of time scales that $\tau_{\rm cor}/\tau_{\rm avg} \sim \epsilon^2$ for some $\epsilon \ll 1$. Since $\bar{T}(t)$ varies on time-scale $\tau_{\rm avg}$ we assume it is relatively unchanged over $\epsilon\tau_{\rm avg}$ so

$$\bar{T}(s,\mathbf{x}) \;\simeq\; \bar{T}(t,\mathbf{x}) \quad , \quad |t-s| \;<\; \epsilon\tau_{\rm avg} \gg \tau_{\rm cor} \quad . \tag{6.13}$$

To take advantage of this time-scale separation at a given time t we choose an earlier time $\tau_x = t - \epsilon\tau_{\rm avg}$. This allows the (approximate) application of (6.13) over $\tau_x < s < t$ and (6.12) for $s < \tau_x$ to re-express the integral of eq. (6.8)

$$\begin{aligned}
\int^t \langle u_i(t)u_j(s)\rangle\,\partial_j\bar{T}(s)\,ds \;&=\; \int_{\tau_x}^t \Big\langle u_i(t)u_j(s)\Big\rangle\,\partial_j\bar{T}(s)\,ds + \overbrace{\int^{\tau_x}\Big\langle u_i(t)u_j(s)\Big\rangle\,\partial_j\bar{T}(s)\,ds}^{\simeq 0 \text{ by eq. (6.12)}} \\
&\simeq\; \int_{\tau_x}^t \Big\langle u_i(t)u_j(s)\Big\rangle\,\partial_j\bar{T}(t)\,ds \;\simeq\; \partial_j\bar{T}(t)\int_{-\infty}^t \Big\langle u_i(t)u_j(s)\Big\rangle\,ds \quad ,
\end{aligned}$$

where a spatial coordinate $\mathbf{x}$ is implicit in every factor. In the final integral the lower bound it taken to $-\infty$ since property (6.12) shows that integrand vanishes over the range $s < \tau_x$. The final integral can be defined as an effective thermal diffusivity or *turbulent conductivity*

$$\tilde{\kappa}_{ij}^{(\rm eff)}(\mathbf{x}) \;\equiv\; \int_{-\infty}^t \Big\langle u_i(\mathbf{x},t)u_j(\mathbf{x},s)\Big\rangle\,ds \;=\; \int_{-\infty}^0 V_{ij}(\mathbf{x},\Delta t)\,d\Delta t \quad . \tag{6.14}$$

This depends on position but, due to our assumption of stationary turbulence, it does not depend on time t.

In terms of this diffusivity the mean heat flux defined by eq. (6.8)

$$\Gamma_i \;=\; \langle u_i T' \rangle \;=\; -\,\tilde{\kappa}_{ij}^{(\mathrm{eff})}\,\partial_j \bar{T}(\mathbf{x},t) \quad . \tag{6.15}$$

is proportional to the gradient of mean temperature. While this seems natural it also reveals how a correlation arises in the fluctuating quantities u_j and T'. Flow down a temperature gradient ($\mathbf{u}\cdot\nabla\bar{T} < 0$) will generate a positive temperature fluctuation, $T' > 0$. This will be anti-correlated with velocity fluctuation which generated it, u_j, leading to a non-vanishing mean $\langle u_j T' \rangle$ directed down the temperature gradient. Using the effective thermal diffusion in eq. (6.3) gives a general heat equation for the evolution of the large-scale temperature $\bar{T}(\mathbf{x},t)$

$$\partial_t \bar{T} \;=\; \partial_i \left(\tilde{\kappa}_{ij}^{(\mathrm{eff})}\,\partial_j \bar{T} \right) \;=\; \nabla\cdot(\underline{\underline{\tilde{\kappa}}}^{(\mathrm{eff})}\cdot\nabla\bar{T}) \quad . \tag{6.16}$$

6B Relation between diffusion and random walks

Diffusion equations such as eq. (6.16) arise naturally from advection equations, such as eq. (6.5), when the velocity $\mathbf{u}(\mathbf{x},t)$ is random. To see this let us consider a case where the velocity statistics are not only stationary (unchanging in time) but are also homogeneous (uniform in space). In this case the diffusion tensor, $\underline{\underline{\tilde{\kappa}}}^{(\mathrm{eff})}$, will not depend on $\mathbf{x}$ and eq. (6.16) becomes

$$\partial_t \bar{T} \;=\; \tilde{\kappa}_{ij}^{(\mathrm{eff})}\,\partial_{ij}^2 \bar{T} \;=\; \tilde{\kappa}_{ij}^{(\mathrm{eff})} \frac{\partial^2 \bar{T}}{\partial x_i \partial x_j} \quad . \tag{6.17}$$

Since $\partial_{ij}^2 \bar{T}$ is a symmetric matrix, any anti-symmetric part of $\tilde{\kappa}_{ij}^{(\mathrm{eff})}$ will vanish upon contraction. Therefore, while eq. (6.14), does not necessarily produce a symmetric matrix, only its symmetric piece

$$\tilde{\kappa}_{ij}^{S} \;=\; \tfrac{1}{2}\tilde{\kappa}_{ij}^{(\mathrm{eff})} + \tfrac{1}{2}\tilde{\kappa}_{ji}^{(\mathrm{eff})} \tag{6.18}$$

will have any effect. We hereafter refer only to the symmetric piece of the diffusion coefficient and omit the superscript, S or (eff) for simplicity.

Equation (6.17) is a simple parabolic equation whose solution can be expressed using the Green's function $G(\mathbf{x},t)$

$$\bar{T}(\mathbf{x},t) \;=\; \int G(\mathbf{x}-\mathbf{x}',t)\,T(\mathbf{x}',0)\,d\mathbf{x}' \quad . \tag{6.19}$$

The Green's function satisfies the diffusion equation,

$$\partial_t G \;=\; \tilde{\kappa}_{ij}\,\partial_{ij}^2 G \quad , \tag{6.20}$$

with initial condition $G(\mathbf{x},0) = \delta(\mathbf{x})$. This initial condition means that eq. (6.17) will have the initial condition $\bar{T}(\mathbf{x},0) = T(\mathbf{x},0)$. The solution to eq. (6.20) is

$$G(\mathbf{x},t) \;=\; \frac{A}{t^{3/2}}\,\exp\left(\mathbf{x}\cdot\underline{\underline{\tilde{\kappa}}}^{-1}\cdot\mathbf{x}/4t \right) \tag{6.21}$$

where $A = (2\pi)^{-3/2}/\sqrt{\det(\underline{\underline{\tilde{\kappa}}})}$ is required for normalization of the δ-function.

6B.1 Random advection

While (6.21) is the Greens function for the diffusion equation, the simple advection equation, eq. (6.2), is solved by the entirely different Green's function

$$G_{\mathbf{u}}(\mathbf{x},t) \;=\; \delta\left[\,\mathbf{x}-\Delta\mathbf{x}(t)\,\right] \quad , \tag{6.22}$$

where

$$\Delta\mathbf{x}(t) \;=\; \int_0^t \mathbf{u}(s)\,ds \quad , \tag{6.23}$$

is the net displacement produced by the velocity field $\mathbf{u}$. Using this Green's function gives the solution for the full temperature field, including random fluctuations,

$$T(\mathbf{x},t) \;=\; \int G_{\mathbf{u}}(\mathbf{x}-\mathbf{x}',t)\,T(\mathbf{x}',0)\,d\mathbf{x}' \quad . \tag{6.24}$$

Taking the average of the equation we find an expression identical to (6.19) but with Green's function

$$G(\mathbf{x},t) \;=\; \langle G_{\mathbf{u}}(\mathbf{x},t)\rangle \quad , \tag{6.25}$$

where the average is over random velocity fields $\mathbf{u}(\mathbf{x},t)$. The random velocity fields produce random displacements, $\Delta\mathbf{x}$, distributed according to some probability distribution function $p(\Delta\mathbf{x})$. The average can be performed as an integral over this pdf,

$$\begin{aligned} G(\mathbf{x},t) \;&=\; \langle G_{\mathbf{u}}(\mathbf{x},t)\rangle \;\equiv\; \int p(\Delta\mathbf{x})G_{\mathbf{u}}(\mathbf{x},t)\,d(\Delta\mathbf{x}) \\ &=\; \int p(\Delta\mathbf{x})\delta\left[\,\mathbf{x}-\Delta\mathbf{x}\,\right]\,d(\Delta\mathbf{x}) \;=\; p(\mathbf{x}) \quad , \end{aligned} \tag{6.26}$$

after using eq. (6.22). While the derivation is somewhat formal the result is intuitive and simple. The contribution of initial point $\mathbf{x}'$ to the mean temperature at position $\mathbf{x}$ is proportional to the probability that the flow has advected the temperature between those points over the intervening time.

Comparing this to (6.21) shows that *a diffusion equation results from advection by a random velocity field whose displacements are distributed as a Gaussian.* Equation (6.21) is a general Gaussian, whose moments are not isotropic. The more familiar version,

$$p(\Delta\mathbf{x}) \;=\; \frac{1}{\sigma^3(2\pi)^{3/2}}\,\exp\left(-\frac{|\Delta\mathbf{x}|^2}{2\sigma^2}\right) \quad , \tag{6.27}$$

has standard deviation, σ, defined so that $\langle\Delta x^2\rangle = \langle\Delta y^2\rangle = \langle\Delta z^2\rangle = \sigma^2$. This means that $\langle|\Delta\mathbf{x}|^2\rangle = 3\sigma^2$. Comparison to (6.21) shows

$$\underline{\underline{\tilde{\kappa}}}^{-1} \;=\; \frac{2t}{\sigma^2}\underline{\underline{\mathrm{I}}} \quad , \tag{6.28}$$

Inverting this gives the thermal diffusion coefficient

$$\underline{\underline{\tilde{\kappa}}} \;=\; \frac{\sigma^2}{2t}\underline{\underline{\mathrm{I}}} \;=\; \frac{\langle|\Delta\mathbf{x}|^2\rangle}{6t}\underline{\underline{\mathrm{I}}} \quad . \tag{6.29}$$

Since the diffusion tensor is proportional to the identity tensor, $\underline{\underline{\mathrm{I}}}$, it is called *isotropic*. We see that this is a direct consequence of adopting the isotropic Gaussian, eq. (6.27), in which $\langle\Delta x^2\rangle = \langle\Delta y^2\rangle = \langle\Delta z^2\rangle$. In the most general form of Gaussian the exponent contains a quadratic form $\mathbf{x}\cdot\underline{\underline{\mathrm{C}}}\cdot\mathbf{x}$, as it does in eq. (6.21). In that case displacements in different directions are of different mean sizes.

6B.2 Random walks

The appearance of t on the right of eq. (6.29) suggests that the diffusion coefficient may be time-dependent. This would, however, contradict our assumption of stationary statistical properties. The displacement, $\Delta\mathbf{x}$, in the numerator increases with time according to eq. (6.23). If $\Delta\mathbf{x} \sim t$, as in simple constant-velocity motion, then we would expect $\tilde{\kappa} \sim t$. It turns out that random motion, such as a random walk or Brownian motion, has the property that the root-mean-squared displacement, $\langle|\Delta\mathbf{x}|^2\rangle^{1/2} \sim t^{1/2}$. After waiting four times as long the particle has only doubled its distance... on average. As a result the right hand side of (6.29) is actually a constant.

This fact can be demonstrated using expression (6.23) for $\Delta\mathbf{x}$, but only with very complex mathematics for dealing with random functions. The same idea can be illustrated more simply if we consider a random walk. In a random walk, the displacement occurs through discrete steps, $\delta\mathbf{x}_1$, $\delta\mathbf{x}_2$, etc. at fixed intervals, δt, rather than smoothly by a velocity function. The displacement after time $t = N\delta t$ is

$$\Delta\mathbf{x}(t) \;=\; \sum_{n=1}^{N} \delta\mathbf{x}_n \quad . \tag{6.30}$$

Steps are random, but each has the same root-mean-squared step-size $\delta\ell$,

$$\langle|\delta\mathbf{x}_1|^2\rangle \;=\; \langle|\delta\mathbf{x}_2|^2\rangle \;=\; \cdots \;=\; \langle|\delta\mathbf{x}_n|^2\rangle \;=\; \delta\ell^2 \quad . \tag{6.31}$$

Furthermore, each step is completely independent of all others. This fact is quantified through the relation

$$\langle\delta\mathbf{x}_m \cdot \delta\mathbf{x}_n\rangle \;=\; \delta\ell^2\,\delta_{mn} \quad . \tag{6.32}$$

Using eqs. (6.30)–(6.32) we find the mean squared displacement

$$\begin{aligned}\langle|\Delta\mathbf{x}|^2\rangle \;&=\; \left\langle\left[\sum_{m=1}^{N}\delta\mathbf{x}_m\right]\cdot\left[\sum_{n=1}^{N}\delta\mathbf{x}_n\right]\right\rangle \;=\; \sum_{m=1}^{N}\sum_{n=1}^{N}\langle\delta\mathbf{x}_m\cdot\delta\mathbf{x}_n\rangle \\ &=\; \sum_{m=1}^{N}\delta\ell^2 \;=\; N\delta\ell^2 \;=\; \frac{\delta\ell^2}{\delta t}\,t \quad .\end{aligned} \tag{6.33}$$

This shows that the rms displacement $\langle|\Delta\mathbf{x}|^2\rangle^{1/2} \sim t^{1/2}$.

Using expression (6.33) in eq. (6.29) gives the scalar diffusion coefficient

$$\tilde{\kappa} \;=\; \frac{1}{6}\,\frac{\delta\ell^2}{\delta t} \;=\; \frac{1}{6}\,v_{\rm rms}^2\,\delta t \quad , \tag{6.34}$$

where $v_{\rm rms} \equiv \delta\ell/\delta t$ is the speed of the random motion. This is related to the trace of our original expression, eq. (6.14),

$$\tilde{\kappa} \;\equiv\; \tfrac{1}{3}\mathrm{Tr}[\underline{\underline{\tilde{\kappa}}}] \;=\; \tfrac{1}{3}\int_0^\infty \langle\mathbf{u}(\mathbf{x},t)\cdot\mathbf{u}(\mathbf{x},t+\Delta t)\rangle\,d(\Delta t) \;=\; \tfrac{1}{3}\langle|\mathbf{u}|^2\rangle\tau_{\rm cor} \quad , \tag{6.35}$$

where $\tau_{\rm cor}$ is defined as the integral of the velocity correlation function

$$C(\Delta t) \;=\; \frac{\langle\mathbf{u}(\mathbf{x},t)\cdot\mathbf{u}(\mathbf{x},t+\Delta t)\rangle}{\langle|\mathbf{u}|^2\rangle} \quad . \tag{6.36}$$

We can compute the correlation function exactly for the case of a random walk with fixed steps of length $\delta t = t_{i+1} - t_i$. Define step i as the one including the first time, t: $t_i \le t \le t_{i+1}$. The time until step $i+1$ begins is $\tau = t_{i+1} - t$. If we assume that the initial time t_0 is chosen at random, then τ will be a random variable, uniformly distributed between 0 and δt. If its maximum value, $\delta t > \Delta t$, then a new step will *definitely* have begun before time $t + \Delta t$ — i.e. with probability $p = 1$. On the other hand, if $\delta t < \Delta t$ then there is a non-zero probability $p = \Delta t/\delta t$ that $\tau < \Delta t$ and that a new step will have begun. If that occurs then $\mathbf{u}(\mathbf{x}, t)$ and $\mathbf{u}(\mathbf{x}, t + \Delta t)$ will be completely uncorrelated. In the opposite case the steps are completely correlated and $\mathbf{u}(\mathbf{x}, t) \cdot \mathbf{u}(\mathbf{x}, t + \Delta t) = v_{\rm rms}^2$. This will occur with probability $1 - p$, for p described above. The correlation function will be equal to this probability,

$$C(\Delta t) \;=\; \mathrm{Prob}\Big[\,(t_{i+1} - t) \;>\; \Delta t\,\Big] \;=\; \begin{cases} 1 - |\Delta t|/\delta t & , \quad |\Delta t| < \delta t \\ 0 & , \quad \text{otherwise} \end{cases} \tag{6.37}$$

It is easy to the that

$$\tau_{\rm cor} \;=\; \int_0^\infty C(\Delta t)\, d(\Delta t) \;=\; \int_0^{\delta t} \left[1 - \frac{\Delta t}{\delta t}\right] d(\Delta t) \;=\; \tfrac{1}{2}\delta t \quad , \tag{6.38}$$

confirming that eqs. (6.34) and (6.35) are quantitatively consistent. We obtain the same turbulent diffusivity using a random walk as using statistics of the continuous velocity field $\mathbf{u}(\mathbf{x}, t)$.

The argument above has shown that a random walk produces displacement growing in accordance with the diffusion equation. To make the full connection between the random walk and diffusion, however, we should demonstrate that the random displacement defined by eq. (6.30), has a Gaussian distribution function. This can be done using a powerful result of probability theory known as the *central limit theorem*. It applies to a random variable, such as $\Delta\mathbf{x}$, defined as the sum of many other random variables, like $\delta\mathbf{x}_n$. If these component variables have identical distributions of almost any kind, and are sufficiently plentiful,[2] then the distribution of the sum will be approximately Gaussian — regardless of how the individual terms are distibuted. This is a very powerful result and explains, in large measure, the ubiquity of the Gaussian distribution in statistics.

In the present case, the displacement $\Delta\mathbf{x}$ is the sum of many random steps, or the integral of a random velocity over many correlation times. In either event this endows $\Delta\mathbf{x}$ with a Gaussian distribution function, $p(\Delta\mathbf{x})$, whose variance is the sum of the variances of its steps. According to eq. (6.26) this distribution function is also the Green's function for the evolution of the average temperature field, $\bar{T}(\mathbf{x}, t)$. A Gaussian is the Green's function for a diffusion equation, so the average temperature field must satisfy a diffusion equation. If the random-walk steps are not themselves isotropic (i.e. steps favor one direction) then the Gaussian distribution is not isotropic and the original diffusion equation involves an anisotropic diffusion coefficient $\underline{\underline{\tilde{\kappa}}}$.

The logic above recurs very often in many physical situation. Indeed, almost every case of diffusion you encounter has at its base a random-walk process of some kind. Here

[2] In practice $N \sim 10$ is typically sufficient.

we have shown that random advection by a turbulent fluid velocity leads to an effective diffusion of turbulence. The random flow will also move momentum randomly leading to a turbulent diffusion of velocity called *turbulent viscosity*. In a later section we will consider the turbulent diffusion of magnetic field which is called *turbulent resistivity*.

In addition to turbulent diffusion, we have seen that the fluid itself has underlying diffusion which lead to the classical thermal diffusion, $\tilde{\kappa}$. This arises from random motions of the individual particles composing the fluid, molecules in a gas or electrons in a plasma. The particles move at their own thermal speed, $v_{\rm th}$, making random steps whose size is called the *mean free path*, $\ell_{\rm mfp}$. Thus the classical thermal diffusion coefficient is given by

$$\tilde{\kappa} \;\simeq\; \frac{1}{6}\ell_{\rm mfp}\, v_{\rm th} \quad . \tag{6.39}$$

In an optically thick plasma heat can also be transported by photons traveling at the speed of light over distances called the *photon mean-free path*, leading to the radiative conductivity found in eq. (1.46). Each random walk produces diffusion with its own diffusion coefficient.

6C Turbulence in stratified atmospheres: mixing length theory

The foregoing presented a general view of turbulent diffusion, applicable to many turbulent flows. In one application of astrophysical interest, namely turbulent diffusion in stellar atmospheres, some modification is required. In that case the mean density, $\bar{\rho}$, varies due to gravitational stratification, over a length-scale similar to the turbulence itself. The turbulent velocity is still subsonic but due to the density stratification it is not incompressible; i.e. $\nabla\cdot\mathbf{u}\neq 0$. Subsonic flows in stratified atmospheres are assumed to be *anelastic*,

$$\nabla\cdot(\bar{\rho}\mathbf{u}) \;=\; 0 \quad , \tag{6.40}$$

rather than incompressible. This modifies the equation for fluctuating temperature to

$$\partial_t T' \;=\; -\,\mathbf{u}\cdot\nabla\bar{T} \;-\; (\gamma-1)\bar{T}\nabla\cdot\mathbf{u} \;=\; -\,\mathbf{u}\cdot\nabla\bar{T} \;+\; (\gamma-1)\frac{\bar{T}}{\bar{\rho}}\mathbf{u}\cdot\nabla\bar{\rho} \quad , \tag{6.41}$$

after dropping the term destined to be neglected in first-order smoothing.

Repeating the steps from eqs. (6.7) to (6.16), but including this new term, leads to a temperature flux

$$\mathbf{\Gamma} \;=\; \langle\mathbf{u}T'\rangle \;=\; \underline{\underline{\tilde{\kappa}}}^{(\rm eff)}\cdot\left[(\gamma-1)\frac{\bar{T}}{\bar{\rho}}\nabla\bar{\rho} \;-\; \nabla\bar{T}\right] \quad , \tag{6.42}$$

where $\underline{\underline{\tilde{\kappa}}}^{(\rm eff)}$ is the same turbulent diffusion tensor, given by eq. (6.14). Because we have assumed $\dot{Q}=0$ there is no genuine heat flux. There is, however, a mean enthalpy flux which behaves like a heat flux,

$$\begin{aligned}\mathbf{q}^{(\rm eff)} \;&=\; \frac{\gamma}{\gamma-1}\langle p'\mathbf{u}\rangle \;=\; \frac{\gamma}{\gamma-1}\frac{k_{\rm B}}{m}\bar{\rho}\,\langle T'\mathbf{u}\rangle \;=\; \frac{\gamma}{\gamma-1}\frac{k_{\rm B}}{m}\bar{\rho}\,\mathbf{\Gamma}\\ &=\; \frac{\gamma}{\gamma-1}\frac{k_{\rm B}}{m}\bar{\rho}\,\underline{\underline{\tilde{\kappa}}}^{(\rm eff)}\cdot\left[(\gamma-1)\frac{\bar{T}}{\bar{\rho}}\nabla\bar{\rho} \;-\; \nabla\bar{T}\right]\end{aligned} \tag{6.43}$$

after using the anelastic approximation, eq. (6.40) to set $\rho' = 0$. The factor multiplying the vector in square brackets matches our expectation of an effective thermal conductivity, except for a factor of γ due to its origin as an enthalpy flux. The vector it multiples is not, however, simply $-\nabla\bar{T}$ as it would be in genuine heat flux. Instead, it seems that stratification leads to heat transport *up* the density gradient as well as down the temperature gradient.

To understand this peculiar situation consider the hypothetical temperature gradient implied by the actual density gradient if the atmosphere were isentropic. In that case $\bar{T} \sim \bar{\rho}^{\gamma-1}$, so the temperature gradient would be

$$\bar{\rho}\,\nabla\bar{T})_{\rm ad} \;=\; (\gamma-1)\bar{T}\nabla\bar{\rho} \quad . \tag{6.44}$$

(The subscript "ad", for adiabatic, is conventional in spite of the fact that we are considering an *isentropic* atmosphere.) This means the effective heat flux

$$\mathbf{q}^{(\rm eff)} \;=\; \frac{\gamma}{\gamma-1}\frac{k_{\rm B}}{m}\bar{\rho}\,\underline{\underline{\tilde{\kappa}}}^{(\rm eff)}\cdot\left[\,\nabla\bar{T})_{\rm ad} \;-\; \nabla\bar{T}\,\right] \quad , \tag{6.45}$$

is proportional to the difference between the isentropic temperature gradient and the *actual* temperature gradient. If the atmosphere is in fact isentropic then the turbulence will not change that. Since the underlying dynamics does not include thermal diffusion it is adiabatic. Thus the turbulence mixes entropy as it mixes fluid parcels. If the entropy was already uniform this mixing will have no effect. If, on the other hand, there was an entropy gradient then the turbulence will diminish it by its mixing action.

The vector difference $\nabla\bar{T})_{\rm ad} - \nabla\bar{T}$ appearing in the turbulent transport is the same one that governs instability in a gravitationally stratified atmosphere discussed in Chapter 5. The dimensionless quantity ξ defined in eq. (5.19) can be used to express

$$\nabla\bar{T})_{\rm ad} \;-\; \nabla\bar{T} \;=\; \frac{m\mathbf{g}}{k_{\rm B}}\,\xi \quad , \tag{6.46}$$

where we have used the gravitational vector $\mathbf{g}$ to remove confusion about direction. Both the actual temperature gradient and the density gradient point downward, parallel to $\mathbf{g}$. Upward heat flux, i.e. against the temperature gradient, therefore requires that $|\nabla\bar{T}| > |\nabla\bar{T})_{\rm ad}|$, the temperature gradient must exceed the adiabatic value; this is often called a *super-adiabataic temperature gradient*. Equation (6.46) shows that this also requires that $\xi < 0$ which we found in Chapter 5 to be the condition for dynamical instability of the stratification. In particular the g-mode is not oscillatory but grows exponentially at a rate

$$\sigma \;=\; \sqrt{-N^2} \;\sim\; g\sqrt{\frac{m}{k_{\rm B}}}\,\bar{T}^{-1/2}\sqrt{-\xi} \;=\; \sqrt{\frac{g}{\bar{T}}\left|\,\nabla\bar{T})_{\rm ad} \;-\; \nabla\bar{T}\right|} \quad , \tag{6.47}$$

from eq. (5.35) after setting $\theta_k = \pi/2$ to obtain the fastest growth — the dominant modes. The final expression assumes a super-adiabatic temperature gradient, as we will henceforth.

The scenario for turbulent heat transport in stars is evident at this point. A super-adiabatic temperature gradient leads to fluid turbulence which leads to a turbulent heat flux (6.43), which then sets the temperature gradient. If turbulent transport were too effective it would flatten the temperature gradient *below* its super-adiabatic value, thereby

quenching the turbulence. Thus the turbulence is maintained at a level sufficient for the turbulent transport to keep the gradient at a level *just* above the adiabatic level. The *mixing length theory* is a quantitative, self-consistent model of this interplay between instability, fluid turbulence and heat transport. It permits the self-consistent determination of the turbulent velocity field which has, up to this point in our presentation, remained *ad hoc*.

Mixing length theory assumes that the g-mode instability generates the turbulence that leads to the turbulent diffusion $\tilde{\kappa}^{(\rm eff)}$. The exponential growth rate σ, in eq. (6.47), was derived using linear dynamics which assumed fluctuations remain small. We wish to use this to infer properties of the *non-linear* phase in which the exponential growth ends. It is a common, albeit heuristic, assumption that the non-linear fluctuations which result from any linear instability will have a correlation time comparable to the growth time of that instability. We will adopt this assumption here and set $\tau_{\rm cor} = 1/\sigma$ where the linear growth rate is given by eq. (6.47).

The size of the non-linear fluctuations, i.e. turbulent eddies, is assumed to be set by the pressure scale height[3]

$$H_p \;=\; \frac{\bar{p}}{\bar{\rho} g} \;=\; \frac{k_{\rm B}\bar{T}}{mg} \quad , \tag{6.48}$$

which is the natural length scale of the equilibrium. The turbulent velocity is $v_{\rm rms} = H_p\sigma$, so the (scalar) turbulent diffusion coefficient becomes

$$\tilde{\kappa}^{(\rm eff)} \;=\; v_{\rm rms}^2\tau_{\rm cor} \;=\; H_p^2\sigma \;=\; \frac{k_{\rm B}^2}{m^2}\left(\frac{\bar{T}}{g}\right)^{3/2}\sqrt{\left|\nabla\bar{T})_{\rm ad} \;-\; \nabla\bar{T}\right|} \quad . \tag{6.49}$$

The diffusion coefficient thus depends on the same gradient that it multiples in the flux. Using this coefficient in eq. (6.43) yields an explicit expression for the mixing length heat flux (Schwartzschild, 1958, §7 gives a more traditional derivation of the same expression)

$$\mathbf{q}^{(\rm mlt)} \;=\; \frac{\gamma}{\gamma-1}\bar{\rho}\frac{k_{\rm B}^3}{m^3}\left(\frac{\bar{T}}{g}\right)^{3/2}\left|\nabla\bar{T})_{\rm ad} \;-\; \nabla\bar{T}\right|^{3/2}\hat{\mathbf{r}} \quad , \tag{6.50}$$

directed outward, as it must be in a star (keeping in mind that this expression applies only where $|\nabla\bar{T}| > |\nabla\bar{T})_{\rm ad}|$). This is the mixing length heat flux, and it turns out to be proportional to the three-halves power of the temperature gradient relative to its adiabatic value. This odd dependence is a combination of the linear dependence explicit in eq (6.45) and another power of one-half in the dependence of the turbulent diffusion coefficient, $\tilde{\kappa}^{(\rm eff)}$.

While eq. (6.50) appears formidable its effect on the atmospheric structure is very simple: the turbulence keeps the actual temperature very close to the adiabatic one. To appreciate this we rework eq. (6.50) to provide the relative departure from an adiabatic temperature over a scale height

$$\frac{H_p}{\bar{T}}\left|\nabla\bar{T})_{\rm ad} \;-\; \nabla\bar{T}\right| \;=\; \frac{1}{gH_p}\left(\frac{\gamma-1}{\gamma}\frac{|\mathbf{q}|}{\bar{\rho}}\right)^{2/3} \quad . \tag{6.51}$$

[3]It is defined more presicesly as a fixed multiple of this. We prefer to stick with a less precise, but clearer account here.

The heat flux carried by the turbulence cannot exceed the total heat flux of the star,

$$|\mathbf{q}| \;=\; \frac{\mathcal{L}_\odot}{4\pi R_\odot^2} \;=\; 6.24\times 10^{10}\,\mathrm{erg/cm^2/s} \quad ,$$

for the case of the Sun. The pressure scale height, H_p, is the local rate the pressure is falling and must therefore be close to the distance from the surface, i.e. the depth $H_p \sim z$. Using the surface value of g gives an approximate value

$$\frac{H_p}{\bar{T}}\Big|\nabla\bar{T})_{\rm ad} \;-\; \nabla\bar{T}\Big| \;\sim\; 2.7\times 10^{-7}\left(\frac{z}{10^{10}\,\mathrm{cm}}\right)^{-1}\left(\frac{\bar{\rho}}{0.1\,\mathrm{g/cm^3}}\right)^{-2/3} \quad , \tag{6.52}$$

where the characteristic values diving z and $\bar{\rho}$ are those near the center of the Sun's convection zone. It therefore seems that the actual temperature gradient departs negligibly from adiabatic except very near the surface: say $\bar{\rho}\sim 10^{-6}\,\mathrm{g/cm^3}$ at $z\sim 10^8$ cm.

The mixing length theory thus leads to a very simple prescription for determining the temperature profile in a star. One steps outward from the core using equations for hydrostatic balance and nuclear energy generation as well as heat flux. At each radius r one knows all quantities including the temperature and luminosity $\mathcal{L}(r)$. Before taking the next step one assumes first that the flux is carried entirely by radiative conduction, from which the temperature gradient would be

$$\left|\frac{dT}{dr}\right|_{\rm rad} \;=\; \frac{|\mathbf{q}|}{\kappa_{\rm rad}} \;=\; \frac{\mathcal{L}(r)}{4\pi r^2\kappa_{\rm rad}} \quad , \tag{6.53}$$

where $\kappa_{\rm rad}$ is the radiative conductivity given by expression (1.46). The step is then made with a temperature gradient (always negative, of course) whose magnitude is

$$\left|\frac{dT}{dr}\right| \;=\; \min\left(\;\left|\frac{dT}{dr}\right|_{\rm rad},\;\left|\frac{dT}{dr}\right|_{\rm ad}\;\right) \quad . \tag{6.54}$$

This naturally makes the appropriate choice between radiative transfer, when it would produce a stable stratification, $|\nabla\bar{T}| < |\nabla\bar{T})_{\rm ad}|$, or turbulent transport when it would not. It tacitly assumes that convection will keep $|\nabla\bar{T})_{\rm ad} - \nabla\bar{T}|$ very small, in the sense of eq. (6.51), and uses the adiabatic gradient instead of solving eq. (6.50).

In a star with mass comparable to or less than that of the Sun the central region is stable and heat flux is radiative, following eq. (6.53); this is the *radiative zone*. As temperature decreases so does $\kappa_{\rm rad}$. This is due both to the explicit temperature dependence in eq. (1.46), and a decrease in the photon mean free path, $\ell_{{\rm mfp},\gamma}$, when more partially ionized scatterers are available. As $\kappa_{\rm rad}$ falls the radiative temperature gradient, eq. (6.53) steepens. At one point it exceeds the adiabatic gradient and convection takes over the heat flux; this is the *convection zone* or convective envelope.

Heavier stars, also called *upper main sequences stars*, are hotter and this transition never occurs. Those stars have no convective envelope. Instead there is convection in the region near the core. This is because their nuclear reactions are more vigorous and also more centrally concentrated.[4] This causes $\mathcal{L}(r)$ to rise more steeply and making eq. (6.53) large near the core. Upper main sequence stars have convective cores and radiative envelopes, instead of radiative cores and convective envelopes found in the lower main sequence.

[4] A majority of the energy in upper main sequence stars is from fusion reactions catalyzed by Carbon, Nitrogen and Oxygen (the so-called CNO-bi-cycle), instead of direct fusion between pairs of protons. The CNO-catalyzed fusion has a higher sensitivity to temperature.

Chapter 7

Shocks

We will motivate our study of shocks by considering the simple problem of a piston moving along a cylinder of fluid initially at rest. The piston begins its motion, $v_p\hat{\mathbf{x}}$, at time $t = 0$ creating a flow $\mathbf{u} = u_x(x,t)\hat{\mathbf{x}}$ ahead of it. The fluid next to the piston will move with it, $u_x = v_p$, while far ahead of it the fluid will still be at rest, as yet "unaware" the piston has begun moving. There will be a moving front at which information about the oncoming piston reaches the stationary fluid, causing it to begin moving. This is the *shock front*. The natural question is: "How fast does the shock front move?" Previous experience leads to expect this information to move at the speed of sound in the unperturbed fluid, c_{s0}. But what if the piston moves even faster than that?

7A Weak shocks: sound waves from a slow piston

Let us begin with the case we already understand: piston motion far below than the sound speed, $v_p \ll c_{s0}$. This is the case we can legitimately treat as a small perturbation to a static homogeneous equilibrium, and solve by linearization. To accommodate conventional shock notation in the future we denote the perturbation with a δ rather than a subscript,

$$p(x,t) \;=\; p_0 \;+\; \delta p(x,t) \;+\; \cdots \quad ,$$

and so forth. In this notation the two linearized equations which couple are

$$\partial_t\, \delta u_x \;=\; -\frac{1}{\rho_0}\,\partial_x\, \delta p \tag{7.1}$$

$$\partial_t\, \delta p \;=\; -\gamma p_0\,\partial_x\, \delta u_x \quad . \tag{7.2}$$

This is a set of telegraphers equations. The fluid is initially equilibrium so

$$\delta u_x(x,0) \;=\; 0 \quad , \quad \delta p(x,0) \;=\; 0 \quad .$$

The perturbation is caused by the piston moving to the right at speed v_p beginning at time $t = 0$. This introduces a boundary condition at position $x_p = v_p t$

$$\delta u_x(x_p,t) \;=\; v_p\,\Theta(t) \;=\; \begin{cases} 0 & , \quad t<0 \\ v_p & , \quad t>0 \end{cases} \tag{7.3}$$

where $\Theta(t)$ is the Heaviside function. The solutions to (7.1) and (7.2) subject to these conditions, are

$$\delta u_x(x,t) = v_p\,\Theta(c_{s0}t - x) \tag{7.4}$$

$$\delta p(x,t) = p_0\,\frac{\gamma v_p}{c_{s0}}\,\Theta(c_{s0}t - x) \quad , \tag{7.5}$$

within the region $x > 0$.

By a time $t > 0$ the piston has moved to position $x_p = v_p t$ and the fluid near it is moving at its same speed $v_p \ll c_{s0}$. This co-moving fluid composes a column extending to a shock front at $x_s = c_{s0}t$. The shock front moves ahead at the speed of sound, c_{s0}, much faster than the fluid in the column. The total length of the column

$$L_c \;=\; x_s \;-\; x_p \;=\; (c_{s0} - v_p)t \quad , \tag{7.6}$$

thus increases in time. The mass density within the column

$$\begin{aligned} \rho_c &= \rho_0 \;+\; \delta\rho \;+\; \cdots \;\simeq\; \rho_0\left(1 \;+\; \frac{\delta\rho}{\rho_0}\right) \\ &= \rho_0\left(1 \;+\; \frac{1}{\gamma}\frac{\delta p}{p_0}\right) \;=\; \rho_0\left(1 \;+\; \frac{v_p}{c_{s0}}\right) \quad , \end{aligned} \tag{7.7}$$

has been enhanced very slightly over its initial value — the smallness of the perturbation is required to justify our linear treatment. The mass within the enhanced column

$$\rho_c L_c \;=\; \rho_0\left(1 \;+\; \frac{v_p}{c_{s0}}\right)(c_{s0} - v_p)t \;=\; \rho_0\, c_{s0}\, t \;+\; \mathcal{O}\left(\frac{v_p^2}{c_{s0}^2}\right) \;\simeq\; \rho_0\, x_s \quad , \tag{7.8}$$

is the same as the mass of undisturbed fluid before the piston moved.

Analysis in a co-moving frame

We now take a new view of our solution by transforming to a reference frame moving rightward at the speed of the shock, $v_s = c_{s0}$. In this reference frame the undisturbed, or *pre-shock*, fluid appears to be moving leftward at

$$u_1 \;=\; 0 \;-\; v_s \;=\; -\,c_{s0} \quad , \tag{7.9}$$

(see fig. 7.1). The flow slows down at the shock to

$$u_2 \;=\; \delta u_x \;-\; v_s \;=\; -\,c_{s0}\left(1 \;-\; \frac{v_p}{c_{s0}}\right) \quad . \tag{7.10}$$

The flow is slowed down by the enhanced pressure on the downstream side, $p_2 > p_1$, which goes along with the enhanced density.[1] The mass flux into the shock, from the right, is

$$\Gamma_\rho \;=\; \rho_1 u_1 \;=\; -\,\rho_0 c_{s0} \quad . \tag{7.11}$$

[1] This appears to raise a chicken-and-egg issue. The initial problem of a piston resolves this nicely.

The flow velocity downstream of the shock, u_2, is slower but the mass density $\rho_2 = \rho_c$ is larger. The total flux

$$\Gamma_\rho \;=\; \rho_2\, u_2 \;=\; -\,\rho_0 \left(1 \;+\; \frac{v_p}{c_{s0}}\right)(c_{s0} - v_p) \;\simeq\; -\,\rho_0 c_{s0} \quad , \tag{7.12}$$

matches that from upstream. This is not a surprising result since mass must be conserved as it crosses the shock. It does, however, illustrate how reduced speed must be accompanied by enhanced density in this co-moving view.

7B Shocks: response to a fast piston

The linear treatment above required the piston to move much sower than the sound speed. Information about the approaching piston traveled at the sound speed, compressing and accelerating the fluid in advance. What would happen if the piston actually moved *faster* than sound: $v_p > c_{s0}$? How would the fluid ahead of it learn it was coming? How much would the fluid be compressed by this supersonic piston?

The answer must clearly be that the piston will compress the fluid by more than a small enhancement. This will require a solution of the full nonlinear equations. The previous linear treatment gives us insight into what the solution will look like (see fig. 7.1). Enhancement to the density and pressure will occur at a shock front moving at some speed, v_s. Ahead of this will be the uniform, undisturbed fluid — the pre-shock region we call region 1. Behind the shock front will be a second uniform region, called region 2, with density, ρ_2, pressure p_2 and velocity v_p to match the piston. Between these uniform regions will be some kind of transition between upstream and downstream states.

The classic solution to the problem is found in a reference frame moving with this shock, at v_s. In this co-moving reference frame we expect the solution to be steady: $\partial/\partial t = 0$ for all quantities. There will be inflow from the right at $u_1 = -v_s < 0$; this is how the stationary fluid appears in the co-moving frame. This inflow will have the properties of the stationary fluid, ρ_1 and p_1. Downstream of the shock will be a second uniform region with density $\rho_2 > \rho_1$, pressure $p_2 > p_1$ and velocity u_2 which has smaller magnitude than u_1 but still negative. Moving with the piston the flow will have velocity $u_2 = v_p - v_s < 0$. Thus the piston problem has boundary conditions ρ_1, p_1, from the stationary fluid, and $u_2 - u_1 = v_p$ form the piston. It is easier to consider the problem with all upstream values, ρ_1, u_1, and p_1, fixed and to solve for the downstream values, ρ_2, u_2 and p_2. We will follow this approach and return in the end to show how $v_p = u_2 - u_1$ can be used to find the correct solution for a given piston.

7B.1 Equations for the shock

We solve the problem using the standard fluid equations, assuming a steady solution, $\partial_t = 0$. This means that whatever shape the transition takes, it appears steady in the co-moving frame. The continuity and momentum equations in conservative form, (C) and (M-c) become

$$\partial_t \rho \;=\; -\partial_x(\rho u) \;=\; 0 \tag{7.13}$$

$$\partial_t(\rho u) \;=\; -\partial_x(\rho u^2) \;-\; \partial_x p \;+\; \partial_x(\tfrac{4}{3}\mu\partial_x u) \;=\; 0 \quad . \tag{7.14}$$

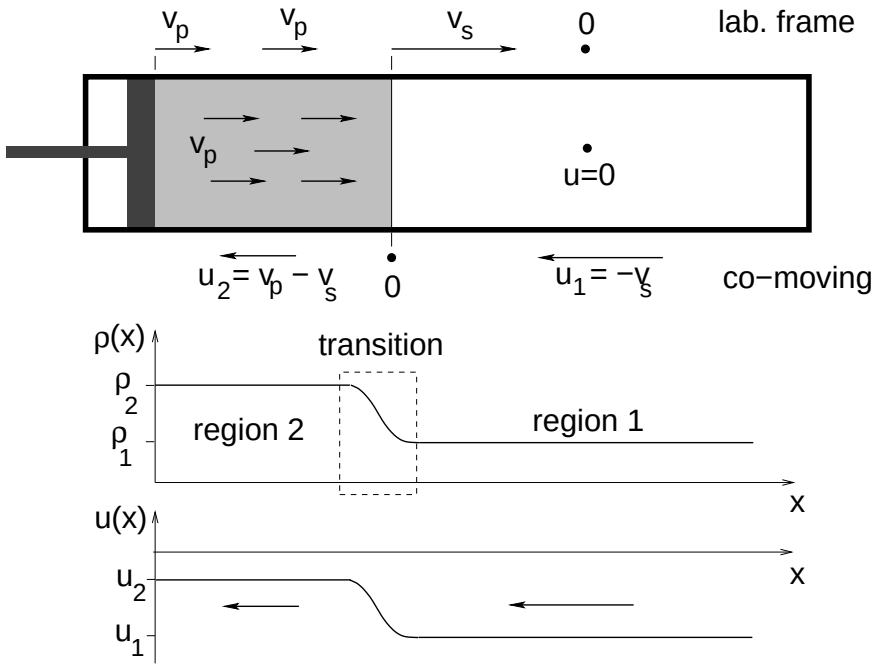

Figure 7.1: Geometry of the piston and shock problems. The piston is shown on top with the column of compressed gas in grey. Arrows above and below the piston show velocities in the lab reference frame (above) and in the frame co-moving with the shock front (below). Axes below plot the solutions $\rho(x)$ and $u(x)$ to the non-linear shock problem posed in the co-moving frame. The solution is composed uniform regions, 1 and 2, separated by a transition region (dashed box).

Viscosity will turn out to be important to a self-consistent treatment and we have used the most accurate form, eqs. (1.35) and (1.36), restricted to a single direction

$$f_x^{(\mathrm{visc})} \;=\; \partial_x \sigma_{xx} \;=\; \partial_x \left[\mu \left(\partial_x u \;+\; \partial_x u \;-\; \tfrac{2}{3} \nabla \cdot \mathbf{u} \right) \right] \;=\; \partial_x (\tfrac{4}{3} \mu \partial_x u) \quad . \tag{7.15}$$

The two equations thus dictate that the fluxes of mass and momentum are constant in the co-moving frame

$$\Gamma_\rho \;=\; \rho\, u \tag{7.16}$$

$$\Gamma_p \;=\; \rho\, u^2 \;+\; p \;-\; \tfrac{4}{3} \mu \partial_x u \quad . \tag{7.17}$$

A third relation must come from the conservation of energy. We use the conservative form of thermal energy conservation, eq. (E-c),

$$\partial_t \left(\frac{p}{\gamma - 1} \right) \;=\; -\, \partial_x \left(u\, \frac{p}{\gamma - 1} \right) \;-\; p\, \partial_x u \;+\; \tfrac{4}{3} \mu \, (\partial_x u)^2 \;=\; 0 \tag{7.18}$$

including viscous heating from eq. (1.47),[2] to recover the kinetic energy sapped by viscosity; we omit thermal conduction for simplicity. The right hand side is not a simple x derivative

[2]The factor $\frac{4}{3}$ is not trivial to arrive at. It is related to the matrix norm $|| \, 2\hat{\mathbf{x}}\hat{\mathbf{x}} - \frac{2}{3}\underline{\underline{\mathbf{I}}} \, ||^2 = \frac{8}{3}$.

of a conserved flux, so it is not the kind of conservation law we need. This is because thermal energy alone is not conserved. To obtain a conservation law for *total* energy we must add the evolution of kinetic energy, found from eq. (7.14),

$$\partial_t(\tfrac{1}{2}\rho u^2) \;=\; -\,\partial_x(\tfrac{1}{2}\rho u^3) \;-\; u\,\partial_x p \;+\; u\,\partial_x(\tfrac{4}{3}\mu\partial_x u) \;=\; 0 \quad . \tag{7.19}$$

Adding this to eq. (7.18) yields an evolution for total energy density

$$\partial_t\left(\frac{p}{\gamma-1}+\tfrac{1}{2}\rho u^2\right) \;=\; -\,\partial_x\left(\frac{\gamma}{\gamma-1}up\right) - \partial_x(\tfrac{1}{2}\rho u^3) + \partial_x(\tfrac{4}{3}\mu\, u\,\partial_x u) \;=\; 0 \quad . \tag{7.20}$$

The right hand side of this is a perfect derivative of the energy flux which is conserved

$$\Gamma_e \;=\; \tfrac{1}{2}\rho u^3 \;+\; \frac{\gamma}{\gamma-1}up \;-\; \tfrac{4}{3}\mu\, u\,\partial_x u \quad . \tag{7.21}$$

The second term is recognizable as enthalpy flux, the first is a flux of kinetic energy and the third is an energy flux due to work by viscous stress.

7C Rankine-Hugoniot relations

The stationary solution we seek consists of a thin transition region over which upstream quantities ρ_1, u_1 and p_1 change to downstream quantities ρ_2, u_2 and p_2. Outside the region all quantities are uniform and the derivative term $\partial_x u$ will vanish. That means that the upstream and downstream values must each be consistent with the conserved fluxes

$$\Gamma_\rho \;=\; \rho_1\, u_1 \;=\; \rho_2\, u_2 \tag{7.22}$$

$$\Gamma_p \;=\; \rho_1\, u_1^2 + p_1 \;=\; \rho_2\, u_2^2 + p_2 \tag{7.23}$$

$$\Gamma_e \;=\; \tfrac{1}{2}\rho_1 u_1^3 \;+\; \frac{\gamma}{\gamma-1}u_1 p_1 \;=\; \tfrac{1}{2}\rho_2 u_2^3 \;+\; \frac{\gamma}{\gamma-1}u_2 p_2 \quad . \tag{7.24}$$

These three conservation laws, conservation of mass, momentum and energy between upstream and downstream flows, are known as the *Rankine-Hugoniot* relations. They can be used to derive downstream quantities ρ_2, u_2 and p_2 from a given set of upstream values, ρ_1, u_1 and p_1.

To facilitate this derivation we combine the fluxes into a single non-dimensional constant

$$\frac{\Gamma_p^2}{2\Gamma_e\Gamma_\rho} \;=\; \frac{(\rho u^2+p)^2}{\rho u[\rho u^3 + 2\gamma p u/(\gamma-1)]} \quad , \tag{7.25}$$

which must be the same on the two sides. We replace $u = M\,c_s$, where M is the local Mach number and c_s is the local sound speed, and rework the expression into a form

$$\frac{\Gamma_p^2}{2\Gamma_e\Gamma_\rho} \;=\; \frac{(\gamma-1)(\gamma M^2+1)^2}{\gamma^2 M^2[(\gamma-1)M^2+2]} \;=\; F(M) \quad , \tag{7.26}$$

involving the local Mach number alone. The Mach number, $M_1^2 = \rho_1 u_1^2/\gamma p_1$, is the only dimensionless combination of state variables, so it is naturally the only one that appears in the dimensionless ratio of fluxes. The Mach number of the upstream flow is thus the single

parameter characterizing the shock; from it all other shock properties follow. It is common to characterize the shock by Mach number M_1, and to call this the *Mach number of the shock.*

Since eq. (7.26) must be the same on the two sides we can invert the function $F(M)$ to find the downstream Mach number

$$M_2 \;=\; F^{-1}\Big[\, F(M_1)\,\Big] \quad , \tag{7.27}$$

strictly in term of the upstream Mach number. The remaining downstream quantities can then be found from upstream quantities and M_2.

The plot in fig. 7.2 reveals that F^{-1} has two branches separated by the minimum of $F(M)$ at $M = 1$. Choosing the same branch for F^{-1} as for $F(M_1)$ would yield the trivial result of no change across the shock: $M_2 = M_1$. There would be no change in any quantity, $\rho_2 = \rho_1$, $u_2 = u_1$ and $p_2 = p_1$, which is clearly consistent with the conservation laws but is not a shock — it is not a transition. The alternative is for M_1 and M_2 to be on opposite sides of the minimum, $M = 1$, as shown in the figure. Thus a genuine shock transition must connect *supersonic* flow to *subsonic* flow. We will see below that the upstream side must be the supersonic flow, $M_1 > 1$, and the downstream side must be subsonic, $M_2 < 1$.

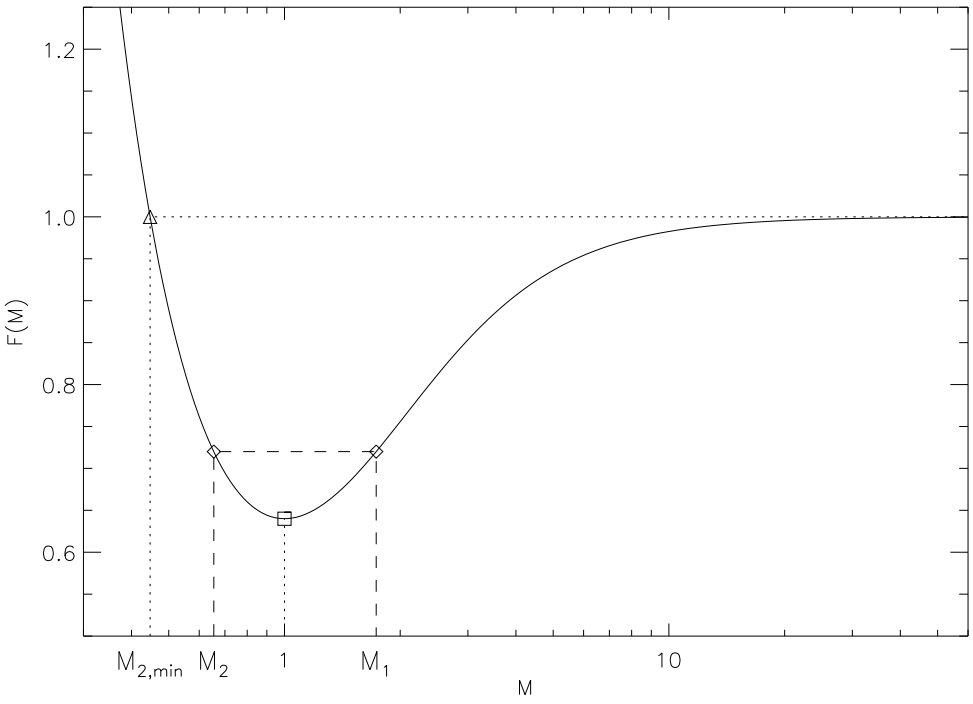

Figure 7.2: A plot of the function $F(M)$ from eq. (7.26) for $\gamma = 5/3$.. A square marks to minimum at $M = 1$. Diamonds show an example of values of M_2 and M_1 satisfying the relation $F(M_2) = F(M_1)$. A horizontal dotted line shows the asymptote, $F(M) \to 1$ as $M \to \infty$. This intersects the left branch of the curve at a point $M_{2,\rm min} = 1/\sqrt{5}$ (triangle), from eq. (7.36), so M_2 will be restricted to the range $M_{2,\rm min} < M_2 \leq 1$.

The inversion can be performed analytically by equating the value of the functions on opposite sides

$$F(M_1) \;=\; \frac{(\gamma-1)(\gamma M_1^2+1)^2}{\gamma^2 M_1^2[(\gamma-1)M_1^2+2]} \;=\; \frac{(\gamma-1)(\gamma M_2^2+1)^2}{\gamma^2 M_2^2[(\gamma-1)M_2^2+2]} \;=\; F(M_2) \quad . \tag{7.28}$$

Cross multiplying by denominators yields

$$\begin{aligned} & 2\gamma(\gamma-1)M_2^4M_1^2 + (\gamma-1)M_2^4 + 2\gamma^2 M_1^4M_2^2 + 2M_2^2 \\ = \;& 2\gamma(\gamma-1)M_1^4M_2^2 + (\gamma-1)M_1^4 + 2\gamma^2 M_2^4M_1^2 + 2M_1^2 \quad , \end{aligned} \tag{7.29}$$

which is a polynomial that factors into

$$\begin{aligned} -2\gamma\Big[M_2^4M_1^4 \;-\; M_1^4M_2^2\Big] \;+\; (\gamma-1)(M_2^4-M_1^4) \;+\; 2(M_2^2-M_1^2) \;&= \\ (M_2^2-M_1^2)\Big[-2\gamma M_2^2M_1^2 \;+\; (\gamma-1)(M_2^2+M_1^2) \;+\; 2\Big] \;&=\; 0 \quad . \end{aligned} \tag{7.30}$$

The leading factor returns the trivial "non-transition", $M_1^2 = M_2^2$. The second leads to the relation

$$M_2^2 \;=\; \frac{(\gamma-1)M_1^2+2}{2\gamma M_1^2-\gamma+1} \quad , \tag{7.31}$$

explicitly relating the downstream Mach number to the upstream Mach number. In agreement with the discussion above we see that in the limit $M_1^2 \to 1$ the down stream Mach number $M_2^2 \to 1$ as well — the two branches meet at the minimum. This is the limit of a weak shock, equivilent to sound wave. We treated the case in the first section finding that in a co-moving frame $u_1 \simeq u_2 \simeq c_{s0}$.

The remainder of the downstream properties can be derived using eq. (7.31) in conjunction with the the original conservation laws. Equation (7.23) dictates that

$$1 \;=\; \frac{\rho_1 u_1^2 + p_1}{\rho_2 u_2^2 + p_2} \;=\; \frac{p_1(\gamma M_1^2+1)}{p_2(\gamma M_2^2+1)} \quad . \tag{7.32}$$

This provides an explicit expression from the pressure ratio

$$\frac{p_2}{p_1} \;=\; \frac{\gamma M_1^2+1}{\gamma M_2^2+1} \;=\; \frac{2\gamma M_1^2-\gamma+1}{\gamma+1} \tag{7.33}$$

where the second expression comes from using (7.31) to replace M_2^2. We can then use the ratio

$$\frac{M_1^2}{M_2^2} \;=\; \frac{u_1^2(\rho_1/p_1)}{u_2^2(\rho_2/p_2)} \;=\; \frac{\rho_1 u_1}{\rho_2 u_2} \times \frac{u_1 p_2}{u_2 p_1} \;=\; \frac{u_1 p_2}{u_2 p_1} \quad . \tag{7.34}$$

This leads to the ratio

$$\frac{u_1}{u_2} \;=\; \frac{\rho_2}{\rho_1} \;=\; \frac{M_1^2}{M_2^2}\frac{p_1}{p_2} \;=\; \frac{(\gamma+1)M_1^2}{(\gamma-1)M_1^2+2} \quad , \tag{7.35}$$

after using eqs. (7.31) and (7.33). These ratios are plotted in fig. 7.3.

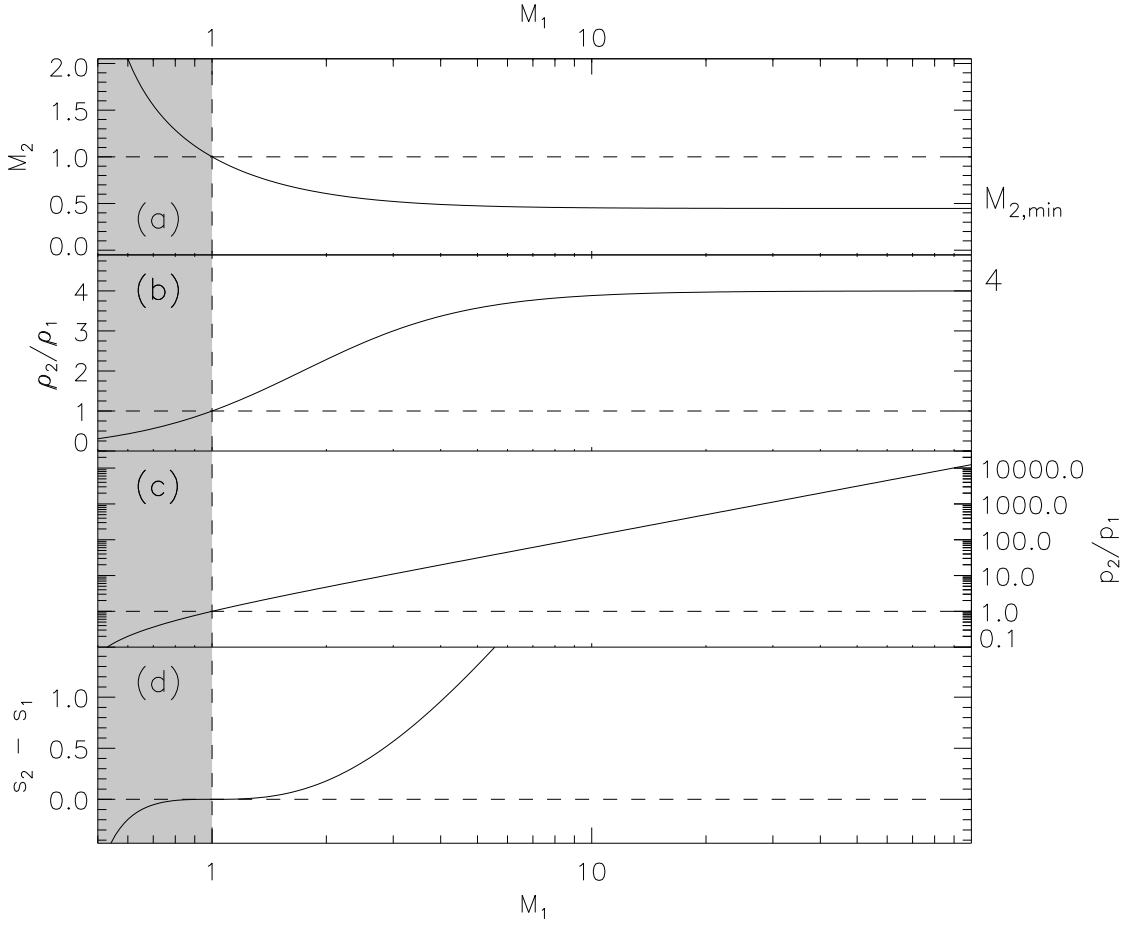

Figure 7.3: Plots of downstream shock properties as a function of upstream Mach number M_1 for $\gamma = 5/3$. (a) The downstream Mach number M_2 given by eq. (7.31), (b) the density ratio ρ_2/ρ_1 given by eq. (7.35), (c) the pressure ratio p_2/p_1 given by eq. (7.33) on a logarithmic axis, and (d) the difference in entropy, $s_2 - s_1$ given by (7.39). The vertical dashed line shows $M_1 = 1$, and to its left is the unphysical region with $M_2 > M_1$ over which the entropy change is negative.

Expressions (7.31), (7.33) and (7.35) have noteworthy behavior in the limit of very strong shocks, $M_1 \to \infty$. These are

$$M_2 \;\to\; \sqrt{(\gamma-1)/2\gamma} \;=\; M_{2,\min} \tag{7.36}$$

$$\frac{p_2}{p_1} \;\to\; \frac{2\gamma}{\gamma+1}\,M_1^2 \tag{7.37}$$

$$\frac{\rho_2}{\rho_1} \;\to\; \frac{\gamma+1}{\gamma-1}\quad . \tag{7.38}$$

Least surprising is that the post-shock pressure can be arbitrarily large as the Mach number is taken to be large. The post-shock density, on the other hand, cannot exceed a fixed multiple of the initial density. For a plasma ($\gamma = 5/3$) the maximum density increase is a factor of four. This also means the flow cannot slow down by more a fixed fraction of its initial value, given by the reciprocal of the density enhancement. Finally, eq. (7.36) gives the value of $M_{2,\min}$, previously noted in fig. 7.2; for the case $\gamma = 5/3$, $M_{2,\min} = 1/\sqrt{5}$.

7D Entropy and the need for dissipation

The specific entropy changes as the shock is crossed

$$\begin{aligned}\frac{(\gamma-1)}{k_{\rm B}/m}(s_2-s_1) \;&=\; \ln\left(\frac{p_2}{p_1}\right) \;-\; \gamma\ln\left(\frac{\rho_2}{\rho_1}\right)\\ &=\; \ln\left[1+\frac{2\gamma}{\gamma+1}(M_1^2-1)\right]+\gamma\ln\left[1+\frac{\gamma-1}{\gamma+1}(M_1^2-1)\right]-\gamma\ln(M_1^2)\quad ,\end{aligned} \tag{7.39}$$

plotted in fig. 7.3d. The difference is an increasing function of M_1 which crosses $s_2 - s_1 = 0$ at $M_1 = 1$. At this point the curve is very flat,

$$\frac{s_2-s_1}{k_{\rm B}/m} \;\simeq\; \frac{2\gamma}{3(\gamma+1)^2}\,(M_1^2-1)^3 \;-\; \frac{2\gamma^2}{(\gamma+1)^3}\,(M_1^2-1)^4 \;+\; \cdots\quad ,$$

but is negative for any value of $M_1 < 1$.

This shows that except in the infinitesimal neighborhood of $M_1 = 1$ the simultaneous conservation of mass, momentum and energy (i.e. eqs. [7.22]—[7.24]) demand there be some heating: $\dot{Q} \neq 0$. While those equations make no *explicit* reference to any form of heating, they cannot be solved without changing the entropy, which requires that some heating be present. Curiously the form of heating has no effect on the final result, although it must be consistent with energy conservation. For example, viscous heating enters the energy eq. (7.20) as a perfect derivative, so it contributes to the energy flux Γ_e but does not add or remove energy. Viscosity converts kinetic energy into thermal energy, but the total energy of the fluid is conserved by this conversion. Thermal conduction would work similarly, adding another term to the conserved energy flux Γ_e. Radiation, on the other hand, is a loss term so its inclusion would make an energy conservation law like (7.21) impossible. Radiative shocks must be analyzed using mass and momentum conservation alone.

The absence of viscosity from the results, but its necessity in the derivation, raises the question of what would happen if the fast piston experiment were posed for a perfectly ideal system, $\mu = 0$, and $\dot{Q} = 0$. Clearly such a system would conserve mass, momentum and

energy so eqs. (7.22)—(7.24) would still obtain. These constitute three equations for three unknowns, ρ_2, u_2, and p_2, so adding a fourth constraint,

$$\frac{p_1}{\rho_1^\gamma} \;=\; \frac{p_2}{\rho_2^\gamma} \quad ,$$

would over-determine the system making solution impossible in general. The resolution of the paradox is that the initial assumption, that a solution exists which is stationary ($\partial/\partial t = 0$) in a co-moving frame, becomes untenable. Instead the region downstream of the disturbance front will fail to settle into a steady, uniform structure — it will include time variations in all variables, even when viewed in a co-moving frame. Dissipative terms, such as viscosity, which convert kinetic energy to heat, turn out to be important for relaxing this motion.

The behavior of the entropy difference, $s_2 - s_1$, plotted in fig. 7.3d, provides the crucial evidence that fluid must traverse a shock from *supersonic to subsonic*: $M_1 > M_2$. While the alternative scenario conserves mass, momentum and energy (these laws are time-reversible), it would require a net *decrease* in the entropy of the system, in violation of the second law of thermodynamics. It is possible for specific entropy to decrease locally, leading to a profile $s(x)$, with positive and negative slope.[3] This steady profile does not change in time so the change in the entropy of the system occurs because the mass in region 2, with specific entropy s_2, is increasing while the mass in region 1, with specific entropy s_1, is decreasing at the same rate. The rate of this transfer is given by the mass flux $|\Gamma_\rho|$ across a shock of area A. This will change the entropy of the system at a rate

$$\frac{dS}{dt} \;=\; A|\Gamma_\rho|\,(s_2 - s_1) \quad , \tag{7.40}$$

which can never be negative according to the second law of thermodynamics. Thus we conclude that the solution where flow speeds up as it crosses the shock, the solutions where the fluid experiences a pressure decrease, are unphysical. Such pressure decreases do occur and are called *rarefaction waves* but they do not satisfy the assumptions of a shock: a transition which appears stationary in a co-moving frame.

7E Return to the piston

The derivation above followed the classic approach in a reference frame co-moving with the shock front. It was motivated by the problem of a piston moving at speed v_p into a stationary medium with sound speed c_{s0}. This drives a shock ahead at speed v_s. The upstream medium, moving at velocity $u_1 < 0$, was actually the stationary medium viewed from the reference frame of the shock. We therefore have $v_s = -u_1$. The downstream medium, moving at u_2 in the shock frame, is actually moving with the piston at

$$v_p \;=\; v_s + u_2 \;=\; u_2 - u_1 \;=\; |u_1 - u_2| \quad . \tag{7.41}$$

[3]Such an entropy decrease does occur inside shocks when thermal conduction is included.

We can divide the piston's speed by the sound speed of the ambient medium to derive a piston Mach number[4]

$$M_p \;=\; \frac{v_p}{c_{s0}} \;=\; \frac{|u_1 - u_2|}{c_{s1}} \;=\; \left(1 - \frac{u_2}{u_1}\right)|M_1| \;=\; \frac{2}{(\gamma+1)}\left(|M_1| - \frac{1}{|M_1|}\right) \quad . \tag{7.42}$$

This expression may be inverted to give the Mach number of the shock from that of the piston

$$|M_1| \;=\; M_s \;=\; \tfrac{1}{4}(\gamma+1)M_p \;+\; \sqrt{\tfrac{1}{16}(\gamma+1)^2 M_p^2 + 1} \quad . \tag{7.43}$$

It is evident from expression (7.43) that a supersonic shock, $M_s > 1$, is created by any piston motion, no matter how slow. This contrasts with our expectation from linear theory that information will propagate at c_{s0} and never faster. The fluid is compressed by the shock, which drives up its temperature and thus its sound speed. The shock front propagates at a speed which is at once *slower* than the sound speed in the downstream medium ($M_2 < 1$) and *faster* than the sound speed in the upstream (ambient) medium.

Extremely subsonic motion, $M_p \ll 1$, generates a weak shock traveling near the sound speed

$$M_s \;=\; \frac{v_s}{c_{s0}} \;\simeq\; 1 \;+\; \tfrac{1}{4}(\gamma+1)M_p \;+\; \cdots \quad , \tag{7.44}$$

as we found in the first section. Placing this into expression (7.33) yields a post-shock pressure

$$\frac{p_2}{p_1} \;\simeq\; 1 \;+\; \gamma\, M_p \;=\; 1 \;+\; \frac{\gamma v_p}{c_{s0}} \quad , \quad v_p \;\ll\; cs0 \quad , \tag{7.45}$$

matching eq. (7.5) form the linear treatment.

An extremely supersonic piston will generate a shock moving at

$$M_s \;\simeq\; \tfrac{1}{2}(\gamma+1)\, M_p \quad , \quad v_p \;\gg\; c_{s0} \quad . \tag{7.46}$$

The shock, in this limit, moves faster than the piston by a fixed factor

$$v_s \;=\; \frac{\gamma+1}{2} v_p \;>\; v_p \quad , \tag{7.47}$$

which is 4/3 for a plasma ($\gamma = 5/3$). This excess speed is needed accommodate a column of compressed fluid growing at a rate

$$\dot{L}_c \;=\; v_s - v_p \;=\; \frac{\gamma-1}{2} v_p \;=\; \frac{\gamma-1}{\gamma+1} v_s \quad , \quad v_p \;\gg\; c_{s0} \quad . \tag{7.48}$$

The front is adding new mass to the growing column with density ρ_1. It adds mass at speed v_s, so the ratio $\dot{L}_c / v_s$ is the inverse of the limiting compression ratio $\rho_2/\rho_1 = (\gamma+1)/(\gamma-1)$, from eq. (7.38), as required by mass conservation.

[4]In previous expression, M_1 and M_2, velocities were divided by their local sound speed. Here we divide the speed of the piston by the sound speed in the *ambient* medium rather than the medium adjacent to the piston.

The piston head, with area A, moves at speed v_p against a force $p_2 A$. It thus does work on the fluid at a rate $\dot{E} = v_p p_2 A$. The column of fluid on which it works grows at a rate $\dot{M} = \rho_1 v_s A$. This endows the material with a specific energy

$$\varepsilon \;=\; \frac{\dot{E}}{\dot{M}} \;=\; \frac{v_p p_2}{v_s \rho_1} \;=\; \frac{v_p}{v_s}\,\frac{\rho_2}{\rho_1}\frac{p_2}{\rho_2} \;\simeq\; \frac{2}{\gamma+1}\left(\frac{\gamma+1}{\gamma-1}\right)\frac{p_2}{\rho_2} \;=\; \frac{2}{\gamma-1}\,\frac{p_2}{\rho_2} \quad . \tag{7.49}$$

Half of this work is converted into thermal energy (per unit mass) of shocked material

$$\varepsilon_T \;=\; \frac{1}{\gamma-1}\,\frac{p_2}{\rho_2} \;=\; \tfrac{1}{2}\varepsilon \quad . \tag{7.50}$$

The other half is converted into bulk kinetic energy, per unit mass,

$$\begin{aligned} \tfrac{1}{2}v_p^2 \;&=\; \frac{1}{2}M_p^2\,\gamma\frac{p_1}{\rho_1} \;=\; \frac{\gamma}{2}\,M_p^2\,\frac{p_1}{p_2}\,\frac{\rho_2}{\rho_1}\,\frac{p_2}{\rho_2} \\ &\simeq\; \frac{\gamma}{2}\left(\frac{2}{\gamma+1}|M_1|\right)^2\left(\frac{\gamma+1}{2\gamma M_1^2}\right)\left(\frac{\gamma+1}{\gamma-1}\right)\frac{p_2}{\rho_2} \;=\; \frac{1}{\gamma-1}\,\frac{p_2}{\rho_2} \;=\; \tfrac{1}{2}\varepsilon \quad , \end{aligned} \tag{7.51}$$

needed for the fluid to move with the piston. Remarkably, the work done by the piston is shared equally between bulk kinetic energy and internal thermal energy of the fluid.

7F Inner structure of the shock

The shock is not a discontinuous jump between states, but a transition whose shape can be found using the constancy of the fluxes including the viscous stress and heating. Equation (7.17) can be used to replace local pressure profile

$$p(x) \;=\; \Gamma_p \;-\; \Gamma_\rho u \;+\; \tfrac{4}{3}\mu\partial_x u \quad . \tag{7.52}$$

Using this in eq. (7.21) gives

$$\begin{aligned} (\gamma-1)\Gamma_e \;&=\; \tfrac{1}{2}(\gamma-1)\Gamma_\rho u^2 \;+\; \gamma\, p\, u \;-\; \tfrac{4}{3}(\gamma-1)\mu\, u\,\partial_x u \\ &=\; -\tfrac{1}{2}(\gamma+1)\Gamma_\rho u^2 \;+\; \gamma\,\Gamma_p\, u \;+\; \tfrac{4}{3}\mu\, u\,\partial_x u \\ &=\; -\tfrac{1}{2}(\gamma+1)\Gamma_\rho u_1^2 \;+\; \gamma\,\Gamma_p\, u_1 \quad , \end{aligned} \tag{7.53}$$

where the final expression comes from evaluating the flux within the upstream region. This yields an equation for the derivative of the velocity

$$\tfrac{4}{3}\frac{\mu}{\Gamma_\rho}\,u\,\partial_x u \;=\; \tfrac{1}{2}(\gamma+1)\,(u^2-u_1^2) \;-\; \gamma\frac{\Gamma_p}{\Gamma_\rho}\,(u-u_1) \quad . \tag{7.54}$$

Introducing the kinematic viscosity, $\nu = \mu/\rho = \mu u/\Gamma_\rho$, the equation becomes

$$\begin{aligned} \tfrac{4}{3}\nu\,\partial_x u \;&=\; \tfrac{1}{2}(\gamma+1)\,(u-u_1)\left[u+u_1-\frac{2\gamma\Gamma_p}{(\gamma+1)\Gamma_\rho}\right] \\ &=\; \tfrac{1}{2}(\gamma+1)\,(u-u_1)(u-u_2) \quad , \end{aligned} \tag{7.55}$$

where the downstream velocity is

$$u_2 \;=\; \frac{2\gamma\Gamma_p}{(\gamma+1)\Gamma_\rho} \;-\; u_1 \;=\; \left[\frac{\gamma-1}{\gamma+1} \;-\; \frac{2}{(\gamma+1)M_1^2}\right] u_1 \quad . \tag{7.56}$$

as can be verified from eq. (7.35).

Equation (7.55) shows that the velocity profile $u(x)$ transitions between the two fixed points, u_2 and u_1 — these are the only two fixed points for velocity. Behavior in the vicinity of the fixed points is found by proposing small perturbations

$$u(x) \;=\; u_1 \;+\; \delta u_+(x) \;=\; u_2 \;-\; \delta u_-(x) \quad .$$

Introducing these into eq. (7.55) yields

$$\tfrac{4}{3}\nu\,\partial_x\,\delta u_\pm \;=\; \pm\,\tfrac{1}{2}(\gamma+1)\,(u_1-u_2)\,\delta u_\pm \quad . \tag{7.57}$$

The perturbations grow or decay with a length scale

$$\lambda \;=\; \frac{8\nu}{3(\gamma+1)\,|u_1-u_2|} \quad . \tag{7.58}$$

As x changes in the direction of the flow the downstream fixed point u_2 will be an attractor *only* if $|u_1| > |u_2|$. This is another piece of evidence that shocks must produce a transition from supersonic to subsonic flow.

If we assume for simplicity that ν is a constant then the solution to eq. (7.55) is

$$u(x) \;=\; \tfrac{1}{2}(u_1+u_2) \;-\; \tfrac{1}{2}|u_1-u_2|\,\tanh(x/2\lambda) \quad . \tag{7.59}$$

The hyperbolic tangent is essentially a smooth version of the step function. Rather than jumping discontinuously the velocity changes from its upstream value, u_1, to its downstream value, u_2, smoothly with a characteristic variation scale length

$$\lambda \;=\; \frac{8\nu}{3(\gamma+1)\,|u_1-u_2|} \;=\; \frac{4}{3}\,\frac{M_1}{M_1^2-1}\,\frac{\nu}{c_{s1}} \quad , \tag{7.60}$$

after using eq. (7.35). In the limit of ideal hydrodynamics, $\nu \sim \mu \to 0$, the shock becomes a genuine discontinuity, $\lambda \to 0$. Paradoxically, the specific entropy still jumps by the same amount, given by eq. (7.39), even in this limit. Taking $\mu \to 0$ does not result in $\dot{Q} \to 0$, since $\partial_x u \sim u/\lambda \to \infty$ at the same time.[5]

Viscosity arises from a random walk of the particles composing the fluid. Motion at typical speed c_s with mean-free-path $\ell_{\rm mfp}$, results in diffusion with coefficient $D \sim c_s\,\ell_{\rm mfp}$. This includes the kinematic viscosity so $\nu/c_s \sim \ell_{\rm mfp}$ the mean free path. The shock's thickness is therefore comparable to the mean-free-path of the fluid particles.

Both extreme limits of expression (7.60) are somewhat problematic. Shocks with extremely large Mach number are far thinner than a mean-free path: $\lambda \sim M_1^{-1}\ell_{\rm mfp}$. Fluid equation are not, however, physically valid on scales smaller than the mean-free-paths of their particles. It thus appears that strong shocks, i.e. $M_1 \gg 1$, cannot be treated using

[5]This is an example of a well-known situation where taking a limit, $\mu \to 0$, yields a result entirely different from setting $\mu = 0$ at the outset.

fluid equations. It is in this limit, however, that we cannot make the assumption that the kinematic viscosity ν is constant. In reality ν increases with temperature and thus the shock thickness may remain greater than the mean-free path.

The shock thickness in expression (7.60) diverges in the opposite limit — a vanishingly weak shock, $M_1 \to 1$. This is is the same limit in which the shock may be treated as a linear perturbation: a sound wave. Adding viscosity to eq. (7.1) introduces diffusion leading to a front with thickness

$$\lambda \sim \sqrt{\nu t} \quad , \tag{7.61}$$

growing with time. This behavior is standard in linear equations with diffusion. It is the steepening effect of the non-linearity that limits the shock thickness to the finite value of eq. (7.60). As the non-linearity is taken to zero the thickness grows without bound. In any actual problem it would grow over time, and if the Mach number was sufficiently small it would fail to reach a steady state.

Chapter 8

Magnetohydrodynamics

8A Plasma: a fluid

A plasma is a gas (hence a fluid) composed of positive ions and free electrons (i.e. not quantum-mechanically bound to the ions). So far we have been careful to develop fluid mechanics without making reference to the individual particles composing the fluid. Everything we have done do far will therefore apply directly to plasmas. In particular the plasma will be characterized by the same five continuous functions of space and time: mass density, $\rho(\mathbf{x}, t)$; pressure, $p(\mathbf{x}, t)$; and three velocity components, $\mathbf{u}(\mathbf{x}, t)$. For a general fluid we defined the velocity field $\mathbf{u}$ using the *center of mass* of a fluid element at a given point — it did not refer to velocities of individual particles composing the fluid. This same logic applies to a plasma: the *plasma velocity* at a point $\mathbf{x}$ is the mass-weighted average of all particle velocities (ions and electrons) in a fluid element surrounding that point. Since ions are far more massive than electrons[1] they dominate the mass-weighted average and $\mathbf{u}(\mathbf{x}, t)$ can be safely regarded as a mean velocity of ions.

The pressure of a plasma should be interpreted with some care. Pressure is defined, in its most basic form, in terms of the thermal energy, or "internal energy", of a fluid element. Since ions and electrons are point particles we can set $\gamma = 5/3$ from now on, and the internal energy of a fluid element with volume δV is

$$E_T \;=\; \frac{3}{2}\, p\, \delta V \quad . \tag{8.1}$$

This is the energy in random (thermal) motions of the particles relative to the element's center of mass. Both the ions and electrons are moving randomly relative to the center of mass, and both contribute to the plasma's thermal energy. The first assumption we shall make, and justify below, is that the electrons and ions interact (collide with one another) enough that they are at the same temperature T. The mean kinetic energy of any particle, be it electron or ion, is therefore the same: $3k_{\rm B}T/2$. Introducing $\bar{m}$ as mean mass of a particle, then the number of particles per unit volume (ion or electron) is $\rho/\bar{m}$. The ideal gas law for a plasma then attributes $k_{\rm B}T$ to each particle to yield a pressure

$$p \;=\; \frac{k_{\rm B}}{\bar{m}}\, \rho\, T \quad . \tag{8.2}$$

[1]The lightest ion, H^+, otherwise known as a proton, is 1836 times heavier than an electron. All other ions are heavier still.

For the simplest plasma, consisting of equal numbers of protons and electrons,

$$\bar{m} \simeq \tfrac{1}{2}m_e + \tfrac{1}{2}m_p \ \simeq \tfrac{1}{2}m_p \quad , \tag{8.3}$$

since $m_e \ll m_p$. Using this in eq. (8.2) shows that the plasma pressure is twice the pressure the ions alone would have at the same temperature. This is because the electrons have equal random kinetic energy (equipartition) and contribute an equal partial pressure.

The foregoing suggests that the dynamics of a plasma are little different than the dynamics of a gas of uncharged particles, often called a "neutral fluid".[2] In many respects this is true, and we have previously used fluid mechanics to treat plasmas such as the Solar wind or the Solar convection zone. Now we see that we were justified in so doing. The main difference between a plasma and a neutral fluid lies in their electro-magnetic properties. The electrons and ion composing the plasma will respond to electric and magnetic fields while neutral particles will not.

8B Plasma: a good conductor

Since the particles composing a plasma are electrically charged they will respond to electric and magnetic fields. We are familiar with such behavior from *electrical conductors* which include a population of electrons free to respond to electric and magnetic fields. The most familiar conductors are solids, like copper, where ions are frozen into a lattice, and do not readily respond to fields. A plasma is more similar to a conducting *fluid*, such as mercury, where the electrons and ions flow independently. In either case the electrons, being much lighter, respond most readily to electric fields and endow the conductor with its characteristic properties. The electrons in a plasma are similarly quick to respond and so plasmas and conventional conductors have very similar electro-magnetic properties. *A plasma is basically an electrically conducting fluid.*

Electrons in any conductor can move to create local charge density concentrations. For example, if they clump up in one place, they will overwhelm the ions to produce a region of negative charge density. According to Gauss's law this will create an electric field directed inward toward the center of the clump. This field has the effect of moving electrons out of the clump and restoring local charge neutrality. Left to itself, the conductor will thus try to remain *charge neutral* at all points — this is one of the defining properties of a conductor.

Departures from neutrality will persist only for as long as it takes the electrons to respond to their own electric field. Provided they are relatively unimpeded by ions (as they almost always are in a plasma) the electrons will cancel out charge density fluctuations in a time called the *electron plasma period*,

$$\tau_{pe} \ = \ \sqrt{\frac{\pi m_e}{e^2 n_e}} \ = \ 3\times 10^{-9}\,\mathrm{sec}\left(\frac{n_e}{10^9 \mathrm{cm}^{-3}}\right)^{-1/2} \quad , \tag{8.4}$$

where e and m_e are the charge and mass of the electrons and n_e is the number density of electrons (i.e. number per unit volume). When electron densities approach those of the solar

[2]The term is deceptive since it suggests a plasma is non-neutral, or charged. As we shall see plasmas are generally *charge neutral*. The term "neutral fluid" is supposed to refer to the charge neutrality of the individual particles composing the fluid.

corona, $n_e \sim 10^9\,\mathrm{cm}^{-3}$, charge imbalances can persist no longer than a few *nanoseconds*. For the much higher electrons densities found in solid copper, $n_e \sim 10^{23}\,\mathrm{cm}^{-3}$, the plasma period is thirty *femtoseconds* (3×10^{-16} sec), matching the period of an X-ray photon with $\lambda = 1.0$ nm. For electromagnetic waves whose periods greatly exceed this, such as visible light (period $\tau = 10^{-13}$ sec), the wave's slowly oscillating electric field will be effectively screened by the mobile electrons; this gives copper is shiny appearance.

For most astrophysical plasmas τ_{pe} is far shorter than evolution of interest. Solar flares, for example, occur over seconds or minutes, during which the nano-second lifetime of any charge density imbalance is negligible. We therefore adopt the approximation that *plasmas remain completely charge neutral*. This matches our experience with more familiar conductors.

While conductors, including plasmas, have no net charge they can support a current resulting from relative motions by their differently charged components. In a metal conductor the electrons move to create the current. In a plasma the ions are free to move, but the electrons still carry most of the current. In either case the current density $\mathbf{J}(\mathbf{x}, t)$ within the conductor quantifies the relative motion.

A plasma is therefore described by eight quantities: its density and pressure, and three components each of a velocity field, $\mathbf{u}(\mathbf{x}, t)$, and current density $\mathbf{J}(\mathbf{x}, t)$. It must be stressed, and constantly borne in mind, that *current* $\mathbf{J}$ *and velocity* $\mathbf{u}$ *are completely independent properties.* This should not seem surprising since it applies to conventional conductors as it does to plasmas. Throwing a copper wire across the room at velocity $\mathbf{u} \neq 0$ *will not* produce a current in wire since the electrons and ions within the wire move together: $\mathbf{J} = 0$. On the other hand, connecting the wire to a battery will create a current carried by a current density $\mathbf{J} \neq 0$ within the wire. Doing this *will not* accelerate the wire so its center of mass will remain stationary: $\mathbf{u} = 0$.

Thus we see that the velocity $\mathbf{u}$ and current density $\mathbf{J}$ of a simple wire conductor are *unrelated to one another*. So it is with a plasma: the velocity field $\mathbf{u}(\mathbf{x}, t)$ and current density field $\mathbf{J}(\mathbf{x}, t)$ are independent properties. The velocity of a fluid element describes its center of mass, so it mainly tells us how the ions are moving. The current density tells us how the electrons are moving *relative* to the ions — $\mathbf{J} = 0$ means the two species are moving together. Thus we can think of $\mathbf{u}$ as characterizing the ions and $\mathbf{J}$ as characterizing the electrons. The two species must have equal and opposite charge density, but are free to move independently, so $\mathbf{u}$ and $\mathbf{J}$ are independent and unrelated.

8C Plasma evolution

The dynamics of the plasma are determined by equations for the time-evolution of each of its eight quantities. The first five quantities, ρ, p, u_x, u_y and u_z, are the same as for a neutral fluid and have the same meaning. Their time-evolution is therefore governed by the same equations: (C), (M) and (E). The mass density, $\rho(\mathbf{x}, t)$ evolves according to the continuity equation. The pressure evolves according to the energy equation (E), with a volumetric heating rate $\dot{Q}$ which includes sources we considered before, such as radiative loss, or thermal conduction, as well as Ohmic heating due to the current density $\mathbf{J}$.

The momentum equation, (M), includes a force density $\mathbf{f}$ with gravitational and viscous stress. Since the plasma is charge-neutral it will not feel a force from an electric field. A

fluid element includes electrons and ions which each feel a force from the electric field. Since the fluid element includes equal charges of both signs the net forces cancel out and there is no *net* force on it from the electric field.

If there is a magnetic field, $\mathbf{B}(\mathbf{x},t)$, then the element will experience a force due its current

$$\mathbf{f}^{(B)} \;=\; \frac{1}{c}\,\mathbf{J}\times\mathbf{B} \quad , \tag{8.5}$$

called the *Lorentz force*. This is the sum of forces on the individual species due to their relative motion.

A complete evolutionary prescription for the plasma requires some equation governing the evolution of current density $\mathbf{J}(\mathbf{x},t)$. Since this field contributes to the evolution of the velocity, through the Lorentz force, the system cannot be evolved self-consistently without a means of determining the current. Since it reflects the velocity of the electrons we might try using a version of Newton's second law, in analogy to our treatment of the center-of-mass velocity $\mathbf{u}$. Electric and magnetic forces, among others, will determine how the electron velocities increase.

This is not, however, the typical approach for finding the current in a conductor. The far more common approach is to relate current to electric field through *Ohm's law*. This is also how the plasma evolution is typically closed.

8D Evolution of J: Ohm's law → the induction equation

If the plasma were a *perfect* conductor (like a superconductor) then the current $\mathbf{J}$ could exist in the absence of any electric field. If the electrons collide with the ions, however, there will be a finite conductivity, σ, relating the current to the electric field. This relation, Ohm's law, applies strictly only within the center of mass frame. Denoting by $\mathbf{E}'$ the electric field in this frame, Ohm's law states that $\mathbf{J}=\sigma\mathbf{E}'$. The case of a perfect conductor corresponds to the limit of infinite conductivity, $\sigma\to\infty$, whereby $\mathbf{J}$ remains finite even as $\mathbf{E}'\to 0$. Indeed, the electric field *must* vanish in a perfect conductor, $\mathbf{E}'=0$, otherwise there would be an infinite current density. This is what we first learn about a conductor: the electric field vanishes within it.

Maxwell's equations, however, applies to the electric and magnetic fields, $\mathbf{E}(\mathbf{x},t)$ and $\mathbf{B}(\mathbf{x},t)$, in the stationary laboratory frame. These relate to the electric field experienced by fluid elements through a kind of Lorentz transformation

$$\mathbf{E}'(\mathbf{x},t) \;=\; \mathbf{E}(\mathbf{x},t) \;+\; \frac{1}{c}\,\mathbf{u}(\mathbf{x},t)\times\mathbf{B}(\mathbf{x},t) \quad , \tag{8.6}$$

albeit in the non-relativistic limit ($u/c\ll 1$). This is the electric field which sets the current according to Ohm's law: $\mathbf{E}'=\mathbf{J}/\sigma$. The electric field in the laboratory frame is therefore

$$\mathbf{E}(\mathbf{x},t) \;=\; -\,\frac{1}{c}\,\mathbf{u}(\mathbf{x},t)\times\mathbf{B}(\mathbf{x},t) \;+\; \frac{1}{\sigma}\,\mathbf{J}(\mathbf{x},t) \quad . \tag{8.7}$$

It is this field to which Faraday's law applies:

$$\frac{\partial\mathbf{B}}{\partial t} \;=\; -\,c\nabla\times\mathbf{E} \;=\; \nabla\times(\mathbf{u}\times\mathbf{B}) \;-\; \nabla\times\left(\frac{c}{\sigma}\,\mathbf{J}\right) \quad . \tag{8.8}$$

This provides a prescription for updating $\mathbf{B}(\mathbf{x},t)$, and since current is related to magnetic field it can be used to update $\mathbf{J}$ as well.

Current and magnetic field are related by Ampère's law,

$$\nabla \times \mathbf{B} \;=\; \frac{4\pi}{c}\mathbf{J} \;+\; \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} \quad . \tag{8.9}$$

The final term is known as the *displacement current* to emphasize its similarity to the term before it: the current. Historically the displacement current was the last to be identified; it was introduced by Maxwell to complete the equations bearing his name. He deduced it theoretically rather than from experiments, because it tends to be extremely small whenever actual currents are present – as they invariably are in laboratory experiments. Since actual currents are generally present in a plasma the displacement current will be very small and will be ignored.

To verify this claim we use eq. (8.20) in the super-conducting limit to estimate $|\mathbf{E}| \sim (u/c)\,B$. For cases of non-relativistic plasmas, to which we confine our attention, $u/c \ll 1$ and the electric field is always much smaller than the magnetic field. The displacement current has relative size

$$\frac{c^{-1}|\partial \mathbf{E}/\partial t|}{|\nabla \times \mathbf{B}|} \;\sim\; \frac{c^{-1}|\mathbf{E}|/\tau}{B/L} \;\sim\; \frac{uL}{c^2\tau} \quad ,$$

where τ and L are typical time-scales and length-scales over which the fields vary. Provided we restrict attention to evolution on time-scales

$$\tau \;\gg\; \frac{uL}{c^2} \;=\; \frac{u}{c}\times\frac{L}{c} \quad , \tag{8.10}$$

the displacement current will be negligible. The factor L/c is the light transit time across the scale of interest. Given its essential role in electromagnetic waves, it is not surprising that the displacement current operates on time-scales comparable to light transit. A solar active region ($L \sim 100,000$ km across) is crossed by light in $L/c \sim 0.3$ sec. Provided we are considering only non-relativistic motions,[3] say $u/c \sim 10^{-3}$, then the displacement current will not affect evolution on time-scales longer than one millisecond. In the same vein as previous assumptions, we restrict ourselves to evolution slower than this and hereafter completely neglect the displacement current. This returns us to the pre-Maxwell version of Ampère's law.

Without the displacement current Ampère's law provides a direct relation between the magnetic field and the current. Given the magnetic field, the current density is

$$\mathbf{J} \;=\; \frac{c}{4\pi}\,\nabla \times \mathbf{B} \quad . \tag{8.11}$$

Since eq. (8.8) provides a prescription for time-advancing the magnetic field, $\mathbf{B}$ is used as a fundamental field in MHD and $\mathbf{J}$ is derived *from it* using eq. (8.11). In fact this equation can be used to *eliminate* $\mathbf{J}$ entirely from MHD, which turns out to be preferable.

[3] If you had not done so until now, please recall that $\mathbf{u}$ is *not* the speed of any single particle but of the center of mass of all particles. The particles composing the plasma could be moving relativistically provided the entire collection moved more slowly.

It can seem odd at first that a property of the plasma, namely the current density $\mathbf{J}$, is dictated by the magnetic field. In more conventional electrodynamics classes we typically *begin* with the current and *find* the magnetic field from it. This leads to the erroneous impression that the current *causes* the magnetic field. Equations such as eq. (8.11) show an equality between two things, but the relationship *cannot* be interpreted as a causal one — the equal sign works in both directions. In MHD we consider the magnetic field to be fundamental and derive the current *from it* using eq. (8.11).

Our intuition about causation has been derailed, essentially, because Ohm's law omits the inertial term (i.e. $m_e\mathbf{a}_e$) whereby forces "drive" electron motion. It assumes instead that the electrons are so light ($m_e \to 0$) that they must forever be in *force-balance* ($\mathbf{F} = 0$). This is the meaning of Ohm's law and its consequence is that the electrons instantaneously adjust the current $\mathbf{J}$ to satisfy eqs. (8.8) and (8.11).

To complete our reasoning we substitute eq. (8.11) into eq. (8.8) to form an equation for the time-evolution of $\mathbf{B}(\mathbf{x}, t)$ called *the induction equation*

$$\frac{\partial \mathbf{B}}{\partial t} \;=\; \nabla \times (\mathbf{u} \times \mathbf{B}) \;-\; \nabla \times (\eta\, \nabla \times \mathbf{B}) \quad . \tag{8.12}$$

The conductivity of the plasma, σ, has be incorporated into the coefficient

$$\eta \;\equiv\; \frac{c^2}{4\pi\sigma} \;\simeq\; 0.94 \times 10^4\,\mathrm{cm}^2/\mathrm{sec}\,\left(\frac{T}{10^6\,\mathrm{K}}\right)^{-3/2} \quad , \tag{8.13}$$

known as *magnetic diffusivity* or (somewhat inaccurately) as *resistivity*.[4] A plasma's resistivity depends on temperature in a surprising way: raising the temperature lowers the resistivity.[5] Faster thermal motions make the electrons and ions less sensitive to collisions, thereby increasing the conductivity. In more conventional conductors, such as copper ($\eta = 13 \times 10^4\,\mathrm{cm}^2/\mathrm{sec}$), electrons interact quantum mechanically and increasing temperature adds disorder to those interactions, *lowering* the conductivity. These two regimes are extremely different even though they both lead (approximately) to an Ohm's law relationship.

The ratio of scales of the two right-hand terms in eq. (8.12)

$$\frac{|\nabla \times (\mathbf{u} \times \mathbf{B})|}{|\nabla \times (\eta \nabla \times \mathbf{B})|} \;\sim\; \frac{uL}{\eta} \;\equiv\; \mathrm{Rm} \quad , \tag{8.14}$$

is known as the *Magnetic Reynold's number*, in analogy with the conventional hydrodynamic Reynold's number, Re=uL/ν. For the solar corona, $T = 10^6$ K, a small feature ($L = 1000$ km $= 10^8$ cm) moving at modest speed ($u = 1$ km/sec $= 10^5$ cm/sec) is characterized by Rm$= 10^9$. This means the resistive term is one billion times smaller than the other two terms. The term can safely be dropped entirely leading to a version of the equation

$$\frac{\partial \mathbf{B}}{\partial t} \;=\; \nabla \times (\mathbf{u} \times \mathbf{B}) \quad , \tag{8.15}$$

[4]The conventional definition of "resistivity" is simply the reciprocal of conductivity. Here the reciprocal is rescaled by a factor of $c^2/4\pi$ — in Gaussian units. This gives it the units of a diffusion coefficient, $\mathrm{cm}^2/\mathrm{sec}$.

[5]This expression assumes a fully-ionized hydrogen plasma in which electrons scatter from ions via classical Coulomb scattering. These are reasonable, and very common, assumptions for most solar and astrophysical plasmas.

known as the *ideal induction equation*. The ideal induction equation characterizes magnetic evolution in a *perfect conductor*,[6] but applies approximately to any plasma where $\mathrm{Rm} \gg 1$ — which is to say almost all astrophysical plasmas.

Just as with viscosity, the chief reason that we mention resistivity at all is that it is sometimes retained to obtain the "curative" properties of diffusion. Expanding the final term in eq. (8.12), and using a vector calculus identity, gives

$$\frac{\partial \mathbf{B}}{\partial t} \;=\; \text{stuff} \;\cdots\; +\; \eta \nabla^2 \mathbf{B} \quad , \tag{8.16}$$

where "stuff" includes only zeroth and first derivatives of $\mathbf{B}$. This makes it clear that resistivity *diffuses* the magnetic field, with diffusion coefficient η. It will tend to "smooth" out sharp features in $\mathbf{B}$, which could otherwise become discontinuities (shocks) were we to insist on $\eta = 0$.

8E The MHD equations

The foregoing discussion has led us to the complete, closed set of MHD equations for the eight fields, $\rho(\mathbf{x},t)$, $\mathbf{u}(\mathbf{x},t)$, $p(\mathbf{x},t)$ and $\mathbf{B}(\mathbf{x},t)$. The fields are time-advanced through the eight equations

$$\frac{\partial \rho}{\partial t} \;+\; \nabla \cdot (\rho \mathbf{u}) \;=\; 0 \quad , \qquad \text{(C) continuity}$$

$$\rho \frac{\partial \mathbf{u}}{\partial t} \;+\; \rho(\mathbf{u}\cdot\nabla)\mathbf{u} \;=\; -\nabla p \;+\; \frac{1}{4\pi}(\nabla \times \mathbf{B}) \times \mathbf{B} \;+\; \mathbf{f} \qquad \text{(MM) momentum}$$

$$\frac{\partial p}{\partial t} \;+\; \mathbf{u}\cdot\nabla p \;=\; -\tfrac{5}{3}\, p(\nabla \cdot \mathbf{u}) \;+\; \tfrac{2}{3}\frac{\eta}{4\pi}|\nabla \times \mathbf{B}|^2 \;+\; \tfrac{2}{3}\dot{Q} \qquad \text{(ME) energy}$$

$$\frac{\partial \mathbf{B}}{\partial t} \;=\; \nabla \times (\mathbf{u} \times \mathbf{B}) \;-\; \nabla \times (\eta\, \nabla \times \mathbf{B}) \quad . \qquad \text{(MI) induction}$$

The factors $\dot{Q}$ and $\mathbf{f}$ once again stand for any number of contributions to volumetric heating and force density; these are included or omitted as the particular problem demands. The Lorentz force has been included explicitly so is is no longer candidate for $\mathbf{f}$. Ohmic heating, $\mathbf{E}'\cdot\mathbf{J} = |\mathbf{J}|^2/\sigma$, has been explicitly included in the energy equation so it will not be considered for $\dot{Q}$. Most significantly, these eight equations (counting separately the components of $\mathbf{u}$ and $\mathbf{B}$) depend only on the same fields they advance — they are closed. It is evident that the fields are completely coupled: $\mathbf{B}$ affects the evolution of $\mathbf{u}$ through the Lorentz force, and $\mathbf{u}$ affects the evolution of $\mathbf{B}$ in the induction equation.

A notable exception to mutual coupling occurs when the magnetic field is relatively weak. The ratio of the pressure force to the Lorentz force scales as

$$\frac{|\nabla p|}{|(\nabla \times \mathbf{B}) \times \mathbf{B}|/4\pi} \;\sim\; \frac{1}{2}\,\frac{8\pi p}{|\mathbf{B}|^2} \;=\; \frac{\beta}{2} \quad , \tag{8.17}$$

[6] Interestingly, the quantum mechanics underlying *super-conductivity* requires $\mathbf{B} = 0$ (the Meisner effect). This is indeed one solution to eq. (8.15), but a rather special one. It seems that super-conductors are only a tiny subset of the hypothetical class of *perfect conductors*.

where $\beta = 8\pi p/B^2$ expresses the ratio of plasma pressure to magnetic pressure. The Lorentz force may be neglected when $\beta \gg 1$, for example in the solar interior. The Ohmic heating term may be neglected for similar reasons. Dropping these terms from (MM) and (ME) reduces them to their hydrodynamic versions, eqs. (M) and (E). In other words the fluid is unaffected by the magnetic field, and behaves exactly as a neutral fluid. The fluid does, however, affect the magnetic field through eq. (MI). It is known as a *passive* magnetic field and (MI) may be solved *after* $\mathbf{u}(\mathbf{x}, t)$ has been found. It is then a linear equation for the unknown, $\mathbf{B}(\mathbf{x}, t)$.

We have not made explicit use of the one of Maxwell's equations stating $\nabla \cdot \mathbf{B} = 0$. This is sometimes included among the MHD equations, but if it were truly independent it would over-determine the system: nine equation for eight unknowns. Taking the divergence of the induction equation, eq. (MI), yields

$$\frac{\partial}{\partial t}(\nabla \cdot \mathbf{B}) \;=\; 0 \quad ,$$

using only the vector identity that the divergence of a curl vanishes. Thus we must *begin* with an initial condition satisfying the condition $\nabla \cdot \mathbf{B} = 0$. After that the induction equation will assure that the condition is satisfied at all later times.

Three fields notably absent from the set of MHD equations above are temperature, T; current density, $\mathbf{J}$; and electric field $\mathbf{E}$. Since the system is closed, these fields are not required to determine the evolution of the plasma. They are *auxiliary quantities* which may be derived from the fundamental quantities should the desire arise:

$$T \;=\; \frac{\bar{m}}{k_{\rm B}}\frac{p}{\rho} \quad , \tag{8.18}$$

$$\mathbf{J} \;=\; \frac{c}{4\pi}\nabla \times \mathbf{B} \quad , \tag{8.19}$$

$$\mathbf{E} \;=\; -\frac{1}{c}\mathbf{u} \times \mathbf{B} \;+\; \frac{\eta}{c}\nabla \times \mathbf{B} \tag{8.20}$$

The approximations used to arrive at the MHD equations are all equivalent to assuming that evolution is *slow* compared to microscopic time scales of the plasma. The fast times scales we have chosen to skip past are: 1) the plasma period τ_{pe} required to re-establish charge neutrality, 2) the ion-electron collision time τ_{ei} required to equalize the temperatures of the electrons and ions, and 3) the light transit time, L/c, on which the displacement current could be significant. Considering only evolution slower than all of these, the plasma behaves as a single conducting fluid whose evolution is governed by the equations of MHD. Following more rapid evolution[7] requires the more sophisticated treatment known as *plasma physics. MHD is the low-frequency limit of plasma physics.* This is the limit most appropriate for astrophysical process since these tend to occur on long time scales.

8F Energy conservation in a plasma

With the inclusion of a magnetic field, which couples to the plasma motion, we must include magnetic energy

$$E_M \;=\; \frac{1}{8\pi}\int |\mathbf{B}(\mathbf{x}, t)|^2 \, d^3x \quad . \tag{8.21}$$

[7]The most common violation arises when density is low enough that τ_{ei} becomes too long to negelct.

Assuming a non-relativistic plasma the energy in the energy of the electric field will be smaller than this by $(u/c)^2 \ll 1$ and can be neglected. The total energy within a fluid parcel $\mathcal{R}$ consists of the bulk kinetic energy, the thermal energy, and the magnetic energy

$$E \;=\; \int\limits_{\mathcal{R}} \left(\tfrac{1}{2}\rho|\mathbf{u}|^2 \;+\; \tfrac{3}{2}\,p \;+\; \tfrac{1}{8\pi}|\mathbf{B}|^2 \right) d^3x \quad . \tag{8.22}$$

We can time-differentiate the energy integral in eq. (8.22) and assume the parcel $\mathcal{R}$ remains fixed, with the possibility of flow through its imaginary boundaries $\partial\mathcal{R}$. Partial time derivatives of each field, ρ, $\mathbf{u}$, p and $\mathbf{B}$, can be replaced using eqs. (C), (MM), (ME) and (MI) respectively. After considerable algebra, involving integration by parts, we arrive at the conservation law

$$\frac{dE}{dt} \;=\; -\oint_{\partial\mathcal{R}} \left(\tfrac{1}{2}\rho|\mathbf{u}|^2\mathbf{u} + \tfrac{5}{2}p\mathbf{u} + \mathbf{S} \right) \cdot d\mathbf{a} \;+\; \int_{\mathcal{R}} \mathbf{u}\cdot\mathbf{f}\, d^3x \;+\; \int_{\mathcal{R}} \dot{Q}\, d^3x \quad . \tag{8.23}$$

This differs from the hydrodynamic version only by the inclusion of a new energy flux term, the Poynting flux,

$$\mathbf{S} \;=\; \frac{c}{4\pi}\,\mathbf{E}\times\mathbf{B} \;=\; \frac{1}{4\pi}|\mathbf{B}|^2\mathbf{u} \;-\; \frac{1}{4\pi}(\mathbf{B}\cdot\mathbf{u})\mathbf{B} \;+\; \frac{\eta}{4\pi}(\nabla\times\mathbf{B})\times\mathbf{B} \quad . \tag{8.24}$$

The final expression, in terms of fundamental MHD quantities, comes from using eq. (8.20) to eliminate $\mathbf{E}$. The Ohmic heating term is included in eq. (ME), in order to compensate for the magnetic energy lost to resistive dissipation. The first term in the final expression resembles the advection of the magnetic energy density, $B^2/8\pi$, by the flow, except that it is twice as large. The factor of two accounts for work done by the flow is analogous to the enthalpy flux. It would correspond to $\gamma/(\gamma-1)$ if the magnetic field were a fluid with $\gamma = 2$. We find below that this is a legitimate interpretation where flow is perpendicular to the field ($\mathbf{u}\cdot\mathbf{B} = 0$).

8G Turbulent magnetic advection: the α-effect

The analysis from Chapter 6 can be repeated for any number of other physical properties influenced by turbulence. One effect warranting special attention is that of a turbulent velocity on a magnetic field through the induction equation. The analysis, sometimes referred to as *mean field electrodynamics*, follows the same steps as in Chapter 6 for temperature, except it involves vectors and many cross products.[8] The result has one interesting new outcome, known as the α-effect, which is frequently invoked in modeling the generation of magnetic fields. This term is not strictly speaking part of MHD, indeed it is derived *using* MHD. Nevertheless, it is sometimes added to the traditional induction equation when fluid turbulences is present. We present here a brief sketch out its derivation.

The magnetic field evolves according to the induction equation, (MI), and its predecessor Faraday's law, can be written in index notation

$$\partial_t B_i \;=\; -\,\epsilon_{ijk}\partial_j(cE_k) \;=\; B_j\partial_j u_i \;-\; u_j\partial_j B_i \quad . \tag{8.25}$$

[8] This leads to liberal use of the totally anti-symmetric tensor ϵ_{ijk}.

The electric field $E_k = -\epsilon_{kmn} u_m B_n/c$, follows from assuming no resistivity (ideal induction equation) which could lead to an electric field in the frame of the plasma. Taking the mean of the first form gives an obvious relation

$$\partial_t \bar{B}_i \;=\; -\,\epsilon_{ijk}\partial_j(c\bar{E}_k) \quad , \tag{8.26}$$

where the mean electric field, or EMF, is

$$\bar{E}_k \;=\; -\,\epsilon_{kmn}\langle u_m B'_n\rangle/c \quad . \tag{8.27}$$

Subtracting this from eq. (16.12), and using first-order smoothing, gives

$$\partial_t B'_i \;= \bar{B}_j\partial_j u_i \;-\; u_j\partial_j\bar{B}_i \quad . \tag{8.28}$$

This can be integrated to give an explicit expression for the fluctuating component of the magnetic field in terms of the (fluctuating) velocity and the mean magnetic field

$$B'_i(t) \;=\; \int^t \bar{B}_j(s)\partial_j u_i(s)\,ds \;-\; \int^t u_j(s)\partial_j\bar{B}_i(s)\,ds \quad , \tag{8.29}$$

where once again $\mathbf{x}$ is implicit in all fields.

Next we use this in the average and invoke the separation of time scales described above. Doing so yields the average

$$\begin{aligned}\langle u_m B'_n\rangle \;&=\; \bar{B}_j \int_{-\infty}^{0} \langle u_m(t+\Delta t)\partial_j u_n(t)\rangle\, d(\Delta t) \\ &\qquad -\;(\partial_j\bar{B}_n)\int_{-\infty}^{0}\langle u_m(t+\Delta t)u_j(t)\rangle\, d(\Delta t) \qquad (8.30)\\ &=\; \bar{B}_j\int_{-\infty}^{0}\langle u_m(t+\Delta t)\partial_j u_n(t)\rangle\, d(\Delta t) \;-\; \tilde{\kappa}^{(\mathrm{eff})}_{mj}\partial_j\bar{B}_n \quad , \qquad (8.31)\end{aligned}$$

after using eq. (6.14) for the integral of the two-time, one-point velocity correlation. The tensor $\tilde{\kappa}^{(\mathrm{eff})}_{ij}$ was originally introduced to relate mean heat flux to mean temperature gradients: it was the turbulent thermal conductivity. It has miraculously re-appeared to describe the effect of turbulence on magnetic fields. Using this in eq. (8.27) gives a simple expression for the mean EMF

$$c\bar{E}_i \;=\; -\,\epsilon_{imn}\langle u_m B'_n\rangle \;=\; \beta_{ijn}\partial_j\bar{B}_n \;-\; \alpha_{ij}\bar{B}_j \quad , \tag{8.32}$$

where the coefficient tensors can be read off

$$\alpha_{ij} \;=\; \int_{-\infty}^{0}\langle u_m(t+\Delta t)\epsilon_{imn}\partial_j u_n(t)\rangle\, d(\Delta t) \tag{8.33}$$

$$\beta_{ijn} \;=\; \epsilon_{imn}\tilde{\kappa}^{(\mathrm{eff})}_{mj} \quad . \tag{8.34}$$

This shows that the mean EMF is proportional to the mean magnetic field and its gradient. The coefficients in this linear relationship are, lamentably, tensors depending on the statistics of the turbulence. We can say little more without knowing more, or assuming more, about the turbulence itself.

The most common way that Physicists say something about a quantity about which they have no specific information is to consider symmetries of the problem. So far we have simplified our analysis by assuming the turbulence to be *stationary* — it is invariant under translation in time. As a result of assuming this symmetry we know that moments of the turbulence, such as α_{ij} and $\tilde{\kappa}^{(\mathrm{eff})}_{mj}$, are not functions of t. Another common assumption is that the turbulence is *homogeneous*, meaning it is invariant under spatial translations. Under this assumptions no property would depend on $\mathbf{x}$.

A third, independent, symmetry of the turbulence is *isotropy*, meaning its properties are unchanged by *rotations* about any axis. This is a powerful assumption to make about properties which are vectors or tensors. In a generic tensor of rank p there are 3^p different components and all of them change when the coordinate axes are rotated. It places severe constraints on these components to demand, as we propose, that every one remain invariant under every possible rotation. Indeed no vectors (rank-1 tensors) can satisfy this condition so there can be no non-trivial vector-valued property under the assumption of isotropy. There is exactly one second-rank tensor and one third-rank tensor which satisfy the condition. Not by coincidence, these are the ones with which you are most familiar, the identity δ_{ij}, and the totally anti-symmetric[9] tensor ϵ_{ijk}. Thus isotropic turbulence must be characterized by tensor moments proportional to these

$$\alpha_{ij} \;=\; \alpha\,\delta_{ij} \quad , \quad \beta_{ijk} = \eta^{(\mathrm{eff})}\,\epsilon_{ijk} \quad , \tag{8.35}$$

with constants of proportionality α and $\eta^{(\mathrm{eff})}$ which are scalars, and thus always isotropic. Thus symmetry has eliminated 8 of the 9 unknowns in α_{ij} and 26 of the 27 in β_{ijk}. It is evidently a very powerful tool, which explains its popularity among Physicists.

While we have reduced our ignorance from 9 to 1 parameter, α, we still do not know how that remaining parameter relates to the statistics of the turbulent velocity field. We evaluate the remaining scalar by taking one-third (since $\delta_{ii} = 3$) of the trace of the original tensor defined in eq. (8.33),

$$\begin{aligned} \alpha \;&=\; \tfrac{1}{3}\alpha_{ii} \;=\; \tfrac{1}{3}\int_{-\infty}^{0} \langle u_m(t+\Delta t)\epsilon_{jmn}\partial_j u_n(t)\rangle\, d(\Delta t) \\ &=\; -\tfrac{1}{3}\int_{-\infty}^{0} \langle \mathbf{u}(t+\Delta t)\cdot\nabla\times\mathbf{u}\rangle\, d(\Delta t) \;\simeq\; -\tfrac{1}{3}\tau_{\mathrm{cor}}\langle \mathbf{u}\cdot\nabla\times\mathbf{u}\rangle \quad . \end{aligned} \tag{8.36}$$

The final expression assumes (with no justification) that $\mathbf{u}$ and $\nabla\times\mathbf{u}$ decorrelate after the same time that $\mathbf{u}$ decorrelates with itself: τ_{cor}.

[9]The term "symmetry" in its name refers to its behavior under exchange of its indices, rather than rotation.

It often surprises people, at first, that the velocity and its curl are correlated at all. It is common expectation that a vector and its curl will lie in orthogonal directions and thus that $\mathbf{u} \cdot \nabla \times \mathbf{u} = 0$ everywhere. A counter example is offered by the simple helical flow

$$\mathbf{u}(r, \phi, z) \;=\; \hat{\mathbf{z}} \pm r\hat{\boldsymbol{\phi}} \;\;, \;\; \nabla \times \mathbf{u} \;=\; \pm 2\hat{\mathbf{z}} \;\;, \tag{8.37}$$

where the upper (lower) signs generate flows whose stream lines are right-handed (left-handed) helices. At least in this flow $\mathbf{u} \cdot \nabla \times \mathbf{u} = \pm 2$ is non-zero, and its sign gives the handedness of the helices. In a more complex, realistic flow there may still be some tendency toward "helical-ness" and that tendency will be characterized, statistically, by the expectation, $\langle \mathbf{u} \cdot \nabla \times \mathbf{u} \rangle \propto -\alpha$. Whenever $\alpha \neq 0$ we refer to the turbulence as *helical turbulence*; when the flow is dominated by right-handed helices, $\alpha < 0$.

Another result of assuming isotropy is that the two-time correlation tensor must also be proportional to the identity and therefore $\tilde{\kappa}_{ij}^{(\mathrm{eff})} = \tilde{\kappa}^{(\mathrm{eff})}\delta_{ij}$. Using this in eq. (8.34)

$$\beta_{ijn} \;=\; \epsilon_{ijn}\tilde{\kappa}^{(\mathrm{eff})} \;=\; \eta^{(\mathrm{eff})}\,\epsilon_{ijn} \;\;, \tag{8.38}$$

shows that $\eta^{(\mathrm{eff})} = \tilde{\kappa}^{(\mathrm{eff})}$.[10] The coefficient $\eta^{(\mathrm{eff})}$ refers to effective diffusion of magnetic field by the turbulence just as $\eta = c^2/4\pi\sigma$ was the microscopic diffusivity due to inter-particle collisions,[11] while $\tilde{\kappa}^{(\mathrm{eff})}$ characterizes the diffusion of heat by turbulence. As discussed above, all diffusion arises from some form of random motions. In the present case both magnetic field and heat are diffused by the random flow of the turbulence, so it is natural that their diffusion coefficients are the same.

Combining these terms into the mean EMF gives a turbulent version of Ohm's law

$$\bar{\mathbf{E}} \;=\; \frac{\eta^{(\mathrm{eff})}}{c}\nabla \times \bar{\mathbf{B}} \;-\; \frac{\alpha}{c}\bar{\mathbf{B}} \;=\; \frac{4\pi\eta^{(\mathrm{eff})}}{c^2}\bar{\mathbf{J}} \;-\; \frac{\alpha}{c}\bar{\mathbf{B}} \;\;, \tag{8.39}$$

where $\bar{\mathbf{J}} = (c/4\pi)\nabla \times \bar{\mathbf{B}}$ is the mean electric current. The first term is the resistivity most commonly associated with Ohm's law, but here generated from the random motion of turbulence. This leads to an effective electrical conductivity $\sigma^{(\mathrm{eff})} = c^2/4\pi\eta^{(\mathrm{eff})}$, where $\eta^{(\mathrm{eff})} = \tilde{\kappa}^{(\mathrm{eff})}$ characterizes the amplitude of the turbulence. Stronger turbulence yields a larger $\eta^{(\mathrm{eff})}$ and hence a lower conductivity (higher resistivity).

The second term on the rhs of eq. (8.39) is unlike any other in traditional electrodynamics: an electric field proportional the the magnetic field. It is referred to as the α-*effect*. In the isotropic case (scalar α) the electric field is either parallel or anti-parallel to the magnetic field (α may be positive or negative). It is worth recalling that it originated as one part of the average product $-\langle \mathbf{u}' \times \mathbf{B}' \rangle / c$. The fluctuating magnetic field, $\mathbf{B}'$ is generated by with the interaction of the fluctuating velocity with the mean magnetic field through the induction equation, eq. (8.28). This is the origin of their correlation and therefore of both terms in (8.39).

[10] This is a remarkable profound result: turbulent velocity diffuses a vector field (i.e. $\mathbf{B}$) at exactly the same rate it diffuses a scalar field (i.e. temperature). When one recalls that diffusion in any form results from a random walk, it is less surprising that the two diffusivities agree when they use the same random walker: turbulence. It is therefore possible to speak generically of *turbulent diffusion*.

[11] It is common to use the scalar β for this, but that is used in a far more importnat role with which we do not wish to cause confusion.

To finish the section we present the equation of magnetic induction supplemented with the α-effect,

$$\frac{\partial \mathbf{B}}{\partial t} \;=\; \nabla \times (\mathbf{u} \times \mathbf{B}) \;+\; \nabla \times (\alpha \mathbf{B}) \;-\; \nabla \times (\eta^{(\mathrm{eff})} \nabla \times \mathbf{B}) \quad . \tag{8.40}$$

We have dropped the over-bars, so $\mathbf{B}$ represents the mean magnetic field and $\mathbf{u}(\mathbf{x}, t)$ the *mean fluid velocity*, which we had previously omitted from our derivation. The turbulent fluctuations have been incorporated into the scalar coefficients α and $\eta^{(\mathrm{turb})}$ defined in eqs. (8.36) and (8.38). These are scalars because we have assumed the turbulence is statistically isotropic. We left them inside spatial derivatives to forego having to assume the turbulence is also homogeneous.[12]

[12]The astute reader may detect a subtle inconsistency in these two assumptions. Without spatial homogeneity statistical moments will have gradients. The direction of the gradient at some point will break the assumed isotropy since that direction will differ from others: it is the direction along the gradient. This inconsistency is genuine, but is often ignored in order to take advantage of the great simplicity of replacing the second and third rank tensors with scalars.

Chapter 9

Tools for MHD intuition

9A Magnetic field lines

Before attempting to solve the rather formidable equations of MHD we can develop a bit of intuition about them. We already had an intuitive feel for solutions to the *hydrodynamic* equations because we've lived our entire lives within a fluid: the air. MHD introduces to these equations only one new element: the magnetic field. Astrophysicists are often guided by intuition about its behavior from one basic concept — the magnetic field line.

A magnetic field line is a space curve which may be expressed a vector valued function of a single parameter, $\mathbf{r}(s)$. As s varies from $-\infty$ to $+\infty$, the vector $\mathbf{r}$ traces a curve through space — this is the field line. In order for this curve to qualify as a magnetic field line the magnetic field, $\mathbf{B}(\mathbf{x})$ must be *tangent* to the curve at every point. In other words it must satisfy a differential equation like

$$\frac{d\mathbf{r}}{ds} \;=\; \mathbf{B}[\mathbf{r}(s)] \quad . \tag{9.1}$$

This equation is solved beginning at some "initial" point, $\mathbf{r}(0) = \mathbf{r}_0$ and itegrating both forward and backward in s.

The parameter s by which we trace the field line is entirely arbitrary. We define it, implicitly, in eq. (9.1) to make the equation as simple as possible. The tangency condition will be satisfied if any scalar constant multiplies $\mathbf{B}$ in eq. (9.1), since it would not change the direction. Such scalar constants come from transforming between parameterizations. A popular alternative is to parameterize the curve using the length ℓ along the curve,

$$\frac{d\mathbf{r}}{d\ell} \;=\; \frac{\mathbf{B}[\mathbf{r}(\ell)}{\Big|\mathbf{B}[\mathbf{r}(\ell)]\Big|} \quad , \tag{9.1$'$}$$

which comes from the transformation $d\ell/ds = |\mathbf{B}|$. In a third choice of parameter, the integrated column mass per flux, μ,

$$\frac{d\mathbf{r}}{d\mu} \;=\; \frac{\mathbf{B}[\mathbf{r}(\mu)}{\rho[\mathbf{r}(\mu)]} \quad , \tag{9.1$''$}$$

where $\rho(\mathbf{x})$ is the mass density. This parameterization, using transformation $d\mu/ds = \rho$, is useful because it moves with the plasma; this is an advantage to which return shortly. All

three equations, (9.1), (9.1'), and (9.1") trace the exact same curve through space — the same field line.

You may have been led to believe that magnetic field lines (otherwise known as lines of force) were simpler to find than by solving an ordinary differential equation. In general they are not. There is one situation where they can be graphed a bit more easily; this is discussed further below, but applies only to very special circumstances. In virtually every case the only way to find a field line is to solve the ODE, eq. (9.1), beginning from some initial condition $\mathbf{r}_0$. Since that initial point can be chosen in an infinite number of ways, there are an infinite number of possible field lines — in practice we would only trace a small number of these. It is immediately evident that there is one field line passing through each and every point in space, with the notable exception of those points where $\mathbf{B}$ vanishes. These points, called *magnetic null points*, if chosen as an initial condition for eq. (9.1), will deliver a solution which is a *single point*, the null point, $\mathbf{r}(s) = \mathbf{r}_0$. These special cases are quite significant in characterizing the structure of all the the other field lines in a particular field.

The field line equation, eq. (9.1), is a coupled set of three first order ODEs (for $x(s)$, $y(s)$ and $z(s)$), and therefore about as easy to solve as any equation. The ease of solution depends entirely on the function $\mathbf{B}(\mathbf{x})$, i.e. the magnetic field. Some fields admit ready solutions while others do not. There are many fields whose field lines are *chaotic* in the full mathematical sense of the word. In fact, due to its importance to heat transport in fusion reactors, there has probably been as much work devoted to studying chaos in magnetic field lines, solving eq. (9.1), as in dynamical systems (see Guckenheimer & Holmes, 1983, for details).

Most students have encountered magnetic field lines in elementary classes on electrodynamics. Unfortunately, many students have been given erroneous information in these classes, which can be easily dismissed using eq. (9.1). Below are three of the more egregious myths often spread by well-meaning, but misinformed electrodynamicists.

- **MYTH: Magnetic field lines have no beginning or end** —A magnetic null point, $\mathbf{r}_\emptyset$, is a fixed point solution of eq. (9.1) since $d\mathbf{r}/ds = \mathbf{B}(\mathbf{r}_\emptyset) = 0$. There are several directions in which it is an attracting fixed point so $\mathbf{r}(s) \to \mathbf{r}_\emptyset$ as $s \to \infty$; there are others where it is repelling fixed point so $\mathbf{r}(s) \to \mathbf{r}_\emptyset$ as $s \to -\infty$. In all of those cases the solution asymptotically approaches $\mathbf{r}_\emptyset$ in parameter s, but reaches that point in finite distance ℓ. Thus in the first case (attracting) the field line *ends* at $\mathbf{r}_\emptyset$ and in the second (repelling) it *begins* at $\mathbf{r}_\emptyset$. This very small set of field lines, curves and surfaces, are called the *spines* and *fan* of the null point respectively. They constitute a clear demonstration that some field lines *do* have a beginning and other have an end.[1]

- **MYTH: The density of magnetic field line shows the strength of the magnetic field** — As explained above there are an infinite number of field lines in space. We always choose to graph only a subset of these by selecting a set of initial points. The density of resulting field lines will depend on how those points are selected and on the nature of the integral curves. It is *not* related in any simple way to $|\mathbf{B}(\mathbf{x})|$.

[1] An even more select group of field lines, called *separators*, have *both* a beginning and an end.

- **MYTH: Magnetic field lines are not real "things"** — We show below that a highly conductive plasma will evolve so that the matter and field lines move together. An equivalent statement, "the field line is made of plasma", proves the fallacy since being made of stuff, such as plasma, is essentially what it means to be "real". To be entirely fair, this falsehood is usually promulgated on victims who have never heard of a plasma or even a conducting fluid. There is no coupling between the magnetic field and a non-conducting fluid, like air, so field lines there are independent of "stuff" and can be regarded as imaginary. Astrophysics, however, typically occurs within a highly conductive plasma ($Rm \gg 1$) and we should be always ready to think of field lines as real things made of "stuff'.

9A.1 Cases of symmetry — the flux function

The one case where it is possible to graph field lines simply is when the magnetic field, $\mathbf{B}(\mathbf{x})$, depends on two variables rather than three. The third (ignorable) direction is called the symmetry direction and we will denote it generically as ζ; the other two will be called ξ and η. The most common symmetries are *Cartesian*, $\zeta = z$ and $(\xi, \eta) = (x, y)$; and *axisymmetry*, $\zeta = \phi$, and $(\xi, \eta) = (r, \theta)$. A magnetic field with no component in the symmetry direction can always be expressed using a *flux function* $f(\xi, \eta)$, as

$$\mathbf{B}(\xi, \eta) \;=\; \nabla \times (f \nabla \zeta) \;=\; \nabla f \times \nabla \zeta \quad . \tag{9.2}$$

The first equality makes clear that $\nabla \cdot \mathbf{B} = 0$ and the second assures there is no component of $\mathbf{B}$ parallel to $\nabla\zeta$. For the two common symmetries eq. (9.2) reduces to

$$\begin{aligned} \mathbf{B}(x, y) \;&=\; \nabla f \times \hat{\mathbf{z}} \;=\; \frac{\partial f}{\partial y}\,\hat{\mathbf{x}} \;-\; \frac{\partial f}{\partial x}\,\hat{\mathbf{y}} \quad , \\ \mathbf{B}(r, \theta) \;&=\; \nabla f \times \nabla \phi \;=\; \frac{1}{r^2 \sin\theta}\frac{\partial f}{\partial \theta}\,\hat{\mathbf{r}} \;-\; \frac{1}{r \sin\theta}\frac{\partial f}{\partial r}\,\hat{\boldsymbol{\theta}} \quad . \end{aligned}$$

The components of the magnetic field are thus specified as partial derivatives of the flux function. From a magnetic field satisfying a given symmetry condition, and $\nabla \cdot \mathbf{B} = 0$, it is always possible to compute the flux function $f(\xi, \eta)$. The flux function is said to "generate" the magnetic field.

The significance of the flux function derives from the fact that *it is constant along any field line*. The derivative of f along a field line $\mathbf{r}(s)$ defined by eq. (9.1) is

$$\frac{df}{ds} \;=\; \frac{d\mathbf{r}}{ds} \cdot \nabla f \;=\; \mathbf{B} \cdot \nabla f \;=\; (\nabla f \times \nabla \zeta) \cdot \nabla f \;=\; 0 \quad . \tag{9.3}$$

To make the last step one must recall that a scalar triple product $\mathbf{a} \times \mathbf{b} \cdot \mathbf{c}$ will vanish if any two elements are equal (or even co-linear). This shows that contours of the function $f(\xi, \eta)$ are field lines, so field lines can be traced by simply contouring the flux function.

Consider, for example, a simple magnetic dipole of dipole moment m. The magnetic field can be written using the vector potential

$$\mathbf{A}_{\rm dip}(r, \theta) \;=\; \frac{m \sin\theta}{r^2}\,\hat{\boldsymbol{\phi}} \;=\; \frac{m \sin^2\theta}{r}\,\nabla\phi \quad . \tag{9.4}$$

The final expression, using $\nabla\phi = \hat{\boldsymbol{\phi}}/r\sin\theta$, makes clear the flux function of a magnetic dipole

$$f_{\rm dip}(r,\theta) \;=\; \frac{m\sin^2\theta}{r} \quad . \tag{9.5}$$

Contours of this function, shown in fig. 9.1, are recognizable as the field lines from an ideal magnetic dipole. Since $f_{\rm dip}$ diverges as $r \to 0$, every single contour, and thus every field line, goes through the origin.

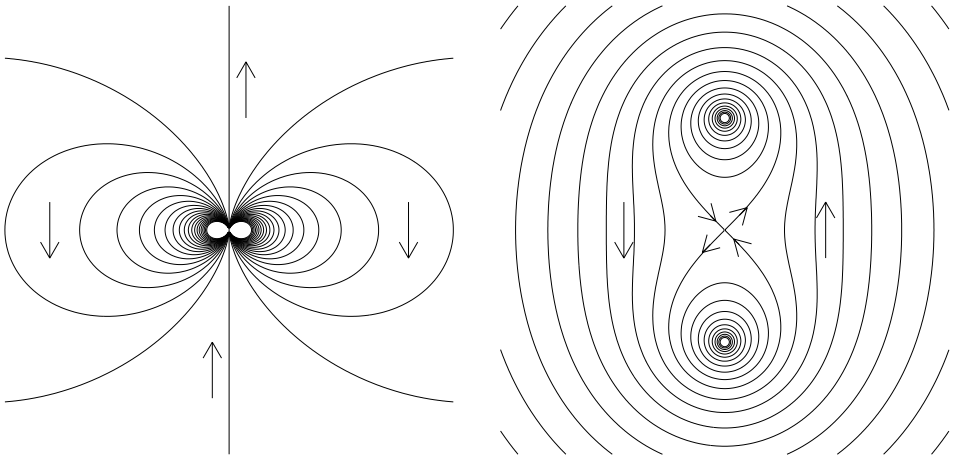

Figure 9.1: Examples of contours of flux functions tracing field lines. The left shows field lines from a point dipole from contours of $f_{\rm dip}(r,\theta)$, eq. (9.5). The right shows field lines from a pair of wires, from contours of $f(x,y)$, eq. (9.6). Both functions diverge at points within the picture, so contours are shown only below some finite upper limit, $f \le f_{\rm max}$; below this value the contour spacing is uniform. Arrows shows the direction of the magnetic field.

For a second example consider the field generated by two parallel wires carrying current $I > 0$, separated by $2d$. The flux function for this field

$$f(x,y) \;=\; -\frac{I}{c}\Big\{\ln[x^2+(y-d)^2] \;+\; \ln[x^2+(y+d)^2]\Big\} \quad , \tag{9.6}$$

shown in fig. 9.1, diverges ($f \to +\infty$) at the wires, $(x,y) = (0,\pm d)$. The saddle point at the origin is a null point since $\nabla f = 0 = \mathbf{B}$ there. The value of the flux function at this point, $f = f_\times = -(4I/c)\ln d$, defines two field lines that both begin and end at the null point. Contours above this level close about one of the two wires; contours below it enclose both wires. In all cases the field traverses the closed contour in the counter-clockwise sense, aseq. (9.1), with $\nabla\zeta = \hat{\mathbf{z}}$, dictates for any local maximum (including a divergence).

The field-line-density-shows-field-strength myth probably originated from considering only the simple cases of fields with Cartesian symmetry: $\mathbf{B}(x,y)$. Since ∇f and $\nabla\zeta$ are orthogonal, the magnitude of eq. (9.2) gives

$$|\mathbf{B}| \;=\; |\nabla f|\,|\nabla\zeta| \quad . \tag{9.7}$$

Contour levels differing by δf, will be separated in space by a distance

$$\Delta \simeq \frac{\delta f}{|\nabla f|} = \frac{|\nabla\zeta|\,\delta f}{|\mathbf{B}|} \quad . \tag{9.8}$$

The density of field lines at a point in the (ξ, η) plane is inversely proportional to distance between them and thus $\propto B/|\nabla\zeta|$. Provided $|\nabla\zeta|$ is a constant, as it is for Cartesian symmetry, the density is clearly proportional to field strength. For axisymmetric field, however, $1/|\nabla\zeta| = r\sin\theta$ and the density is proportional to field strength *times* the distance from the symmetry axis. Even for this very simple case field line density is not strictly proportional to field strength.

The ease of plotting field lines is certainly one reason for the popularity of two-dimensional MHD in apparent disregard for our three-dimensional universe. If we insist on working with fully three-dimensional fields (i.e. lacking any symmetry) there is no simple way to trace magnetic field lines. One must solve a set of three coupled ODEs, eq. (9.1) for example, beginning from one or more initial points. This procedure, called *field line tracing*, must be done — there is no short-cut in general.

9A.2 Flux tubes

A field line is a curve through space: it has zero cross section. A related construct is a *flux tube*, which consists of a bundle of field lines with finite cross section. A flux tube may be constructed by first defining a small surface $\mathcal{S}$ with a surface normal $\hat{\mathbf{n}}$. The field should pierce the surface in once sense so $\hat{\mathbf{n}}\cdot\mathbf{B} > 0$ everywhere on $\mathcal{S}$. The net flux crossing the surface is

$$\Phi_{\mathcal{S}} = \int_{\mathcal{S}} \mathbf{B}\cdot d\mathbf{a} > 0 \quad . \tag{9.9}$$

The flux tube is then constructed by using every point $\mathbf{r}_0 \in \mathcal{S}$ as an initial condition to solve eq. (9.1). The union of all resulting field lines is a volume $\mathcal{T}$ which is the flux tube. It follows naturally that the surface of the tube, $\partial\mathcal{T}$, consists of field lines initiated at every point on $\partial\mathcal{S}$. It can also be shown that no field lines cross the surface of the flux tube: $\mathbf{B}\cdot\hat{\mathbf{n}}$ on $\partial\mathcal{T}$.

The basic use for a flux tube is as a finite-width manifestation of a field line. To achive this objective $\mathcal{S}$ is taken to be small enough that all field lines initiating in it stay relatively close together. That means we expect the corss section of $\mathcal{T}$ to remain fairly small. Provided it remains in tact, it can readily be shown that any cross section will have the same flux $\Phi_{\mathcal{S}}$ as the initial cross section.

9B Frozen field lines

As already alluded to, the main intuitive power of magnetic field lines derives from the so-called *frozen-field-line* theorem. This holds that for a highly conductive conductive plasma the fluid and the magnetic field evolve together in such a way that *field lines and plasma remain frozen together*. An equivalent statement, which is easier to prove, is the following.

> **Frozen field lines:** In a perfectly conducting fluid, if two fluid elements lie on the same field line at one time they will lie on a common field line at all earlier and later times.

This statement holds only in the case of perfect electrical conduction, $\mathbf{E}' = 0$ or $Rm \to \infty$, also called *ideal MHD*. Since most astrophysical plasmas have astronomically large magnetic Reynold's numbers (albeit not infinite) the frozen field line theorem will be an extremely good approximation.

In the limit of perfect conduction, the induction eq. (MI), becomes

$$\frac{\partial \mathbf{B}}{\partial t} \;=\; \nabla \times (\mathbf{u} \times \mathbf{B}) \;=\; -\,(\mathbf{u}\cdot\nabla)\mathbf{B} \;+\; (\mathbf{B}\cdot\nabla)\mathbf{u} \;-\; \mathbf{B}\,(\nabla\cdot\mathbf{u}) \quad , \tag{9.10}$$

after a standard vector identity. This is the equation governing the evolution of $\mathbf{B}$, *regardless* of what goes into determining the flow velocity $\mathbf{u}(\mathbf{x}, t)$. The frozen-field-line theorem is a direct mathematical consequence of eq. (9.10) alone; it makes no assumptions about the nature of the plasma velocity field $\mathbf{u}$, or what forces may govern its evolution.

The theorem before us refers to two steps which involve solving ODEs: tracing a field line, and following fluid elements. The first step, as we have discussed above, involves solving the field line equation, for which we will use form (9.1"). This requires knowledge of the mass density, $\rho(\mathbf{x}, t)$, which will be assumed to evolve according to the continuity equation, (MC). The second step is to follow a fluid element along its trajectory $\mathbf{r}(t)$, which satisfies the ODE, eq. (1.5),

$$\frac{d\mathbf{r}}{dt} \;=\; \mathbf{u}[\mathbf{r}(t), t] \quad . \tag{9.11}$$

from an initial position, $\mathbf{r}(0) = \mathbf{r}_0$.

The frozen-flux-theorem refers to a process illustrated in fig. 9.2. The field line is traced, at time $t = t_i$, from a point, $\mathbf{r}_{a,i}$ to a second point, $\mathbf{r}_{b,i}$. This is done by solving eq. (9.1") with the magnetic field $\mathbf{B}(\mathbf{x}, t_i)$. Next the trajectories of these two fluid elements are followed, using eq. (9.11), to their new positions $\mathbf{r}_{a,f}$ and $\mathbf{r}_{b,f}$ at subsequent time $t_f > t_i$. The theorem itself makes the following assertion about these two points: the field line from the new magnetic $\mathbf{B}(\mathbf{x}, t_f)$ beginning at point $\mathbf{r}_{a,f}$ will cross through the point $\mathbf{r}_{b,f}$.

The above assertion is equivalent to the observation that the point $\mathbf{r}_{b,f}$ can be obtained from $\mathbf{r}_{a,i}$ in two different ways. We can trace the field line $\mathbf{r}_{a,i} \to \mathbf{r}_{b,i}$ first and then follow the trajectory $\mathbf{r}_{b,i} \to \mathbf{r}_{b,f}$. Or we can follow the trajectory $\mathbf{r}_{a,i} \to \mathbf{r}_{a,f}$ first *and then* trace the field line $\mathbf{r}_{a,f} \to \mathbf{r}_{b,f}$. Either order of operations will lead us to the exact same point $\mathbf{r}_{b,f}$. This is to say the operation of field-line-tracing and the operation of trajectory-following *commute* with one another.

The operations of field-line-tracing and trajectory-following are intimately related to the differential operators

$$\frac{d}{d\mu} \;=\; \left(\frac{\mathbf{B}}{\rho}\right)\cdot\nabla \quad , \quad \frac{D}{Dt} \;=\; \frac{\partial}{\partial t} \;+\; \mathbf{u}\cdot\nabla \quad , \tag{9.12}$$

which are said to "generate" them. The variations of a general function of space and time, $g(\mathbf{x}, t)$, along a field line or along a trajectory are given by the derivatives, $dg/d\mu$ and Dg/Dt respectively. Using the component of the position vector $g(\mathbf{x}, t) = x_j$, and the identity

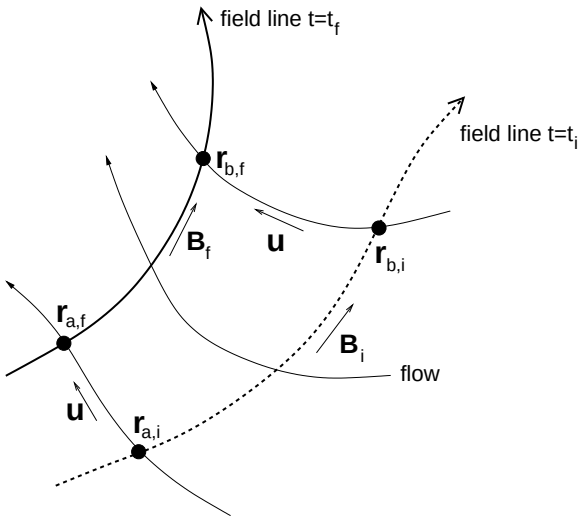

Figure 9.2: The processes of field-line-tracing and trajectory-following used to demonstrate the frozen-field-line theorem.

$\partial x_j/\partial x_i = \delta_{ij}$, returns the ODEs (9.1") and (9.11) defining field lines and trajectories. In brief, the field-line-following operator consists of repetitions of the fundamental operator $d/d\mu$, and trajectory-follow of repetitions of the operator D/Dt.

The intimate relation can be appreciated from how repeated differentiation, combined into a Taylor series, will propagate a function to a point further along the field line

$$g[\mathbf{r}(\mu_b),t] = g[\mathbf{r}(\mu_a),t] + (\mu_b-\mu_a)\left.\frac{dg}{d\mu}\right|_{\mu_a} + \tfrac{1}{2}(\mu_b-\mu_a)^2\left.\frac{d^2g}{d\mu^2}\right|_{\mu_a} + \cdots \quad . \tag{9.13}$$

Introducing $g(\mathbf{x},t) = x_j$ into this expression generates the space curve of the field line itself. A similar Taylor series, involving repeated use of the advective derivative, D/Dt, propagates information along a trajectory.

Without delving an further into abstract mathematical details,[2] the two operations, field-line-tracing and trajectory-following, will commute only if this is true of the derivatives which generate them. Their commutator is readily found

$$\begin{aligned}\left[\frac{D}{Dt},\frac{d}{d\mu}\right] &= \frac{D}{Dt}\frac{d}{d\mu} - \frac{d}{d\mu}\frac{D}{Dt} = \left\{\frac{D}{Dt}\left(\frac{\mathbf{B}}{\rho}\right) - \frac{d\mathbf{u}}{d\mu}\right\}\cdot\nabla \\ &= \frac{1}{\rho}\left\{\frac{D\mathbf{B}}{Dt} - \frac{\mathbf{B}}{\rho}\frac{D\rho}{Dt} - (\mathbf{B}\cdot\nabla)\mathbf{u}\right\}\cdot\nabla \\ &= \frac{1}{\rho}\left\{\frac{\partial\mathbf{B}}{\partial t} + (\mathbf{u}\cdot\nabla)\mathbf{B} + \mathbf{B}(\nabla\cdot\mathbf{u}) - (\mathbf{B}\cdot\nabla)\mathbf{u}\right\}\cdot\nabla \quad ,\end{aligned} \tag{9.14}$$

after using the Lagrangian form of the continuity equation, $D\rho/Dt = -\rho(\nabla\cdot\mathbf{u})$, from eq. (C-L). The commutator will vanish if the factor in curly braces vanishes. It will vanish identically provided the magnetic field satisfies the ideal induction equation, eq. (9.10).

[2] If you would prefer to do so, they fall under the theory of *continuous groups* or *Lie groups*.

In other words, the ideal induction equation ($\mathbf{E}' = 0$) assures the interchangeability of field-line-tracing and trajectory-following.

The two-fluid-element statement proven above can be generalized to hold for fluid elements along an entire field line. If we mark all the fluid elements along a field line at one time, $t = t_i$, and then follow the marked fluid elements as they flow, then they will always trace out a single field line. This field line and the marked fluid are the same. The entire field line is made of stuff and is therefore real, provided only that the magnetic field evolves according to eq. (9.10).

9B.1 Symmetric cases

In cases of symmetry, field lines can be quickly found using flux functions defined in eq. (9.2). Equating the partial time derivative of this definition to the ideal induction equation

$$\frac{\partial \mathbf{B}}{\partial t} \;=\; \nabla \times \left(\frac{\partial f}{\partial t} \nabla \zeta \right) \;=\; \nabla \times (\mathbf{u} \times \mathbf{B}) \quad , \tag{9.15}$$

provides a constraint on the evolution of the flux function. Ideal evolution of the magnetic field can be assured by equating the two expressions inside curls

$$\frac{\partial f}{\partial t} \nabla \zeta \;=\; \mathbf{u} \times \mathbf{B} \;=\; \mathbf{u} \times (\nabla f \times \nabla \zeta) \;=\; (\mathbf{u} \cdot \nabla \zeta) \nabla f - (\mathbf{u} \cdot \nabla f) \nabla \zeta \quad . \tag{9.16}$$

The component of this expression perpendicular to the symmetry direction, $\nabla\zeta$, requires $\mathbf{u} \cdot \nabla \zeta = 0$, at least where there is magnetic field – $\nabla f \neq 0$. In other words the field will only remain perpendicular to $\nabla\zeta$, as expression (9.2) demands, if the flow is also perpendicular.

Provided that condition is satisfied, ideal induction can be achieved if the flux function is advected with the flow

$$\frac{Df}{Dt} \;=\; \frac{\partial f}{\partial t} \;+\; \mathbf{u} \cdot \nabla f \;=\; 0 \quad . \tag{9.17}$$

In other words, each fluid elements carries with it the value of the flux function. Any two fluid elements which lie on the same field line must have the same value of f. Since these values are preserved as the fluid elements move, they will continue to lie on the same field line. Thus we see that the froze-field-line theorem is particularly straight forward in cases where flux functions exist.

9B.2 Frozen flux

There is a related concept, often conflated with frozen field lines, which is actually a weaker form of the theorem. It states that the flux inside any closed curve, $\mathcal{C}$, which is advected with the flow, will not be changed provided $\mathbf{E}' = 0$. This is a straightforward consequence of Faraday's law in integral form, provided proper account is taken of the moving path. Naturally the proper way to account for that motion is to use the electric field in the moving frame

$$\frac{\Phi_{\mathcal{C}}}{dt} \;=\; -\, c \oint_{\mathcal{C}} \mathbf{E}' \cdot d\mathbf{l} \;=\; 0 \quad . \tag{9.18}$$

The frozen flux theorem can be applied to a flux tube by setting $\mathcal{C} = \partial\mathcal{S}$, the perimeter of the cross section. This shows that if the cross section moves with the flow it will continue

to enclose the same flux. This will be true of every cross section of the flux tube. There still remains, however, a gap between enclosing the same net flux and enclosing the same *field lines*. It is in this sense that frozen flux is weaker than frozen field lines. Once the latter is established it is clear that each field line composing the flux tube will move with the plasma, so that means the entire bundle (the flux tube) must do so too.

9B.3 Who's in charge?

The discussion of frozen field lines has been intentionally silent on the question of whether field line moves the plasma or the plasma moves the field line. So far we have shown only that they move *together*. In fact, it is often dangerous to anthropomorphize fluids, assigning to one property the role of leader. In the present situation, however, there are two limits where doing so provides legitimate insight. These are cases of very large and very small plasma β.

We have already discussed the fact that very large β ($\beta \gg 1$) means the magnetic field exerts no significant force on the fluid. The fluid therefore undergoes the same dynamics as a neutral fluid — simple hydrodynamics. The velocity field $\mathbf{u}(\mathbf{x}, t)$ can thus be found without knowing anything about the magnetic field. At this point, provided the plasma is highly conductive, we can "paint in" field lines and follow them as if they were lines of tracer dye. The field lines are said to be *passively* advected by the flow.

The solar interior is a region of extremely large β ($\beta \sim 10^5$) where flows are driven by thermal convection and Coriolis effects. These generate the well-known differential rotation of the Sun as well as other internal motions. The magnetic field will have no effect on this motion; it will instead be passively moved around by those motions. This leads to the standard introductory picture of the solar dynamo, where differential rotation deflects North-South (poloidal) field lines into the East-West (toroidal) direction.

In the opposite limit, $\beta \ll 1$, magnetic forces are far stronger than pressure forces and the latter can be neglected. If the flow velocities are also small — small enough that $M^2\beta \ll 1$ — then inertia will also be negligible in comparison to magnetic forces. This means the magnetic field will seek a force-free equilibrium. $(\nabla \times \mathbf{B}) \times \mathbf{B} = 0$, The equilibrium might change slowly in response to boundaries and the plasma is forced to go wherever the field lines choose to go. In this limit the magnetic field lines move the plasma. A commonly invoked picture is that of small beads on a stiff wire. This picture makes clear the frozen-field-line theorem permits plasma to slide freely along the field line — it lives in a one dimensional world called a flux tube.

Low β plasmas include the solar corona ($\beta \sim 10^{-3}$), the solar wind, the inner portion of the Earth's magnetosphere, and the rarified coronae above astrophysical disks. In all of these cases the low-density, low-pressure plasma is moved in response to motion of magnetic field lines. For example, the magnetospheric plasma must co-rotate with the Earth because the field lines do so. Here we must be careful not to imagine a field line extending all the way down to the Earth's surface. The bulk of the atmosphere is *not* a good conductor and field lines are *not* real objects there. There is an upper layer called the *ionosphere* which moves with the rest of the atmosphere and has a rather high conductivity. Field lines are real there and will move as it moves — rotating once per day. The magnetosphere above it is far lighter ($\beta \ll 1$) and is forced to move with the field lines.

9C Magnetic pressure & magnetic tension

The previous section showed how magnetic field lines provide some intuitive insight into the induction equation, eq. (MI), at least in the ideal limit $Rm \gg 1$. The other significant modification MHD makes to the hydrodynamics is the Lorentz force (actually a force density)

$$\mathbf{f}_M \;=\; \frac{1}{4\pi}(\nabla\times\mathbf{B})\times\mathbf{B} \;=\; \frac{1}{4\pi}(\mathbf{B}\cdot\nabla)\mathbf{B} \;-\; \nabla\left(\frac{|\mathbf{B}|^2}{8\pi}\right) \quad , \tag{9.19}$$

after using a standard vector calculus identity. The term on the very right has a form identical to the force density from a pressure gradient. It thus appears that part of the Lorentz force takes the form of *magnetic pressure*

$$p_{\rm mag} \;=\; \frac{B^2}{8\pi} \quad . \tag{9.20}$$

This is quite natural since magnetic field has an energy density, $B^2/8\pi$, and a gas has an energy density $p/(\gamma-1)$. Pushing the analogy still further tempts us to assign $\gamma = 2$ to magnetic field; this is useful in certain cases, but as we will soon see, magnetic field pressure behaves in ways fundamentally different from a real gas.

It is also worth recognizing that the dimensionless parameter, β, is defined as the ratio of plasma pressure to magnetic pressure

$$\beta \;=\; \frac{p}{p_{\rm mag.}} \;=\; \frac{p}{B^2/8\pi} \quad . \tag{9.21}$$

To develop a intuition about the other term in eq. (9.19), we introduce a unit vector directed along the magnetic field at each point in space

$$\hat{\mathbf{b}}(\mathbf{x}) \;=\; \frac{\mathbf{B}(\mathbf{x})}{|\mathbf{B}(\mathbf{x})|} \quad . \tag{9.22}$$

The full vector may now be written $\mathbf{B} = B\hat{\mathbf{b}}$ and introduced into eq. (9.19) to give

$$\begin{aligned}\mathbf{f}_M \;&=\; \frac{B^2}{4\pi}(\hat{\mathbf{b}}\cdot\nabla)\hat{\mathbf{b}} \;+\; \frac{B}{4\pi}\hat{\mathbf{b}}(\hat{\mathbf{b}}\cdot\nabla)B \;-\; \nabla\left(\frac{B^2}{8\pi}\right) \\ &=\; \frac{B^2}{4\pi}(\hat{\mathbf{b}}\cdot\nabla)\hat{\mathbf{b}} \;-\; \left(\underline{\underline{\mathrm{I}}} - \hat{\mathbf{b}}\hat{\mathbf{b}}\right)\cdot\nabla\left(\frac{B^2}{8\pi}\right) \quad ,\end{aligned} \tag{9.23}$$

after combining the two terms proportional to ∇B^2. The pressure term is modified in this new version of the force: it is multiplied by the projection matrix, $\underline{\underline{\mathrm{I}}} = \hat{\mathbf{b}}\hat{\mathbf{b}}$, which eliminates the parallel component of the magnetic pressure gradient. This makes sense since the original Lorentz force, $\mathbf{J}\times\mathbf{B}$, was strictly perpendicular to the magnetic field. Thus regardless of how the magnetic pressure varies, there will be no force along the magnetic field. This differs from gas pressure which produces a force $-\nabla p$.

The first term on the rhs of eq. (9.23) can be interpreted using expression (9.1') for a magnetic field line $\mathbf{r}(\ell)$. The forces density is

$$\mathbf{f}_{\rm mt} \;=\; \frac{B^2}{4\pi}(\hat{\mathbf{b}}\cdot\nabla)\hat{\mathbf{b}} \;=\; \frac{B^2}{4\pi}\left(\frac{d\mathbf{r}}{d\ell}\cdot\nabla\right)\hat{\mathbf{b}} \;=\; \frac{B^2}{4\pi}\frac{d\hat{\mathbf{b}}}{d\ell} \quad . \tag{9.24}$$

Naturally, this force is also perpendicular to the magnetic field because the vector

$$\mathbf{K} \;=\; \frac{d\hat{\mathbf{b}}}{d\ell} \;=\; \frac{d^2\mathbf{r}}{d\ell^2} \quad , \tag{9.25}$$

is perpendicular:

$$\hat{\mathbf{b}}\cdot\mathbf{K} \;=\; \hat{\mathbf{b}}\cdot\frac{d\hat{\mathbf{b}}}{d\ell} \;=\; \frac{1}{2}\frac{d|\hat{\mathbf{b}}|^2}{d\ell} \;=\; 0 \quad .$$

The vector $\mathbf{K}$, i.e. the second derivative, is known as the *curvature vector* of the space curve $\mathbf{r}(\ell)$. There is a circle which osculates $\mathbf{r}(\ell)$ at a given point. $\mathbf{K}$ points toward the center of that circle and $|\mathbf{K}|$ is the inverse of its radius — called the radius of curvature R_c. (The radius of curvature of a straight line is $R_c \to \infty$, while a kink or corner has $R_c = 0$.) This geometrical interpretation makes clear that $\mathbf{K}$ will be perpendicular to the curve's tangent vector $\hat{\mathbf{b}}$.

The force $\mathbf{f}_{\rm mt}$, in eq. (9.24), is analogous to the force in a string under tension and it is therefore called the *magnetic tension*. Recall that a string under tension T, has a force $T\hat{\mathbf{b}}$ along it. If the string is deformed the tension creates a force-per-length $T\mathbf{K}$ in the direction of curvature. Since $\mathbf{f}_{\rm mt}$ is a force per volume, our analogy is that a magnetic field line is like a string with tension-per-area $B^2/4\pi$. When a field line is bent, there is a tension force directed toward the center of curvature, with magnitude, $B^2/4\pi R_c$.

The complete picture we have derived is that magnetic field behaves like a gas with pressure $B^2/8\pi$ in directions *perpendicular* to itself, and like strings under tension (per area) $B^2/4\pi$ parallel to itself. Both effects produce forces perpendicular to the field lines, and often in opposing directions. The field lines of the magnetic dipole, shown on the left of fig. 9.1, are curved toward the origin. This inward magnetic tension force is, in this case, exactly balanced by an outward pressure gradient ($B^2 \sim r^{-6}$). We could have deduced the balance more easily by noting that $\mathbf{J} = 0 = \nabla\times\mathbf{B}$ in this particular field. Still, it is often useful to think in terms of magnetic pressure and magnetic tension.

The conventional derivation above can be achieved more quickly using the Maxwell stress tensor. Equation (9.19) can be written as the divergence of a tensor

$$\mathbf{f}_M \;=\; \frac{1}{4\pi}(\mathbf{B}\cdot\nabla)\mathbf{B} \;-\; \nabla\left(\frac{|\mathbf{B}|^2}{8\pi}\right) \;=\; \nabla\cdot\left(\frac{1}{4\pi}\mathbf{BB} \;-\; \frac{B^2}{8\pi}\,\hat{I}\right) \quad , \tag{9.26}$$

after using $\nabla\cdot\mathbf{B} = 0$ and using the identity tensor to write the gradient as a divergence.[3] The tensor

$$\underline{\underline{\mathrm{T}}} \;=\; \frac{1}{4\pi}\mathbf{BB} \;-\; \frac{B^2}{8\pi}\underline{\underline{\mathrm{I}}} \;=\; \frac{B^2}{8\pi}\hat{\mathbf{b}}\hat{\mathbf{b}} \;-\; (\underline{\underline{\mathrm{I}}} - \hat{\mathbf{b}}\hat{\mathbf{b}})\frac{B^2}{8\pi} \quad , \tag{9.27}$$

is the Maxwell stress tensor with no electric field.[4]

Stress tensors occur throughout fluid dynamics where forces act only on the surface of a fluid element. This is famously the case with the viscous stress sensor, eq. (1.35), but is also true of pressure: $\underline{\underline{\sigma}} = -p\underline{\underline{\mathrm{I}}}$. The surface element $da\,\hat{\mathbf{n}}$ experiences a force $\underline{\underline{\sigma}}\cdot\hat{\mathbf{n}}\,da$, where $\hat{\mathbf{n}}$ is the *outward* surface normal. When the force is directed inward, as it is for pressure,

[3] You may recall using this trick before in deriving momentum conservation in hydrodynamics. This is no coincidence — both terms come form pressures.

[4] The absence of electric forces are due to the fact that $E/B \sim v^2/c^2 \ll 1$ in our non-relativistic plasma.

$\underline{\underline{\sigma}} \cdot \hat{\mathbf{n}} = -p\hat{\mathbf{n}}$, it is *compressive*. Outward directed forces are therefore *tensile* — pulling outward on the surface element. Forces tangent to the surface are called *shear* forces.

A force at a surface, i.e. a stress tensor, is equivalent to a flux of momentum density. For the Maxwell stress tensor this momentum is carried by the electromagnetic field. When it crosses the surface it carries momentum into the fluid element, thereby exerting a force. Thus, the Lorentz force is just an expression of how electro-magnetic momentum flows into and out of a fluid element. It is true, though not at all obvious from expression eq. (**??**), that unless there is a current density, $\mathbf{J} \neq 0$, within the fluid element the electromagnetic flux into and out of a fluid element will exactly balance.[5] It is evident from expression (9.27), that the Maxwell stress tensor is compressive perpendicular to $\hat{\mathbf{b}}$ (the projection tensor appears once more) and tensile parallel to $\hat{\mathbf{b}}$. This is the origin of magnetic pressure and magnetic tension, respectively.

9D Example

A very simple example can illustrate the ideas described above. For the example we assume the plasma velocity has the form of a simple sinusoidal shear

$$\mathbf{u} = v_0 \sin(2\pi y/L)\hat{\mathbf{x}} \quad , \tag{9.28}$$

where L is the shearing length. This will act on a magnetic field lying in the x–y plane, $\mathbf{B} = B_x\hat{\mathbf{x}} + B_y\hat{\mathbf{y}}$ through the ideal induction equation, eq. (9.10),

$$\begin{aligned} \frac{\partial B_x}{\partial t} &= -v_0 \sin(2\pi y/L)\frac{\partial B_x}{\partial x} + \frac{2\pi v_0}{L}\cos(2\pi y/L)\, B_y \\ \frac{\partial B_y}{\partial t} &= -v_0 \sin(2\pi y/L)\frac{\partial B_y}{\partial x} \end{aligned}$$

If the initial configuration is a uniform vertical field $\mathbf{B}(\mathbf{x}, 0) = B_0\hat{\mathbf{y}}$, then the solution to the induction equation is

$$\mathbf{B}(x, y, t) = B_0 \frac{2\pi v_0 t}{L} \cos(2\pi y/L)\, \hat{\mathbf{x}} + B_0\hat{\mathbf{y}} \quad . \tag{9.29}$$

We expect solution eq. (9.29) to be equivalent to passive threads dragged along with the flow. As stated previously, it is possible to parameterize a field line with any monotonic function of position. Since B_y is always positive the y coordinate is just such a parameter and we can write field lines $x(y)$. The y-parameterized field line satisfies the equation

$$\frac{dx}{dy} = \frac{B_x[x(y), y]}{B_y[x(y), y]} = \frac{2\pi v_0 t}{L}\cos(2\pi y/L) \quad . \tag{9.30}$$

Beginning at $x(0) = x_0$ this equation has the solution

$$x(y) = v_0 t\, \sin(2\pi y/L) + x_0 \quad . \tag{9.31}$$

[5] This is how even the considerable Maxwell stresses from the magnetic field around us does not actually exert any force on us.

We can use eq. (9.28) to write this as

$$\mathbf{r}(y) \;=\; \mathbf{u}(y)\,t \;+\; \mathbf{r}_0 \quad , \tag{9.32}$$

which is the position of a particle moving at constant velocity for time t. This is exactly what the frozen-in-flux theorem said would happen: each bit of the field line moves with the fluid.

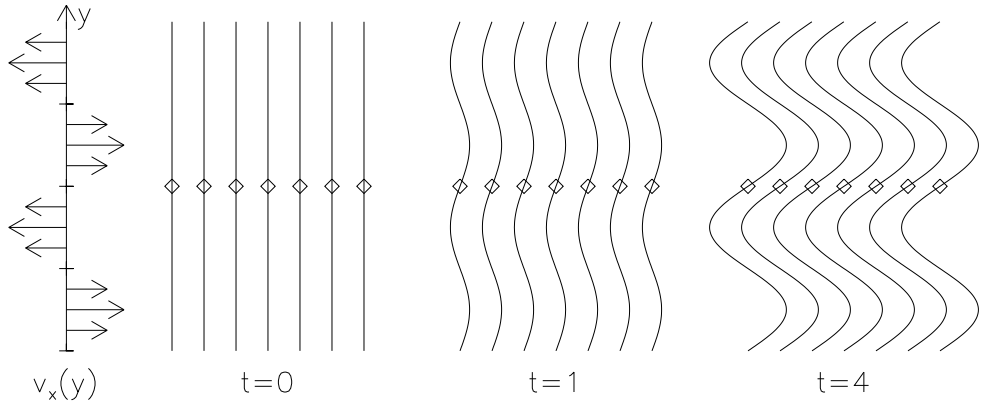

Figure 9.3: The shear flow (left) and some field lines of the evolving magnetic field. The initial points for these field lines (diamonds) lie along $y = 0$ where the velocity vanishes.

The direction of the current resulting from eq. ((9.29) can be directly found

$$\nabla \times \mathbf{B} \;=\; \nabla \times \Big[\, B_0 v_0 t k \cos(ky)\hat{\mathbf{x}} \Big] \;=\; B_0 v_0 t k^2 \sin(ky)\,\hat{\mathbf{z}} \quad , \tag{9.33}$$

after introducing $k = 2\pi/L$. The current is therefore directed normal to the x–y plane in which the bent field lines lie. It is largest where the field line curvature is greatest, $y = \pm n\pi/k = \pm nL/2$, for odd n. These are also the points of greatest Lorentz force-density

$$\frac{1}{4\pi}(\nabla \times \mathbf{B}) \times \mathbf{B} \;=\; \frac{1}{4\pi}\,(B_0 v_0 t)^2 k^3 \sin(ky)\cos(ky)\hat{\mathbf{y}} \;-\; \frac{1}{4\pi}\,B_0^2 k^2 v_0 t \sin(ky)\hat{\mathbf{x}} \quad . \tag{9.34}$$

The first term is a vertical force directed away from regions of stronger magnetic field: $ky = m\pi$ — the is the pressure force. The second term acts horizontally, and is strongest where the field is most curved $ky = \pm\pi/2,\ \pm 3\pi/2$. Referring to fig. 9.3 we see that the force opposes the bending of the field lines — it is leftward at $ky = \pi/2$, rightward at $ky = -\pi/2$ etc. This is the tension force. To see this relation mathematically we can re-write the above equation as

$$\begin{aligned}
\frac{1}{4\pi}(\nabla \times \mathbf{B}) \times \mathbf{B} \;&=\; -\frac{1}{8\pi}\hat{\mathbf{y}}\frac{\partial}{\partial y}\left[B_0 v_0 k t \cos(ky)\right]^2 \;+\; \frac{1}{4\pi}\,B_0 \frac{\partial}{\partial y}\left[B_0 k v_0 t \cos(ky)\hat{\mathbf{x}}\right] \\
&=\; -\frac{1}{8\pi}\nabla|\mathbf{B}|^2 \;+\; \frac{1}{4\pi}\,(\mathbf{B}\cdot\nabla)\mathbf{B}
\end{aligned}$$

The problem so far has used a prescribed flow velocity, eq. (9.28). Whatever maintains this flow (not identified) must do so against an ever-increasing force. Note that (9.34)

has a horizontal component exactly opposing $\mathbf{u}$ and growing linearly in time. To work self-consistently we use eq. (9.31) to replace $v_0 t \sin(ky) = x(y,t) - x_0$ and discard terms proportional to $(x - x_0)^2$. This leaves only the tension force

$$\frac{1}{4\pi}(\nabla \times \mathbf{B}) \times \mathbf{B} \simeq -\frac{B_0^2 k^2}{4\pi} [\, x(y,t) - x_0 \,] \hat{\mathbf{x}} = \frac{B_0^2}{4\pi} \frac{\partial^2 x}{\partial y^2} \hat{\mathbf{x}} \tag{9.35}$$

The horizontal velocity is $\mathbf{u} = \hat{\mathbf{x}} \partial x / \partial t$ and the horizontal momentum equation becomes

$$\rho \frac{\partial^2 x}{\partial t^2} = \frac{B_0^2}{4\pi} \frac{\partial^2 x}{\partial y^2} \quad , \tag{9.36}$$

which is the wave equation. This has the periodic solution

$$x(y,t) - x_0 = A \sin(ky) \sin(k v_{\rm A} t) \quad , \tag{9.37}$$

where

$$v_{\rm A} = \frac{B_0}{\sqrt{4\pi\rho}} \quad , \tag{9.38}$$

is called the Alfvén speed. This is a standing wave of wavelength $2\pi/k = L$, of a kind called an *Alfveń wave*.

This scenario is similar to every other oscillation or wave in Physics: motion leads to a restoring force, which reverses the motion, overshoots and leads to the opposite restoring force. Here the shear flow drags the field lines as depicted in fig. 9.3, according to the frozen-in-flux theorem. The distorted field produces a restoring force, magnetic tension, which opposes the shear flow, eventually reversing it. The reversed flow then undoes the distortion (fig. 9.3 form right to left), overshooting and creating the opposite distortion. Thus the Alfvén wave is one in which the restoring force is magnetic tension. We learn in the next section of other modes, called magnetosonic waves, in which magnetosonic pressure serves as a restoring force.

Chapter 10

Magnetostatic equilibria

Equilibria solutions of the MHD equations, (M-C)—(M-I), which are static ($\mathbf{u} = 0$) must satisfy the single equation

$$-\nabla p \ + \ \frac{1}{4\pi}(\nabla \times \mathbf{B}) \times \mathbf{B} \ + \ \rho \mathbf{g} \ = \ 0 \ , \tag{10.1}$$

where we have included gravity. In the case $\beta \gg 1$ the magnetic contribution (the Lorentz force) can be neglected, leaving the standard hydrostatic equations. We have already discussed these at length so we will consider the opposite limit, $\beta \ll 1$ where the magnetic field is dominant. Provided that the solutions are $L \ll H_p/\beta$, where H_p is the pressure scale height, we han drop the gravitational term as well as the pressure term leaving the equation

$$(\nabla \times \mathbf{B}) \times \mathbf{B} \ = \ 0 \ . \tag{10.2}$$

Equation (10.2) is a constraint on the magnetic field alone. A field that satisfies this condition is known as a *force-free field*.[1] The condition requires that $\mathbf{B}(\mathbf{x})$ be a vector field which is everywhere parallel to its own curl: $\nabla \times \mathbf{B} = \alpha \mathbf{B}$. The constant of proportionality, α, can be a function of space in general. There is a condition related to maintaining the requirement $\nabla \cdot \mathbf{B} = 0$. *To wit*

$$\nabla \cdot (\nabla \times \mathbf{B}) \ = \ 0 \ = \ \nabla \cdot (\alpha \mathbf{B}) \ = \ \mathbf{B} \cdot \nabla \alpha \ + \ \alpha \underbrace{\nabla \cdot \mathbf{B}}_{=0} \ .$$

A force-free field must therefore satisfy the set of equatons

$$\nabla \times \mathbf{B} \ = \ \alpha \mathbf{B} \tag{10.3}$$

$$\mathbf{B} \cdot \nabla \alpha \ = \ 0 \tag{10.4}$$

$$\nabla \cdot \mathbf{B} \ = \ 0 \ . \tag{10.5}$$

These are equations for the unknown fields, $\mathbf{B}(\mathbf{x})$ and $\alpha(\mathbf{x})$, and are non-linear in these. Solutions are known as *non-linear force-free fields* (NLFFs). Since they are non-linear in the unknowns they are extremely difficult to solve. Several assumptions are typically made to make them more solvable we cover these below.

[1] All equilibria are by definition force-free. The term force-free field refers to the magnetic field which is in equilibrium by itself.

10A Potential fields

The most common assumption is that there is no current within the corona: $\nabla \times \mathbf{B} = 0$. When coupled with the requirement that $\nabla \cdot \mathbf{B} = 0$ the magnetic field satisfies the same conditions associated with an electrostatic filed in vacuum. The same methods can thus be used to solve for the field. Because $\nabla \times \mathbf{B} = 0$ we can express the magnetic field using a scalar potential, $\chi(\mathbf{x})$,

$$\mathbf{B}(\mathbf{x}) \;=\; -\nabla\chi \quad , \tag{10.6}$$

motivating the term *potential field*. In order to satisfy the second, equation the scalar potential must satisfy Laplace's equation,

$$\nabla \cdot \mathbf{B} \;=\; -\nabla^2\chi \;=\; 0 \quad . \tag{10.7}$$

A major attraction of potential fields is the many well-known methods of solving Laplace's equation. Given a vertical field along a plane the potential can be found using a Green's function.

$$\chi(x,y,z) \;=\; \frac{1}{2\pi}\int \frac{B_z(x',y',0)\,dx'\,dy'}{\sqrt{(x-x')^2+(y-y')^2+z^2}} \quad . \tag{10.8}$$

When the vertical field vanishes outside of a small region centered at the origin, a case approximating a sunspot, the potential field away from that region

$$\mathbf{B}(\mathbf{x}) \;\simeq\; \frac{\Phi}{2\pi}\frac{\mathbf{x}}{|\mathbf{x}|^3} \quad , \quad \Phi \;=\; \int B_z(x,y,0)\,dx\,dy \quad , \tag{10.9}$$

analogous to the electric field from a charge concentration. In this sense sunspots appear to the space above the solar surface like point magnetic charges,[2] at least when that space is assumed to be current free.

The global magnetic field of the Sun is often approximated using a potential field, this time outside the solar surface. The influence of plasma pressure and solar wind flow cannot be neglected to all radii: these forces are, after all, responsible for opening to let out the solar wind. The simplest means of accommodating this complex effect is to designate a spherical surface, known as the *source surface*, at $r = R_{ss}$, where the field is required to be purely radial

$$B_\theta \;=\; -\frac{1}{r}\left.\frac{\partial\chi}{\partial\theta}\right|_{r=R_{ss}} \;=\; 0 \quad , \quad B_\phi \;=\; -\frac{1}{r\sin\theta}\left.\frac{\partial\chi}{\partial\phi}\right|_{r=R_{ss}} \;=\; 0 \quad . \tag{10.10}$$

This can be assured by satisfying the homogeneous Dirichlet condition,

$$\chi(R_{ss},\theta,\phi) \;=\; 0 \quad . \tag{10.11}$$

The solution to Laplace's equation can be written using spherical harmonics

$$\chi(r,\theta,\phi) \;=\; \sum_{\ell=1}^{\infty}\sum_{m=-\ell}^{m=\ell}\left(\frac{R_{ss}^{2\ell+1}}{r^{\ell+1}} \;-\; r^\ell\right) Y_\ell^m(\theta,\phi)\,(g_{\ell m}+ih_{\ell m}) \quad , \tag{10.12}$$

[2] It is actually the end of a long tube of magnetic flux confined by high pressure beneath the solar surface. This makes it mathematically similar to a model of a magnetic monopole proposed by Dirac.

where $g_{\ell m}$ and $h_{\ell m}$ are real coefficients. The Neumann condition at the solar surface, $r = R_\odot$,

$$B_r(R_\odot,\theta,\phi) \;=\; \sum_{\ell,m} R_\odot^{\ell-1}\left[(\ell+1)\frac{R_{ss}^{2\ell+1}}{R_\odot^{2\ell+1}} \;+\; \ell\right] Y_\ell^m(\theta,\phi)\,(g_{\ell m}+ih_{\ell m}) \quad , \tag{10.13}$$

can be used to determine the coefficients given a map of the radial field at all points of the surface at one time. Such a solution is known as a *potential field source-surface* (PFSS) model.

An axisymmetric potential dipole, $\ell = 1$ and $m = 0$, is generated by the flux function

$$f(r,\theta) \;=\; \frac{\Phi_{\rm open}}{6\pi}\left(\frac{2R_{ss}}{r}+\frac{r^2}{R_{ss}^2}\right)\sin^2\theta \quad , \tag{10.14}$$

whose contours are shown in fig. 10.1. The magnetic field

$$\begin{aligned}\mathbf{B} \;&=\; \nabla f\times\nabla\phi \;=\; \frac{1}{r^2\sin\theta}\frac{\partial f}{\partial\theta}\hat{\mathbf{r}} \;-\; \frac{1}{r\sin\theta}\frac{\partial f}{\partial r}\hat{\boldsymbol{\theta}} \\ &=\; \frac{\Phi_{\rm open}}{3\pi\,R_{ss}^2}\left[\hat{\mathbf{r}}\left(\frac{2R_{ss}^3}{r^3}+1\right)\cos\theta \;+\; \hat{\boldsymbol{\theta}}\left(\frac{R_{ss}^3}{r^3}-1\right)\sin\theta\right] \quad ,\end{aligned}$$

clearly satisfies condition (10.10). That it is a potential field, $\nabla\times\mathbf{B} = 0$, can be verified by taking the curl of the expression. Integrating the radial field at the source surface, $r = R_{ss}$, over either hemisphere yields the total open magnetic flux, $\Phi_{\rm open}$.

Field lines in a PFSS equilibrium are closed if both ends are rooted in the surface, $r = R_\odot$, or open if they intersect the source surface $r = R_{ss}$; the shaded region in fig. 10.1 consists of open field lines. The boundary between these classes of field line is a surface called a *separatrix*. In the dipole field the separatrix connects to the circle of magnetic null points ($B_r = B_\theta = B_\phi = 0$) along the equator of the source surface ($r = R_{ss}$ and $\theta = \pi/2$). This surface is defined by the level set

$$f(r,\theta) \;=\; f_{\rm sx} \;=\; f(R_{ss},0) \;=\; \frac{\Phi_{\rm open}}{2\pi} \quad .$$

This connects to the solar surface $R_\odot$ at latitude $\theta_{\rm sx}$

$$\sin^2\theta_{\rm sx} \;=\; \frac{3}{2R_{ss}/R_\odot + R_\odot^2/R_{ss}^2} \;=\; \frac{3R_\odot}{2R_{ss}}\,\frac{1}{1\;+\;R_\odot^3/2R_{ss}^3} \quad . \tag{10.15}$$

The case $R_{ss} = 2.5R_\odot$ has a separatrix at $\theta_{\rm sx} = 49.7°$. The open magnetic field lines, anchored to polar angles, $\theta < \theta_{\rm sx}$ and $\theta > \pi - \theta_{\rm sx}$ are the source of the entire solar wind in this simple configurations. These polar caps are referred to as *coronal holes*, and shown with darker shading in fig. 10.1.

10B Constant-α fields

The next level of sophistication is the constant-α field. As the name suggests, α is taken to be spatially uniform — it is a parameter. This choice satisfies eq. (10.4) immediately since $\nabla\alpha = 0$. We then take a curl of eq. (10.3) to get

$$\nabla\times(\nabla\times\mathbf{B}) \;=\; -\,\nabla^2\mathbf{B} \;=\; \alpha\nabla\times\mathbf{B} \;=\; \alpha^2\mathbf{B} \quad . \tag{10.16}$$

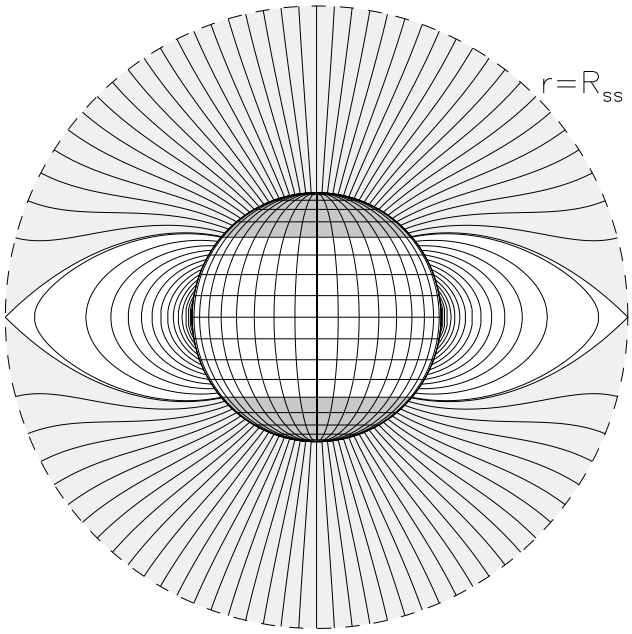

Figure 10.1: Field lines from the dipole PFSS field generated by eq. (10.14), for $R_{SS} = 2.5R_\odot$. Field lines are initialized at polar angles spaced evenly, 2.5° apart. The shaded region contains open field lines — they connect to the source surface. The boundary between open and closed field lines, the separatrix, is anchored to polar angle $\theta_{\rm sx} = 50°$ (i.e. $\pm 40°$ latitude). The regions poleward of these boundaries, the *coronal holes*, are shaded more darkly.

The Laplacian of a vector is shorthand for Laplacians of each *Cartesian* component

$$\nabla^2 B_x = -\alpha^2 B_x \tag{10.17}$$
$$\nabla^2 B_y = -\alpha^2 B_y \tag{10.18}$$
$$\nabla^2 B_z = -\alpha^2 B_z \ ; \tag{10.19}$$

each of these is a *Helmholtz* equation. They are linear in the unknown, so a constant-α field is also known as a *linear force-free field*.

Some feel for constant-α field comes from a solution over a periodic $L \times L$ box in the $z = 0$ plane. The vertical field at the lower boundary can be expressed as a Fourier series

$$B_z(x, y, 0) = \sum_{\mathbf{k}} \hat{b}_{\mathbf{k}} \, e^{i\mathbf{k}\cdot\mathbf{x}_\perp} \ , \tag{10.20}$$

where the wave vectors

$$\mathbf{k} = \frac{2\pi}{L} m \, \hat{\mathbf{x}} + \frac{2\pi}{L} n \, \hat{\mathbf{y}} \quad , \quad m, n \ \text{ integers} \ , \tag{10.21}$$

form a grid in the (x, y) plane and $\mathbf{x}_\perp = x\hat{\mathbf{x}} + y\hat{\mathbf{y}}$ is the position in the $z = 0$ plane. The coefficients of the expansion are

$$\hat{b}_{\mathbf{k}} = \frac{1}{L^2} \int_0^L \int_0^L B_z(x, y, 0) \, e^{-i\mathbf{k}\cdot\mathbf{x}_\perp} \, dx \, dy \ . \tag{10.22}$$

The most reasonable solution to eq. (10.19) is

$$B_z(x, y, z) = \sum_{\mathbf{k}} \hat{b}_{\mathbf{k}} \, \exp\Big[i\mathbf{k}\cdot\mathbf{x}_\perp - z\sqrt{|\mathbf{k}|^2 - \alpha^2} \Big] \ , \tag{10.23}$$

since it will decay with height provided the radical is real: i.e. $|\alpha| < |\mathbf{k}|$. If this does not happen, because there is some wave-number whose magnitude is *smaller* than $|\alpha|$, then the magnetic field strength will not decrease with height. As a consequence the magnetic energy density $|\mathbf{B}|^2/8\pi$ will not decrease and if the box extends infinitely high the magnetic energy within it will be infinite. This is an unphysical situation.

The grid of wave numbers in eq. (10.21), includes $\mathbf{k} = 0$ which will be smaller than any possible values of $|\alpha|$ except if $\alpha = 0$ — i.e. a potential field. The coefficient of this term

$$\hat{b}_{\mathbf{0}} = \frac{1}{L^2} \int_0^L \int_0^L B_z(x, y, 0) \, dx \, dy = \frac{\Phi}{L^2} \ , \tag{10.24}$$

will vanish provided the $L \times L$ box contains *no net flux*. This then is a requirement for realizable linear force free field. The next largest wave number is $|\mathbf{k}| = 2\pi/L$ from either $(m, n) = (\pm 1, 0)$ or $(m, n) = (0, \pm 1)$. It is unlikely that the coefficients of *both* of these will vanish so a linear force-free field is realistic only when the parameter falls in the range

$$|\alpha| \leq \frac{2\pi}{L} \ . \tag{10.25}$$

This is possible within the finite box but taking $L \to \infty$ to obtain a solution in an unbounded domain leaves only a potential field, $\alpha = 0$. While potential fields are possible in unbounded

domains, linear force-free fields are possible only in bounded domains, and then only for sufficiently small twist $|\alpha|$.

For any value of α within the permitted range there is a a unique field for any flux-balanced (i.e. $\Phi = 0$) distribution of vertical field, $B_z(x, y, 0)$. The horizontal components of this field

$$\mathbf{B}_\perp(x,y,z) \;=\; \sum_{\mathbf{k}} \frac{i\hat{b}_{\mathbf{k}}}{k^2}\left(\alpha\mathbf{k}\times\hat{\mathbf{z}} - \mathbf{k}\sqrt{k^2-\alpha^2}\right)\exp\Big[\,i\mathbf{k}\cdot\mathbf{x}_\perp - z\sqrt{|\mathbf{k}|^2-\alpha^2}\,\Big] \quad . \tag{10.26}$$

are determined entirely from the vertical field at the boundary. Even at the $z = 0$ surface the horizontal components are determined from the vertical component.

10C Equilibria with symmetry: the Grad-Shafranov equation

When there is a symmetry direction with coordinate ζ, the magnetic field can be written in terms of a flux function $f(\mathbf{x})$. If the field has a component in the symmetry direction it can be written

$$\mathbf{B} \;=\; \nabla f\times\nabla\zeta \;+\; B_\zeta\nabla\zeta \quad . \tag{10.27}$$

Field lines remain within surfaces where $f(\mathbf{x})$ is constant.

The Lorentz force from the symmetric field can be worked out in several pieces using ample vector calculus. The first term is

$$\begin{aligned}
(\nabla\times\mathbf{B})\times(\nabla f\times\nabla\zeta) &= \Big[(\nabla\times\mathbf{B})\cdot\nabla\zeta\Big]\nabla f - \Big[(\nabla\times\mathbf{B})\cdot\nabla f\Big]\nabla\zeta \\
&= \nabla\cdot\Big[\mathbf{B}\times\nabla\zeta\Big]\nabla f - \nabla\cdot\Big[\mathbf{B}\times\nabla f\Big]\nabla\zeta \\
&= \nabla\cdot\Big[(\nabla f\times\nabla\zeta)\times\nabla\zeta\Big]\nabla f - \nabla\cdot\Big[(\nabla f\times\nabla\zeta)\times\nabla f\Big]\nabla\zeta \\
&\quad -\nabla\cdot\Big[B_\zeta\nabla\zeta\times\nabla f\Big]\nabla\zeta \\
&= -\nabla\cdot\Big[|\nabla\zeta|^2\,\nabla f\Big]\nabla f - \nabla\cdot\Big[|\nabla f|^2\,\nabla\zeta\Big]\nabla\zeta - \nabla\cdot\Big[B_\zeta\nabla f\times\nabla\zeta\Big]\nabla\zeta \\
&= -\nabla\cdot\Big[|\nabla\zeta|^2\,\nabla f\Big]\nabla f - \Big[\nabla B_\zeta\cdot(\nabla f\times\nabla\zeta)\Big]\nabla\zeta
\end{aligned} \tag{10.28}$$

and

$$\begin{aligned}
(\nabla\times\mathbf{B})\times B_\zeta\nabla\zeta &= B_\zeta\Big[\,\nabla\times(B_\zeta\nabla\zeta)\,\Big]\times\nabla\zeta \;=\; B_\zeta\Big[\,\nabla B_\zeta\times\nabla\zeta\,\Big]\times\nabla\zeta \\
&= -B_\zeta\,|\nabla\zeta|^2\,\nabla B_\zeta \;=\; -\tfrac{1}{2}|\nabla\zeta|^2\,\nabla B_\zeta^2 \quad .
\end{aligned} \tag{10.29}$$

Using these in the equilibrium equation

$$\begin{aligned}
4\pi\nabla p &= (\nabla\times\mathbf{B})\times\mathbf{B} \\
&= -\nabla\cdot\Big[|\nabla\zeta|^2\,\nabla f\Big]\nabla f - \Big[\nabla B_\zeta\cdot(\nabla f\times\nabla\zeta)\Big]\nabla\zeta - \tfrac{1}{2}|\nabla\zeta|^2\,\nabla B_\zeta^2 \quad .
\end{aligned} \tag{10.30}$$

The above equation must obtain in every direction. We first take the component in the ignorable direction by contracting with $\nabla\zeta$. Due to symmetry

$$\nabla\zeta\cdot\nabla p \;=\; \nabla\zeta\cdot\nabla f \;=\; \nabla\zeta\cdot\nabla B_\zeta^2 \;=\; 0 \quad .$$

leaving only a single term in this component of the force

$$\nabla B_\zeta \cdot (\nabla f \times \nabla \zeta) \;=\; \nabla \zeta \cdot (\nabla B_\zeta \times \nabla f) \;=\; 0 \quad . \tag{10.31}$$

The gradients of f and B_ζ must be parallel everywhere in the solution. It can be shown that this requires B_ζ to depend on $\mathbf{x}$ only through a dependence on $f(\mathbf{x})$. In other words

$$B_\zeta(\mathbf{x}) \;=\; B_\zeta\Big[\, f(\mathbf{x}) \,\Big] \quad , \tag{10.32}$$

so that

$$\nabla B_\zeta \;=\; \frac{dB_\zeta}{df}\,\nabla f \;=\; B'_\zeta(f)\,\nabla f \quad , \tag{10.33}$$

by the chain rule. It is a peculiarity of this problem that it involves unspecified functions of the unknown function — this is the nature of the beast. Using this in eq. (10.30)

$$4\pi \nabla p \;=\; -\,\nabla \cdot \Big[|\nabla\zeta|^2\,\nabla f \Big] \nabla f - \tfrac{1}{2}|\nabla\zeta|^2 \frac{dB_\zeta^2}{df}\,\nabla f \quad . \tag{10.34}$$

In order for the left hand side to be parellel to ∇f everywhere it is also necessary for p to depend on $\mathbf{x}$ only through $f(\mathbf{x})$. The flux function must then satisfy the differential equation

$$\nabla \cdot \Big[|\nabla\zeta|^2\,\nabla f \Big] \;=\; -\,4\pi \frac{dp}{df} \;-\; \tfrac{1}{2}|\nabla\zeta|^2 \frac{dB_\zeta^2}{df} \quad . \tag{10.35}$$

This is a single equation for the flux function $f(\mathbf{x})$, known as the Grad-Shafranov equation. It is an equation of elliptic type, but involves two, as-yet-unspecified, functions of the unknown, $p(f)$ and $B_\zeta(f)$.

Cartesian symmetry

When the symmetry direction is $\hat{\mathbf{z}}$ then $\zeta = z$, $|\nabla\zeta|^2 = 1$, and eq. (10.35) becomes

$$\nabla^2 f \;=\; -\,4\pi \frac{d}{df}\left[p(f) + \frac{1}{8\pi} B_z^2 \right] \;=\; -\,4\pi P'(f) \quad , \tag{10.36}$$

so the two functions combine into a single function $P'(f)$. To be an equilibrium a flux function must have the property that the contours of its Laplacian coincide with its own contours. This is not a common property of functions. Solving this equation, the Grad-Shafranov equation, requires one to find a function with this unusual property.

The property is guaranteed by certain additional symmetry. For example, any cylindrical magnetic field can be an equilibrium given the appropriate distribution of axial field and pressure. Cylindrical fields are generated by flux functions $f(r)$ where $r = \sqrt{x^2 + y^2}$ is the cylindrical radius. This generates a field

$$\mathbf{B} \;=\; B_\phi(r)\,\hat{\boldsymbol{\phi}} \;+\; B_z(r)\,\hat{\mathbf{z}} \quad , \tag{10.37}$$

where $B_\phi(r) = f'(r)$. The contours of $f(r)$ will naturally be concentric circles, as will be the contours of its Laplacian. Thus there is always a function $P'(f)$, possibly multi-valued, for which eq. (10.36) is satisfied.

To obtain this function, in the cylindrical case, we multiply eq. (10.36) by $f'(r)$ to get

$$f'\,\nabla^2 f \;=\; f'(r)\,f''(r) \;+\; \frac{2f'^2(r)}{r} \;=\; -\,4\pi f'(r)\,P'(f) \;=\; -\,4\pi\,\frac{d}{dr}\Big[\,P(f)\,\Big] \quad . \tag{10.38}$$

Integrating this gives the pressure function

$$P(r) \;=\; p + \frac{1}{8\pi}B_z^2 \;=\; P(0) \;-\; \frac{1}{8\pi}f'^2(r) \;-\; \int_0^r \frac{f'^2(r')}{2\pi\,r'}\,dr' \quad . \tag{10.39}$$

Replacing $f'(r)$ with B_ϕ shows that the total pressure (plasma plus magnetic) decreases outward to balance inward magnetic tension

$$p \;+\; \frac{|\mathbf{B}|^2}{8\pi} \;=\; P(0) \;-\; \int_0^r \frac{B_\phi^2(r')}{2\pi\,r'}\,dr' \quad . \tag{10.40}$$

The salient point is that for any azimuthal field $B_\phi(r)$ there is a function $P(r)$ which makes it an equilibrium. The equilibrium may be achieved with plasma pressure alone ($B_z = 0$), axial field alone ($p = 0$, in which case it is a non-linear force-free field), or any combination of the two.

Genuinely two-dimensional solutions to eq. (10.36) can be found for different choices of the function $P'(f)$. Choosing a linear function, $P'(f) = C\,f$ results in a linear equation, namely Helmholtz equation,

$$\nabla^2 f \;=\; -\,4\pi C\,f \;=\; -\,\alpha^2\,f \quad , \tag{10.41}$$

after defining a new constant, $\alpha^2 = 4\pi C$. The solution to this equation will depend on boundary condition. Solutions with cartesian and cylindrical boundaries are

$$\begin{aligned} f(x,y) &= \cos(kx)\,e^{-y\sqrt{k^2-\alpha^2}} \quad , \quad y \;\geq\; 0 \\ f(r,\phi) &= J_m(\alpha r)\,\cos(m\phi) \quad , \end{aligned}$$

where k and m are constants, and $k^2 > \alpha^2$. In the absence of of plasma pressure this choice requires

$$P(f) \;=\; \frac{B_z^2}{8\pi} \;=\; \frac{\alpha^2\,f^2}{8\pi} \quad , \tag{10.42}$$

so $B_z = \alpha f$, and the solution is a constant-α force-free field.

If, however, $P'(f)$, is chosen to be a non-linear function of its argument then eq. (10.36) will be a non-linear PDE for $f(x,y)$. There is no general method of solving such equations, so not much can be said in general about non-linear force-free fields. There are, however, many isolated examples of functions satisfying the equation. Such a solution can be identified by taking its Laplacian and verifying that it is a function of $f(x,y)$ alone. Consider, as an example, the chain of magnetic islands generated by the flux function

$$f(x,y) \;=\; \frac{B_0}{k}\,\ln\Big[\,\cosh(ky) \;-\; \xi\,\cos(kx)\,\Big] \quad , \tag{10.43}$$

whose contours are shown in fig. 10.2; k, B_0 and ξ are parameters and $0 \leq \xi < 1$. Differentiation of this expression reveals that

$$\nabla^2 f \;=\; \frac{(1-\xi^2)kB_0}{\Big[\cosh(ky) - \xi\,\cos(kx)\Big]^2} \;=\; (1-\xi^2)\,kB_0\,e^{-2kf/B_0} \quad . \tag{10.44}$$

It is clear from the right hand expression that $\nabla^2 f$ is a function of f alone and therefore satisfies eq. (10.36). That function is the derivative of

$$P(f) \;=\; P(0) + \frac{B_0^2}{8\pi}(1-\xi^2)\, e^{-2kf/B_0} \;=\; P(0) + \frac{B_0^2}{8\pi}\;\frac{1-\xi^2}{\left[\cosh(ky) - \xi\cos(kx)\right]^2} \quad . \qquad (10.45)$$

The total pressure in eq. (10.45) could be achieved with pressure alone ($B_z = 0$) or with axial field alone ($p = 0$, this is then a *non-linear force-free field*), or by any combination of the two.

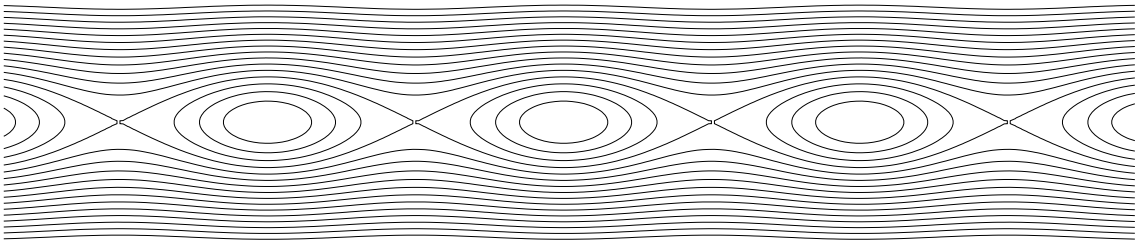

Figure 10.2: A chain of magnetic islands generated by the flux function eq. (10.43).

The island-chain equilibrium is interesting by itself, but is offered here as an example of a non-linear equilibrium. It would be extraordinarily difficult to start from the non-linear PDE

$$\nabla^2 f \;=\; -\,e^{-f} \quad ,$$

and obtain eq. (10.43) as a solution. This explains the extreme scarcity of non-linear equilibria with simple expressions.

Chapter 11

MHD waves

The equations of ideal MHD (without viscosity, thermal conduction or resistivity) omitting gravity are

$$\begin{aligned}
\partial_t \rho &= -\nabla \cdot (\mathbf{u}\rho) && (11.1)\\
\rho\Big[\, \partial_t \mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u}) \,\Big] &= -\nabla p \;+\; \frac{1}{4\pi}(\nabla \times \mathbf{B}) \times \mathbf{B} && (11.2)\\
\partial_t p + \mathbf{u}\cdot\nabla p &= -\gamma p \nabla \cdot \mathbf{u} && (11.3)\\
\partial_t \mathbf{B} &= \nabla \times (\mathbf{u} \times \mathbf{B}) && (11.4)
\end{aligned}$$

These represent eight equations for the time-evolution of the eight fields $\rho(\mathbf{x},t)$, $u_x(\mathbf{x},t)$, $u_y(\mathbf{x},t)$, $u_z(\mathbf{x},t)$, $p(\mathbf{x},t)$, $B_x(\mathbf{x},t)$, $B_y(\mathbf{x},t)$ and $B_z(\mathbf{x},t)$. They compose a non-linear system which we will linearize exactly as we did the hydrodynamic equations in Chapter 4. Following that system we introduce a column of the eight fields,

$$\underline{\mathrm{U}}(\mathbf{x},t) = [\rho, u_x, u_y, u_z, p, B_x, B_y, B_z]^T \quad ,$$

to represent the entire set of unknowns.

The first step in linearization is to identify an equilibrium solution to eqs. (11.1)–(11.4). There are an infinite number of such solutions from which we will restrict ourselves to a very limited subset: *homogeneous, static* equilibria. This means $\mathbf{u}_0 = 0$ and ρ_0, p_0, and $\mathbf{B}_0$ are all independent of space and time. These can be seen to satisfy (11.1)–(11.4). In particular since $\nabla p_0 = 0$ and $\nabla \times \mathbf{B}_0 = 0$, magnetostatic balance,

$$0 \;=\; -\nabla p_0 \;+\; \frac{1}{4\pi}(\nabla \times \mathbf{B}_0) \times \mathbf{B}_0 \quad , \qquad (11.5)$$

is satisified. Linearizing about this particular equilibrium yields the system of equations

$$\begin{aligned}
\partial_t \rho_1 &= -\rho_0 \nabla \cdot \mathbf{u}_1 && (11.6)\\
\rho_0 \partial_t \mathbf{u}_1 &= -\nabla p_1 + \frac{1}{4\pi}(\nabla \times \mathbf{B}_1) \times \mathbf{B}_0 && (11.7)\\
\partial_t p_1 &= -\gamma p_0 \nabla \cdot \mathbf{u}_1 && (11.8)\\
\partial_t \mathbf{B}_1 &= \nabla \times (\mathbf{u}_1 \times \mathbf{B}_0) \quad , && (11.9)
\end{aligned}$$

where variables with subscript 1 are the unknown perturbations, which will be combined into the column vector $\underline{\mathrm{U}}_1(\mathbf{x},t)$,

The rationale for our very restrictive choice of equilibrium was discussed at length in Chapter 4. A homogeneous equilibrium is one in which the spatial coordinate $\mathbf{x}$ does not appear. The linear equations, (11.6)–(11.9), will commute with each spatial derivative so plane waves

$$\underline{\mathrm{U}}_1(\mathbf{x},t) \;=\; \underline{\hat{\mathrm{U}}}\, e^{i\mathbf{k}\cdot\mathbf{x}-i\omega t} \quad , \tag{11.10}$$

will always be able to satisfy the, system. $\hat{U}_1 = \hat{\rho}_1$, $\hat{U}_2 = \hat{u}_{x,1}$, etc., are complex coefficients. Any spatial derivative, $\partial/\partial x_j$ acting on such an eigenmode will simply "pull down" a factor of ik_j. In other words the symbol ∇ can be simply replaced by the vector $i\mathbf{k}$. The linearized equations therefore become

$$\begin{aligned}
-i\omega\hat{\rho}_1 &= -i\rho_0(\mathbf{k}\cdot\hat{\mathbf{u}}_1) && (11.11)\\
-i\omega\rho_0\hat{\mathbf{u}}_1 &= -i\mathbf{k}\hat{p}_1 + \frac{1}{4\pi}(i\mathbf{k}\times\hat{\mathbf{B}}_1)\times\mathbf{B}_0 \\
&= -i\mathbf{k}\hat{p}_1 + \frac{1}{4\pi}(i\mathbf{k}\cdot\mathbf{B}_0)\hat{\mathbf{B}}_1 - \frac{1}{4\pi}i\mathbf{k}(\mathbf{B}_0\cdot\hat{\mathbf{B}}_1) && (11.12)\\
-i\mathbf{k}\hat{p}_1 &= -i\gamma p_0(\mathbf{k}\cdot\hat{\mathbf{u}}_1) && (11.13)\\
-i\omega\hat{\mathbf{B}}_1 &= i\mathbf{k}\times(\hat{\mathbf{u}}_1\times\mathbf{B}_0) \\
&= i(\mathbf{k}\cdot\mathbf{B}_0)\hat{\mathbf{u}}_1 - i(\mathbf{k}\cdot\hat{\mathbf{u}}_1)\mathbf{B}_0 \quad , && (11.14)
\end{aligned}$$

after using the vector identity $(\mathbf{a}\times\mathbf{b})\times\mathbf{c} = (\mathbf{a}\cdot\mathbf{c})\mathbf{b} - (\mathbf{b}\cdot\mathbf{c})\mathbf{a}$.

Our magneto-hydrostatic (MHS) equilibrium has in common with the hydrostatic (HS) equilibrium homogeneity — the position, $\mathbf{x}$, does not appear in either. In that respect, all positions in space are equivalent and we are free to choose the origin anywhere that is convenient. The HS equilibrium is also *isotropic*, meaning every *direction* is equivalent. This means we are free to *orient* our coordinate axis in any way that is convenient. We took advantage of this freedom in Chapter 4 to choose $\hat{\mathbf{x}}$ to be parallel to the wave vector $\mathbf{k}$. The wave vector is not part of the equilibrium, but introduced in a later part of the analysis. The MHS equilibrium differs from the HS equilibrium because it is *not* isotropic — the magnetic field points in a direction which is therefore different from all other directions. We can, for convenience, choose the coordinate system so one axis, say $\hat{\mathbf{z}}$, points in this direction. In other words $\mathbf{B}_0 = B_0\hat{\mathbf{z}}$.

The anisotropy of the MHS equilibrium deprives us of the freedom to align the axis along $\mathbf{k}$, since we have already chosen to align it along $\mathbf{B}_0$. While the equilibrium is not isotropic, it does possess a symmetry known as *gyrotropy* (it is a *gyrotropic equilibrium*), meaning it is unchanged by rotation about a single axis: in this case the direction of $\mathbf{B}_0$. All directions perpendicular to this symmetry axis are equivalent. This means were are free to align the coordinates so that $\mathbf{k}$ lies within the x–z plane. In other words we defined

$$\mathbf{k} \;=\; k\,\hat{\mathbf{k}} \;=\; k[\sin\theta\,\hat{\mathbf{x}} \;+\; \cos\theta\,\hat{\mathbf{z}}] \quad , \tag{11.15}$$

where θ is the angle between the equilibrium magnetic field and the wave vector.

The choices described above allow us to simplify the appearance of eqs. (11.11)–(11.14) in several ways. First it can be readily seen that every term on the rhs is proporitonal to k

and we can divide by this. This leaves, on the left, $v = \omega/k$, the phase velocity of the plane wave.[1] Multiplying all equations by i we make the entire system real

$$
\begin{aligned}
v\,\hat{\rho}_1 &= \rho_0(\sin\theta\,\hat{u}_x + \cos\theta\,\hat{u}_z) && (11.16)\\
v\,\hat{u}_{x,1} &= \frac{1}{\rho_0}\sin\theta\,\hat{p}_1 - \frac{B_0}{4\pi\rho_0}\cos\theta\,\hat{B}_{x,1} + \frac{B_0}{4\pi\rho_0}\sin\theta\,\hat{B}_{z,1} && (11.17)\\
v\,\hat{u}_{y,1} &= -\frac{B_0}{4\pi\rho_0}\cos\theta\,\hat{B}_{y,1} && (11.18)\\
v\,\hat{u}_{z,1} &= \frac{1}{\rho_0}\cos\theta\,\hat{p}_1 && (11.19)\\
v\,\hat{p}_1 &= \gamma p_0(\sin\theta\,\hat{u}_{x,1} + \cos\theta\,\hat{u}_{z,1}) && (11.20)\\
v\,\hat{B}_{x,1} &= -B_0\cos\theta\,\hat{u}_{x,1} && (11.21)\\
v\,\hat{B}_{y,1} &= -B_0\cos\theta\,\hat{u}_{y,1} && (11.22)\\
v\,\hat{B}_{z,1} &= B_0\sin\theta\,\hat{u}_{x,1}\quad . && (11.23)
\end{aligned}
$$

The system naturally consists of 8 equations for eight unknown vector components, $\underline{\hat{\mathrm{U}}}$, *and* an unknown phase velocity, v. This can be written symbollocally

$$v\,\underline{\hat{\mathrm{U}}} \;=\; \underline{\underline{\hat{\mathrm{F}}}}\cdot\underline{\hat{\mathrm{U}}}\quad , \tag{11.24}$$

so the phase velocity is an *eigenvalue* of the 8×8 matrix. We show below that the eigenvalues of $\underline{\underline{\mathrm{F}}}$ are roots of the $8^{\rm th}$ order polynomial

$$\det(\underline{\underline{\mathrm{F}}} - v\underline{\underline{\mathrm{I}}}) \;=\; v^2\,(v^2 - v_{\rm A}^2)[v^4 - (v_{\rm A}^2 + c_s^2)v^2 + v_{\rm A}^2 c_s^2\cos^2\theta] \;=\; 0\quad , \tag{11.25}$$

where $v_{\rm A} = B_0/\sqrt{4\pi\rho_0}$ is the Alfvén speed and $c_s = \sqrt{\gamma p_0/\rho_0}$ is, as before, the sound speed. The eight roots form the 8 branches of the dispersion relation for MHD waves.

11A Alfvén waves

It is a daunting task to find all eight eigenvalues of an arbitrary 8×8 matrix. Fortunately there are aspects of the system (11.16)– (11.23), which make the task a bit easier. First, the two unknowns $\hat{u}_{y,1}$ and $\hat{B}_{y,1}$ appear only in the two eqs. (11.18) and (11.22). This means out 8×8 matrix consists of a 2×2 matrix and an independent 6×6 matrix. The 2×2 system is

$$v\begin{bmatrix}\hat{u}_{y,1}\\ \hat{B}_{y,1}\end{bmatrix} \;=\; \begin{bmatrix}0 & -\cos\theta\,B_0/4\pi\rho_0\\ -\cos\theta\,B_0 & 0\end{bmatrix}\cdot\begin{bmatrix}\hat{u}_{y,1}\\ \hat{B}_{y,1}\end{bmatrix}\quad . \tag{11.26}$$

The eigenvalues of this 2×2 matrix are readily found to be

$$v \;=\; \pm\frac{B_0}{\sqrt{4\pi\rho_0}}\cos\theta \;=\; \pm\,v_{\rm A}\,\cos\theta\quad , \tag{11.27}$$

[1] This remarkable property, that neither ω nor k appear independently but only in the combination, ω/k, follows from the fact that the MHD equations, (11.1)—(11.4), do not define a characteristic length or time scale. In other words it is not possible to form the variables, ρ, u, p and B, into any combinations with units of length or time (try it!). It *is* possible to make several combinations with units of velocity — $\sqrt{p/\rho}$ or $B/\sqrt{\rho}$ are two — and the ratio ω/k will naturally combine these in some way (see below).

where $v_{\rm A} = B_0/\sqrt{4\pi\rho_0}$ is called the Alfvén speed. This set of plane wave normal modes are known as *shear Alfvén waves* or simply Alfvén waves. The eigenvectors corresponding to the upper and lower signs are

$$\left[\begin{array}{c} \hat{u}_{y,1} \\ \hat{B}_{y,1} \end{array} \right] \;=\; \left[\begin{array}{c} v_{\rm A} \\ \mp B_0 \end{array} \right] \quad . \tag{11.28}$$

The velocity perturbation of the Alfvén wave, $\mathbf{u}_1$, is directed perpendicular to the wave vector and thus it is called a *transverse* wave. One consequence of this is that

$$\nabla \cdot \mathbf{u}_1 \;\rightarrow\; i\mathbf{k}\cdot\hat{\mathbf{u}}_1 \;=\; 0 \quad ,$$

so the perturbation is *incompressible*. It is for this reason that neither density nor pressure is perturbed by the wave: $\rho_1 = p_1 = 0$.

The Alfvén wave can be though of as an electromagnetic wave, in the same general category as familiar electromagnetic (EM) radiation in vacuum. The electric field is related to the plasma velocity through eq. (8.7),

$$\mathbf{E}_1(\mathbf{x},t) \;=\; -\frac{1}{c}\mathbf{u}_1(\mathbf{x},t)\times\mathbf{B}_0 \;=\; -\frac{B_0}{c}\,u_{y,1}(\mathbf{x},t)\,\hat{\mathbf{x}} \quad . \tag{11.29}$$

Using eq. (11.28) to replace $u_{y,1} = \mp(v_{\rm A}/B_0)B_{y,1}$, yields

$$\mathbf{E}_1(\mathbf{x},t) \;=\; \pm\,\frac{v_{\rm A}}{c}\,\mathbf{B}_1(\mathbf{x},t)\times\hat{\mathbf{z}} \quad , \tag{11.30}$$

showing that $\mathbf{E}_1$ and $\mathbf{B}_1$ are orthogonal to one another, just as in an EM wave in vacuum. Compared to those waves, where $E_1/B_1 = 1$ (in cgs), an Alfvén wave has a far weaker electric field: $E_1/B_1 = v_{\rm A}/c \ll 1$. The source of the difference is that the displacement current

$$\frac{1}{4\pi}\frac{\partial\mathbf{E}_1}{\partial t} \;=\; \pm\,\frac{v_{\rm A}}{4\pi c}\,\frac{\partial\mathbf{B}_1}{\partial t}\times\hat{\mathbf{z}} \;=\; \mp\, i\,\frac{\omega v_{\rm A}}{4\pi c}\mathbf{B}_1\times\hat{\mathbf{z}} \;=\; -\,ik_z\,\frac{v_{\rm A}^2}{4\pi c}\mathbf{B}_1\times\hat{\mathbf{z}} \quad , \tag{11.31}$$

is far smaller than the actual current

$$\mathbf{J}_1 \;=\; \frac{c}{4\pi}\nabla\times\mathbf{B}_1 \;=\; ik\,\frac{c}{4\pi}\,\hat{\mathbf{k}}\times\mathbf{B}_1 \quad , \tag{11.32}$$

by a factor $v_{\rm A}^2/c^2 \ll 1$. Thus the actual current replaces the displacement current, critical to EM waves in vacuum, to give rise to Alfvén waves propagating much more slowly. Recall that the displacement current is discarded from MHD because it supports only very high frequency dynamics: $\omega \sim kc \gg kv_{\rm A}$.

11A.1 Flow of information and energy flux

The orthogonal electric and magnetic fields in the Alfvén wave gives rise to a Poynting flux

$$\mathbf{S}_2 \;=\; \frac{c}{4\pi}\mathbf{E}_1\times\mathbf{B}_1 \;=\; \pm\,\frac{B_1^2}{4\pi}\,v_{\rm A}\,\hat{\mathbf{z}} \quad . \tag{11.33}$$

The energy flux is along the equilibrium magnetic field, regardless of the direction of the wave vector $\mathbf{k}$. This is another contrast with EM waves in vacuum where $\mathbf{S} \propto \hat{\mathbf{k}}$ – energy flows in the direction of the wave vector.

Due to the anisotropy of the equilibrium the phase velocity, eq. (11.27), depends on the direction of the wave vector. Since the isotropy is broken by the direction of the equilibrium magnetic field, the dependence is on the angle between the wave vector and the magnetic field, θ. The dispersion relation, in more general form, is

$$\omega(\mathbf{k}) \;=\; \pm\, v_{\rm A}\, \hat{\mathbf{b}}_0 \cdot \mathbf{k} \quad , \tag{11.34}$$

where $\hat{\mathbf{b}}_0$ is the direction of $\mathbf{B}_0$, previously assumed to be $\hat{\mathbf{z}}$. The phase velocity is a vector[2] found from the derivative

$$\mathbf{v}_g \;=\; \frac{\partial \omega}{\partial \mathbf{k}} \;=\; \pm\, v_{\rm A}\, \hat{\mathbf{b}}_0 \quad . \tag{11.35}$$

This is always parallel to (or anti-parallel to) the direction of the magnetic field — *regardless of the direction of the wave vector* $\mathbf{k}$. It is no coincidence that the group velocity is parallel to the energy flux — both go purely along the equilibrium magnetic field.

If our equilibrium had more more complex, $\mathbf{B}_0(\mathbf{x})$, then we would not have been able to solve the linear equations using plane waves. Perturbations on length scales much shorter than the scales of the equilibrium can, however, be treated using the Eikenol approximation of Sec. 4F. The perturbation propagates along rays corresponding to the plane-wave solutions. Thus there will be a shear Alfvén ray which, according to eq (4.90), will propagate at the group velocity eq. (11.35). This is perfectly parallel to the local magnetic field direction. Since the group velocity is independent of the wave vector it is not necessary to solve eq. (4.91). Moreover, since $\mathbf{v}_{\rm g}$ is along the magnetic field we see from a comparison of eqs. (4.90) and (9.1), that *magnetic field lines are ray paths of shear Alfvén waves*, at least in the Eikenol limit. This gives another physical interpretation of field lines — an impulsive disturbance at a point will be felt along the entire magnetic field line from that point — and puts another nail into the coffin of field line myth #3.

11A.2 Intuitive view of Alfvén waves

An intuitive perspective on Alfvén waves is offered by considering the field line that passes through the origin at time $t = 0$. The fluid element at this point moves according to

$$\frac{d\mathbf{r}_0}{dt} \;=\; \mathbf{u}[\mathbf{r}_0(t), t] \;=\; u_{y,1}[\mathbf{r}_0(t), t]\, \hat{\mathbf{y}} \;=\; \hat{u}_{y,1}\, \cos[k_x x_0(t) + k_z z_0(t) - \omega t]\, \hat{\mathbf{y}} \quad . \tag{11.36}$$

This fluid element moves only along the y axis, remaining always at $x_0 = z_0 = 0$, so

$$\frac{dy_0}{dt} \;=\; \hat{u}_{y,1}\, \cos(\omega t) \quad , \tag{11.37}$$

and therefore

$$\mathbf{r}_0(t) \;=\; \frac{\hat{u}_{y,1}}{\omega}\, \sin(\omega t)\, \hat{\mathbf{y}} \quad , \tag{11.38}$$

[2]This is one of the seemingly rare instances where common usage conforms to correct terminology. Phase *speed*, $v = \omega/|\mathbf{k}|$, is a scalar giving the speed the phase fronts move perpendicular to themselves. Group *velocity*, $v_{g,j} = \partial\omega/\partial k_j$, is a vector giving the speed *and direction* a wave packet propagates.

The field line passing through this moving point satisfies

$$\frac{\partial y}{\partial z} \;=\; \frac{B_y}{B_z} \;=\; \frac{\hat{B}_{y,1}}{B_0}\,\cos(k_z z - \omega t) \quad . \tag{11.39}$$

Integrating this from $z = 0$ gives the field line

$$\begin{aligned} y(z,t) \;&=\; y_0(t) \;+\; \frac{\hat{B}_{y,1}}{B_0}\int_0^z \cos(k_z z' - \omega t)\,dz' \\ &=\; \frac{\hat{u}_{y,1}}{\omega}\,\sin(\omega t) \;+\; \frac{\hat{B}_{y,1}}{B_0 k_z}\Big[\sin(k_z z - \omega t) + \sin(\omega t)\Big] \quad . \end{aligned} \tag{11.40}$$

Equation (11.28) can be used to replace $\hat{B}_{y,1} = \mp(B_0/v_{\rm A})\hat{u}_{y,1}$; eq. (11.28) to replace $k_z = \pm\omega/v_{\rm A}$. The result is the field line

$$y(z,t) \;=\; -\,\frac{\hat{u}_{y,1}}{\omega}\,\sin(k_z z - \omega t) \;=\; -\,\frac{\hat{u}_{y,1}}{\omega}\,\sin[k_z(z - \omega t/k_z)] \quad , \tag{11.41}$$

which is a sinusoidal distortion propagating at speed $\omega/k_z = \pm v_{\rm A}$ along the z direction. The distortion has peaks and troughs where $B_{y,1}(z,t) = 0$ — where the field line is horizontal. Nodes in the field line ($y = 0$) occur at extrema of $B_{y,1}(z,t) \propto \cos(k_z z - \omega t)$ — where the magnetic field is most deflected. The perturbation pattern will move left or right according to the sign of $\omega/k_z = \pm v_{\rm A}$. This is related to the phase relation between $B_{y1,}$ and $u_{y,1}$, which is equivalent to the relation between trough locations and downward motion: a trough moves toward the side where adjacent fluid elements are moving downward.

The picture above is exactly analogous to a small, sinusoidal perturbation propagating on a string under tension. The field line is a material curve, just like a string, and magnetic field provides a tension force, just like a string. This analogy makes clear why energy and information flow purely along the field direction, $\hat{\mathbf{z}}$.

The phasing for the above traveling wave differs from that of the standing Alfvén wave from eqs. (9.28) and (9.29). In the standing wave the anti-nodes of the magnetic field occur at nodes of the velocity ($y = \pm L, \pm 2L, \ldots$), and *vice versa*. This contrasts with the traveling wave where the anti-nodes of both fields occur at the same points and their relative sign determines the direction of propagation. A super-position of two counter-propagating traveling waves of equal amplitude produces a standing wave.

All waves require a restoring force. The restoring force for the Alfvén wave is magnetic tension, which opposes bending of magnetic field lines. Field lines are bent by perturbations varying in z, along the line, and thus the oscillation frequency depends on k_z. Variation in x, from k_x, will produce a line of troughs making an angle $\tan^{-1}(k_x/k_z)$ with $\hat{\mathbf{x}}$. Each field line responds independently to its own bend causing its trough to move at speed ω/k_z along itself. As neighboring field lines respond independently, but at the same rate, the line of troughs will appear to move at a speed ω/k normal to itself. This is the phase speed of the wave, but can be regarded as an illusion since the field lines are really propagating distortions at a higher speed along their own length.

We can extend the reasoning above to the case where $\mathbf{k} = k\hat{\mathbf{x}}$ so the velocity is the same at all points on a single field line: $\mathbf{u}_1 = \hat{u}_{y,1}\cos(k_x x)\hat{\mathbf{y}}$. This will not bend the field line and thus there will be no restoring force. The wave is a hydrodynamic *shear mode* with $\omega = 0$.

11B Zero-frequency modes

The linear system (11.16)–(11.23), has solution with zero frequency, and thus zero phase velocity: $\omega = 0$ so $v = 0$, just as the hydrodynamic system did. Since the unknown component, $\hat{\rho}_1$, appears nowhere on the rhs, the *entropy mode*

$$\underline{\hat{\mathrm{U}}}^{(7)} = \begin{bmatrix} 1, & 0, & 0, & 0, & 0, & 0, & 0, & 0 \\ \hat{\rho}_1 & \hat{u}_{x,1}, & \hat{u}_{y,1}, & \hat{u}_{z,1}, & \hat{p}, & \hat{B}_{x,1}, & \hat{B}_{y,1}, & \hat{B}_{z,1} \end{bmatrix}^T \tag{11.42}$$

is an MHD eigenmode, to which we assign the number 7. This represents an initial perturbation to the density distribution which does not affect the pressure. There is no restoring force without a pressure perturbation, so the density perturbation persists.

A second mode can be found by noting that $\hat{B}_{x,1}$ and $\hat{B}_{z,1}$ enter only into eq. (11.17). They enter this equation in the combination, $-\cos\theta \hat{B}_{x,1} + \sin\theta \hat{B}_{z,1}$ so the mode

$$\underline{\hat{\mathrm{U}}}^{(8)} = \begin{bmatrix} 0, & 0, & 0, & 0, & 0, & \sin\theta, & 0, & \cos\theta \\ \hat{\rho}_1 & \hat{u}_{x,1}, & \hat{u}_{y,1}, & \hat{u}_{z,1}, & \hat{p}, & \hat{B}_{x,1}, & \hat{B}_{y,1}, & \hat{B}_{z,1} \end{bmatrix}^T \tag{11.43}$$

will lead to zero phase velocity: it is a zero-frequency mode. The meaning of this mode is seen from the fact that

$$\nabla \cdot \mathbf{B}_1 \;\rightarrow\; i\mathbf{k} \cdot \hat{\mathbf{B}}_1 \;=\; ik \;\neq\; 0 \quad .$$

Recall that $\nabla \cdot \mathbf{B} = 0$ is not part of the MHD system, eqs. (11.1)–(11.4). Instead it is used as a constraint on the initial condition. If the initial condition satisfies it then so will the solution of the time-dependent MHD equation. Normal mode $\underline{\hat{\mathrm{U}}}^{(8)}$ represents a perturbation to the initial condition which violates the constraint. This violation *persists* as a static violation of the constraint. It is an *unphysical mode* which enters our analysis because we did not impose the constraint $\nabla \cdot \mathbf{B}_1$ on our perturbations. We take this step below in order to reduce our system from 8 to 7 unknowns.[3]

11C Magnetosonic waves - fast & slow

To eliminate the unphysical mode, $\underline{\hat{\mathrm{U}}}^{(8)}$, from further consideration we restrict $\hat{\mathbf{B}}_1$ to be orthogonal to $\mathbf{k}$ (this assures $\nabla \cdot \mathbf{B} = 0$). One acceptable choice, $\hat{\mathbf{B}} = \hat{\mathbf{y}}$, leads to Alfvén modes which we have already discussed. The second choice, to which we now turn, is a vector in the x–z plane

$$\hat{\mathbf{B}}_1 \;=\; \hat{B}_{\perp,1}\, \hat{\mathbf{y}} \times \hat{\mathbf{k}} \;=\; \hat{B}_{\perp,1} \Big[\cos\theta \hat{\mathbf{x}} \;-\; \sin\theta\, \hat{\mathbf{z}} \Big] \quad , \tag{11.44}$$

[3]Counting makes the problem appear more avoidable than it actually is. The three unknowns, $B_{x,1}$, $B_{y,1}$, and $B_{z,1}$, are subject to a constraint, $\nabla \cdot \mathbf{B}_1 = 0$, so there are actually only *two* independent unknowns. Unfortunately it is not possible in general to use the constraint to eliminate one variable, say $B_{z,1}$. Unless we can do this our system will exhibit an unphysical zero-frequency mode corresponding to the extra unknown we failed to eliminate.

where $\hat{B}_{\perp,1}$ is the amplitude of the magnetic field in the $\hat{\mathbf{y}} \times \hat{\mathbf{k}}$ direction. In terms of this amplitude eq. (11.17) can be written

$$v\,\hat{u}_{x,1} \;=\; \frac{1}{\rho_0}\sin\theta\,\hat{p}_1 - \frac{B_0}{4\pi\rho_0}\,\hat{B}_{\perp,1} \quad . \tag{11.45}$$

A linear combination of eqs. (11.21) and (11.23) yields

$$v\,\hat{B}_{\perp,1} \;=\; -\,B_0\,\hat{u}_{x,1} \quad . \tag{11.46}$$

This single equation can replace the two it combined.

We can also eliminate the entropy mode from further consideration by setting

$$\hat{\rho}_1 \;=\; \frac{\rho_0}{\gamma p_0}\,\hat{p}_1 \quad , \tag{11.47}$$

which corresponds to an adiabatic perturbation according to eq. (**??**). Introducing this into eq. (11.16) converts into into a copy of eq. (11.20) and can thus be dropped.

The steps above describe how we can reduce our system by two equations and by two unknowns. We also drop eqs. (11.18) and (11.22), and unknowns $\hat{u}_{y,1}$ and $\hat{B}_{y,1}$, which decouple to form Alfvén modes. This leaves a 4×4 system

$$\underbrace{\begin{bmatrix} -v & 0 & \sin\theta/\rho_0 & -B_0/4\pi\rho_0 \\ 0 & -v & \cos\theta/\rho_0 & 0 \\ \gamma p_0 \sin\theta & \gamma p_0 \cos\theta & -v & 0 \\ -B_0 & 0 & 0 & -v \end{bmatrix}}_{\underline{\underline{\mathrm{M}}}} \cdot \begin{bmatrix} \hat{u}_{x,1} \\ \hat{u}_{z,1} \\ \hat{p}_1 \\ \hat{B}_{\perp,1} \end{bmatrix} = 0 \quad . \tag{11.48}$$

For this to have a non-trivial solution it is necessary that $\det(\underline{\underline{\mathrm{M}}}) = 0$. This leads to a quartic equation in v whose four solutions give the four remaining normal modes.

The determinant of a 4×4 matrix can be written as a sum of 3×3 determinants using a technique known as Laplace expansion (see Riley et al., 2006, §8.9). This takes advantage of the fact that any determinant is linear in the terms of any one row. In particular, the determinant of $\underline{\underline{\mathrm{M}}}$ is a linear combination of the two non-zero terms in its second row

$$\begin{aligned} \det(\underline{\underline{\mathrm{M}}}) \;=\; & (-v)\,\det\begin{bmatrix} -v & 0 & \sin\theta/\rho_0 & -B_0/4\pi\rho_0 \\ 0 & 1 & 0 & 0 \\ \gamma p_0 \sin\theta & \gamma p_0 \cos\theta & -v & 0 \\ -B_0 & 0 & 0 & -v \end{bmatrix} \\ & + \left(\frac{\cos\theta}{\rho_0}\right)\,\det\begin{bmatrix} -v & 0 & \sin\theta/\rho_0 & -B_0/4\pi\rho_0 \\ 0 & 0 & 1 & 0 \\ \gamma p_0 \sin\theta & \gamma p_0 \cos\theta & -v & 0 \\ -B_0 & 0 & 0 & -v \end{bmatrix} \quad . \end{aligned}$$

Each of these individual 4×4 determinants contains a row (row 2) with a single non-zero entry, namely 1. That determinant can be recast as the determinant of the 3×3 matrix which excludes that row and the column with the one (called the *minor* of that element). This

is done by by successively swapping rows and columns until the 1 is moved into the lower right corner. According to the rules of determinants (see Riley et al., 2006, for example), each row or column swap changes the sign of the determinant. Thus the 3×3 determinant is multiplied by $+1$ (-1) if the number of swaps required was even (odd). The result is

$$\begin{aligned}
\det(\underline{\underline{\mathrm{M}}}) &= -v \det \begin{bmatrix} -v & \sin\theta/\rho_0 & -B_0/4\pi\rho_0 \\ \gamma p_0 \sin\theta & -v & 0 \\ -B_0 & 0 & -v \end{bmatrix} \\
&\quad + (-1)\frac{\cos\theta}{\rho_0} \det \begin{bmatrix} -v & 0 & -B_0/4\pi\rho_0 \\ \gamma p_0 \sin\theta & \gamma p_0 \cos\theta & 0 \\ -B_0 & 0 & -v \end{bmatrix} \\
&= -v\Big[-v^3 \;+\; v\,c_s^2 \sin^2\theta \;+\; v\,v_{\mathrm{A}}^2\Big] - \frac{\cos\theta}{\rho_0}\Big[v^2 \gamma p_0 \cos\theta \;-\; \gamma p_0\, v_{\mathrm{A}}^2\, \cos\theta\Big] \\
&= v^4 \;-\; v^2\, v_{\mathrm{A}}^2 \;-\; v^2\, c_s^2 \;+\; c_s^2\, v_{\mathrm{A}}^2\, \cos^2\theta \;=\; 0 \quad . \qquad (11.49)
\end{aligned}$$

This is the quartic equation defined by eq. (11.48).

Equation (11.49) is a simple quadratic equation in the variable v^2 and has two distinct solutions

$$v_{f,s}^2(\theta) \;=\; \tfrac{1}{2}(v_{\mathrm{A}}^2 + c_s^2) \;\pm\; \tfrac{1}{2}\sqrt{(v_{\mathrm{A}}^2 + c_s^2)^2 - 4c_s^2\, v_{\mathrm{A}}^2 \cos^2\theta} \qquad (11.50)$$

known as *fast* (upper sign) and *slow* (lower sign) *magnetosonic waves* — sometimes simply called *fast modes* and *slow modes*. It can be shown that for no choice of θ, v_{A} or c_s, is either expression negative. Thus each mode has a positive and negative frequency (as it must for real solutions) with no imaginary part (no damping or growth). These are just two more plane wave normal modes. They join the shear Alfvén wave and the zero-frequency entropy mode to form the complete spectrum of a magnetized plasma.

The variation of the phase speed with direction of wave vector, $\mathbf{k}$, is traditionally depicted using *Friedrichs diagrams*, such as those in fig. 11.1. Each of the wave modes is depicted by a different curve. Distance from the origin in a particular direction, θ from upward, represents the phase speed of that wave for that direction. Both the Alfvén mode and slow mode have zero phase speed perpendicular to the field ($\theta = \pm\pi/2$) and thus intersect the origin rather than crossing the horizontal axis.

The three examples in fig. 11.1 show how the inter-relation of the wave modes changes as the sound speed is greater or less than the Alfén speed. Both situations are possible depending on the plasma β

$$\beta \;=\; \frac{p_0}{B_0^2/8\pi} \;=\; \frac{2}{\gamma}\,\frac{c_s^2}{v_{\mathrm{A}}^2} \quad . \qquad (11.51)$$

The case of $\beta \gg 1$, equivalent to $c_s \gg v_{\mathrm{A}}$, is one where the magnetic field is too weak to significantly influence the fluid. Taking that limit in eq. (11.50), gives a fast mode speed

$$v_f(\theta) \;\simeq\; c_s \quad , \quad c_s \;\gg\; v_{\mathrm{A}} \quad , \qquad (11.52)$$

like a sound wave from hydrodynamics — the magnetic field has not changed it. The wave is isotropic, as it was in hydrodynamics. While the magnetic field does break this symmetry, it is too weak to have any actual effect.

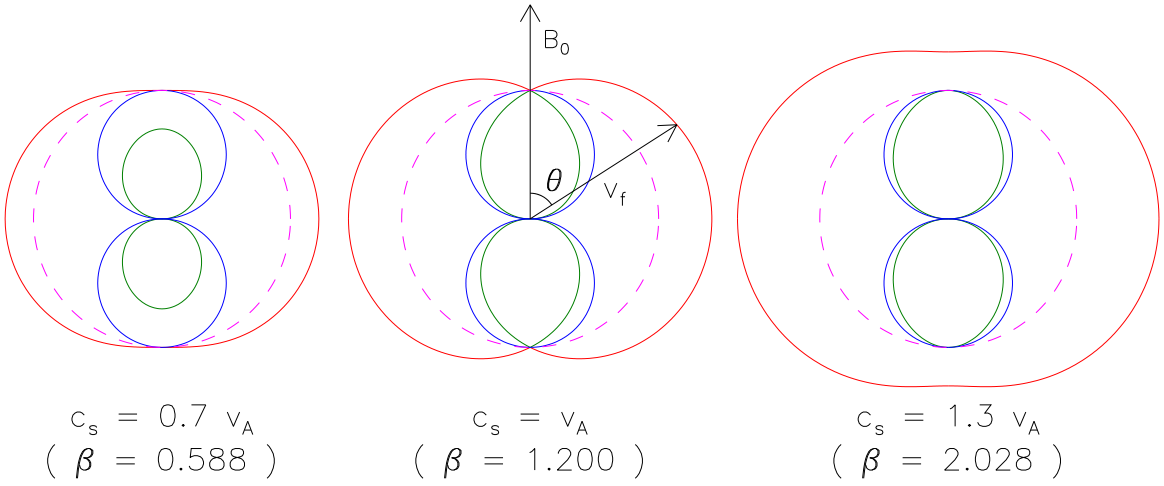

Figure 11.1: Plots of the phase speeds of the three MHD waves in three different parameter regimes: $c_s = 0.7\,v_{\rm A}$ (left), $c_s = v_{\rm A}$ (center), $c_s = 1.3\,v_{\rm A}$ (right). Wave modes are traced by solid curves: fast magnetosonic (red, outer), shear Alfvén (blue, middle), and slow magnetosonic (green, inner). The phase speed at a given angle is represented by the distance from the origin. A black arrow shows the axis from the origin along the direction of the equilibrium magnetic field ($\mathbf{B}_0$). A magenta dashed line shows a circle of radius $v_{\rm A}$.

The slow mode speed, eq. (11.50), can be expanded in the small parameter $v_{\rm A}/c_s$,

$$\begin{aligned} v_s^2(\theta) &= \tfrac{1}{2}(v_{\rm A}^2 + c_s^2) \;-\; \tfrac{1}{2}c_s^2\sqrt{1 + 2v_{\rm A}^2/c_s^2 - 4(v_{\rm A}^2/c_s^2)\cos^2\theta + \cdots} \\ &\simeq \tfrac{1}{2}v_{\rm A}^2 \;-\; \tfrac{1}{2}\Big(\,v_{\rm A}^2 \;-\; 2v_{\rm A}^2\cos^2\theta\,\Big) \;=\; v_{\rm A}^2\cos^2\theta \quad , \end{aligned} \tag{11.53}$$

to match that of the Alfvén wave. Weak as it is the magnetic field provides a small restoring force to the shear mode, giving these erstwhile zero-frequency modes a finite frequency. Each polarization of shear mode, $\hat{\mathbf{u}}_1 \propto \hat{\mathbf{y}}$ and $\hat{\mathbf{u}}_1 \propto \hat{\mathbf{y}} \times \hat{\mathbf{k}}$, gives rise to a different polarization of Alfvén wave. All all incompressible, $\nabla \cdot \mathbf{u}_1 \to i\mathbf{k} \cdot \hat{\mathbf{u}}_1 = 0$, like the hydrodynamic shear mode.

11C.1 Parallel propagation

Further intuition can be gained considering specific directions of propagation, corresponding to specific cuts through fig. 11.1. The case of parallel propagation, $\hat{\mathbf{k}} = \hat{\mathbf{z}}$ or $\theta = 0$, corresponds to vertical cuts. Taking that limit in eq. (11.50) becomes

$$v_{f,s}^2 \;=\; \tfrac{1}{2}(v_{\rm A}^2 + c_s^2) \;\pm\; \tfrac{1}{2}|v_{\rm A}^2 - c_s^2| \quad . \tag{11.54}$$

This means $v_f = \max(v_{\rm A}, c_s)$ and $v_s = \min(v_{\rm A}, c_s)$. Thus when $v_{\rm A} > c_s$, (see left panel of fig. 11.1) the fast magnetosonic wave and shear Alfvén wave both propagate along the field at the Alfvén speed; the slow magnetosonic wave propagates at the sound speed. In

the other case, $v_{\rm A} < c_s$, (see right panel of fig. 11.1) the slow magnetosonic wave and shear Alfvén wave both propagate at the Alfvén speed and the fast magnetosonic wave propagates at the sound speed.

Setting $\theta = 0$ in eq. (11.48) decouples into a pair of 2×2 systems

$$\begin{bmatrix} -v & -B_0/4\pi\rho_0 \\ -B_0 & -v \end{bmatrix} \cdot \begin{bmatrix} \hat{u}_{x,1} \\ \hat{B}_{\perp,1} \end{bmatrix} = 0 \tag{11.55}$$

$$\begin{bmatrix} -v & 1/\rho_0 \\ \gamma p_0 & -v \end{bmatrix} \cdot \begin{bmatrix} \hat{u}_{z,1} \\ \hat{p}_1 \end{bmatrix} = 0 \tag{11.56}$$

The determinant in the two matrices are $v^2 - v_{\rm A}^2$ and $v^2 - c_s^2$ respectively. Only one of these will vanish for a particular wave, the other system demands a *trivial solution*. The mode propagating at $v = \pm c_s$ will thus have $\hat{u}_{x,1} = \hat{B}_{\perp,1} = 0$. It will be a longitudinal wave with no magnetic disturbance, exactly like a hydrodynamic sound wave (which is why it has the same speed). The other wave, propagating at $v = \pm v_{\rm A}$ must have $\hat{u}_{z,1} = \hat{p}_1 = 0$. It is an incompressible, transverse mode exactly like the shear Alfvén mode. This velocity in this mode is $\mathbf{u}_1 \propto \hat{\mathbf{x}}$ direction, in contrast to the mode in section 11A, which had $\mathbf{u}_1 \propto \hat{\mathbf{y}}$. The two are, in the case of parallel propagation, two polarizations of the same wave. Ordinarily we distinguish the shear Alfvén mode by its velocity normal to the plane spanned by $\mathbf{B}_0$ and $\mathbf{k}$. In the case where $\mathbf{B}_0 \parallel \mathbf{k}$ there is no plane and one of the magnetosonic modes (the one with $v \to v_{\rm A}$) becomes degenerate with the shear Alfvén wave.

11C.2 Perpendicular propagation

Perpendicular propagation ($\theta = \pi/2$) gives phase speed

$$v_f = \sqrt{v_{\rm A}^2 + c_s^2} \equiv v_{\rm FMS} \quad , \quad v_s = \frac{v_{\rm A} c_s}{v_{\rm FMS}} \cos\theta \to 0 \ . \tag{11.57}$$

$v_{\rm FMS}$ is known as the *fast magnetosonic speed*, although it only characterizes this wave when it propagates perpendicular to the equilibrium field; it propagates more slowly in any other direction.

The eigenmode for perpendicular propagation is found from three rows of eq. (11.48),

$$\begin{bmatrix} \hat{u}_{x,1} \\ \hat{u}_{z,1} \\ \hat{p}_1 \\ \hat{B}_{\perp,1} \end{bmatrix} = \begin{bmatrix} \hat{u}_{x,1} \\ \hat{u}_{z,1} \\ \hat{p}_1 \\ -\hat{B}_{z,1} \end{bmatrix} = \begin{bmatrix} -v \\ 0 \\ \rho_0(v_{\rm A}^2 - v^2) \\ B_0 \end{bmatrix} \tag{11.58}$$

The phase relation between perturbations to gas pressure and magnetic pressure

$$\frac{p_1}{B_0 B_{z,1}/4\pi} = \frac{\rho_0(v^2 - v_{\rm A}^2)}{B_0^2/4\pi} = \frac{v^2}{v_{\rm A}^2} - 1 \ . \tag{11.59}$$

In the fast mode $v^2 = v_{\rm A}^2 + c_s^2$, both pressures have the same sign: they are in phase. The magnetic pressure and plasma pressure *both* provide restoring forces so the mode propagates

at a speed exceeding both $v_{\rm A}$ and c_s. This has a velocity perturbation in the $\hat{\mathbf{x}} = \hat{\mathbf{k}}$ direction, so it is a longitudinal wave.

In the slow mode, by contrast ($v = 0$), the pressures have opposite sign and equal magnitude. Their forces therefore cancel each other out giving the mode zero frequency. It also has no velocity, $\hat{u}_{x,1} = 0$. This mode is related to the creation of a new equilibrium where constant total pressure is achieved by varying the plasma pressure and magnetic pressure out of phase.

11C.3 Waves at other angles

We have noted at several points when a wave was either longitudinal or transverse. These terms are applied generally to plane waves which include non-vanishing perturbation to vector fields, such as velocity, magnetic field, or electric field. In a transverse wave, such as an EM wave in vacuum, all vector perturbations ($\mathbf{E}_1$ and $\mathbf{B}_1$ in this case) are perpendicular to the wave vector $\mathbf{k}$. In a longitudinal wave they are co-linear (parallel or anti-parallel) with $\mathbf{k}$. This is the case, for example, in a sound wave perturbation to a homogeneous hydrostatic equilibrium.

The categorization into either transverse or longitudinal waves is inevitable for *isotropic systems*, such as a vacuum or homogeneous hydrostatic equilibrium. In those cases the only option is for vector perturbations to point either parallel to or perpendicular to the only direction of significance: the wave vector. There are no other directions available because the equilibrium system defines no direction — it is isotropic. For transverse waves, all perpendicular directions are equivalent so there will be two polarizations with identical dispersion relations — *degenerate* polarizations.[4]

When the equilibrium is not isotropic there is no reason to expect a wave to fall into one of these to simple categories. The equilibrium provides one direction, and the wave vector a second. Each vector perturbation may be in an direction combining the parallel or perpendicular to either of these. This is the case for the magnetosonic modes whose velocity perturbations lie within the plane spanned by $\mathbf{k}$ and $\mathbf{B}_0$, but with no simple relationship to either vector. When the fast mode propagates perpendicular to $\mathbf{B}_0$ it fails to fall into either category since $\mathbf{u}_1$ is parallel to $\mathbf{k}$ while $\mathbf{B}_1$ is perpendicular to it. As the wave vector is swept 90° to be parallel to $\mathbf{B}_0$, the fast mode becomes either a transverse or longitudinal mode depending on whether c_s is less than or greater to $v_{\rm A}$.

Between these limits the fast mode is neither longitudinal nor transverse. To see this consider the particular limit $c_s/v_{\rm A} \to 0$ in which $v_f = v_{\rm A}$ independent of θ (confusingly, the dispersion relations becomes isotropic in this limit). The limit can be obtained by setting $p_0 \to 0$ in eq. (11.48) to obtain $\mathbf{u}_1 = v_{\rm A}\,\hat{\mathbf{x}}$ and while $|\mathbf{B}_1| = B_0$. The magnetic field perturbation must always remain perpendicular to $\mathbf{k}$, in order to assure $\nabla \cdot \mathbf{B}_1 = 0$. The velocity perturbation, however, remains in the $\hat{\mathbf{x}}$ direction regardless of $\mathbf{k}$. This is natural since, in absence of pressure, the only force accelerating plasma is the Lorentz

[4]There is one further symmetry we have not discussed: invariance under mirror reflection or coordinate inversion $\mathbf{x} \to -\mathbf{x}$. It is possible for the system to be invariant under all rotations, it is isotropic, but not be invariant under reflection. Transverse waves can be superposed into left-handed and right-handed circular polarizations, and when mirror is absent these need not be degenerate. This occurs for EM waves propagating through solutions of certain organic molecules lacking mirror symmetry.

force, perpendicular to $\mathbf{B}_0$. Since the direction of $\mathbf{u}_1$ is independent of $\mathbf{k}$, there is no fixed angle between them: it takes on every value from 0° to 90°.

Magnetized equilibria, $\mathbf{B}_0 \neq 0$, provide a direction which can combine with the wave vector to span a plane in which the perturbation vectors ($\mathbf{u}_1$ and $\mathbf{B}_1$) may lie. It is also possible for the perturbations to be perpendicular to this plane and thus remain always perpendicular to $\mathbf{k}$. This is the case for the shear Alfvén mode which is therefore a transverse wave: both $\mathbf{u}_1$ and $\mathbf{B}_1$ are perpendicular to $\mathbf{k}$. There is, however, a single normal to the plane so this transverse wave does not have degenerate polarizations — in general. The only polarization for an Alfvén wave is in the direction $\mathbf{k} \times \mathbf{B}_0$.

The above argument fails in the case were $\mathbf{k}$ is co-linear with $\mathbf{B}_0$ (i.e. parallel propagation). This one situation is similar to that of isotropic equilibria in that the only preferred direction is provided by $\mathbf{k}$ and perturbations must be either perpendcular to it (2 degenerate transverse polarizations) or parallel to it (longitudinal waves). This is why fig. 11.1 shows pairs of curves meeting at the vertical ($\mathbf{B}_0$). These are the degenerate transverse modes, including the shear Alfvén wave. In the middle panel all three curves meet at a single point due to the "accidental degeneracy" when $c_s = v_\mathrm{A}$.

Chapter 12

MHD instability

12A Linearized equations

We have found all linear modes for homogeneous magnetostatic equilibrium: $p_0(\mathbf{x}) = p_0$ and $\mathbf{B}_0(\mathbf{x}) = B_0\hat{\mathbf{z}}$. The linear modes are plane waves with real eigenfrequencies in 4 branches: entropy mode ($\omega = 0$), shear Alfvén waves $\omega = \pm\mathbf{k}\cdot\hat{\mathbf{z}}\,B_0/\sqrt{4\pi\rho}$, and fast and slow magnetosonic waves. The situation becomes more complicated when more complicated equilibria are considered: for example an equilibrium with magnetic field $\mathbf{B}_0(\mathbf{x})$ which is not uniform.

We consider next a more complicated case. We restrict the complexity by limiting ourselves to cases with $\beta \ll 1$. Equilibria in this case are force-free magnetic fields

$$\nabla\times\mathbf{B}_0 \;=\; \alpha\,\mathbf{B}_0 \quad , \quad \mathbf{B}_0\cdot\nabla\alpha \;=\; 0 \quad . \tag{12.1}$$

Due to the low β, the pressure term may be dropped from the momentum equation. It therefore becomes unnecessary to solve the energy equation. The perturbed mass density does not enter into any of the equations so it is also unnecessary to solve the continuity equation. Linearization of the remaining equations are

$$\begin{aligned} -i\omega\,\hat{\mathbf{u}}_1 &= \frac{1}{4\pi\rho_0(\mathbf{x})}(\nabla\times\mathbf{B}_0)\times\hat{\mathbf{B}}_1 \;+\; \frac{1}{4\pi\rho_0(\mathbf{x})}(\nabla\times\hat{\mathbf{B}}_1)\times\mathbf{B}_0 \\ &= \frac{\mathbf{B}_0(\mathbf{x})}{4\pi\rho_0(\mathbf{x})}\times\Big[\,\alpha(\mathbf{x})\,\hat{\mathbf{B}}_1 - \nabla\times\hat{\mathbf{B}}_1\,\Big] \qquad (12.2) \\ -i\omega\,\hat{\mathbf{B}}_1 &= \nabla\times(\hat{\mathbf{u}}_1\times\mathbf{B}_0) \quad , \qquad (12.3) \end{aligned}$$

after taking the usual normal mode form

$$\mathbf{u}_1(\mathbf{x},t) \;=\; \hat{\mathbf{u}}_1(\mathbf{x})\,e^{-i\omega t} \quad , \quad \mathbf{B}_1(\mathbf{x},t) \;=\; \hat{\mathbf{B}}_1(\mathbf{x})\,e^{-i\omega t} \quad . \tag{12.4}$$

Without assuming any spatial symmetry in the equilibrium field, $\mathbf{B}_0(\mathbf{x})$, we cannot predict how the mode amplitudes, $\hat{\mathbf{u}}_1(\mathbf{x})$ and $\hat{\mathbf{B}}_1(\mathbf{x})$ will depend on space.

These two can be combined into a single vector equation for eigenvalue ω

$$\omega^2\,\hat{\mathbf{u}}_1 \;=\; \frac{\mathbf{B}_0(\mathbf{x})}{4\pi\rho_0(\mathbf{x})}\times\Big[\,\nabla\times\nabla\times(\hat{\mathbf{u}}_1\times\mathbf{B}_0) \;-\; \alpha\nabla\times(\hat{\mathbf{u}}_1\times\mathbf{B}_0)\Big] \;=\; \mathbf{F}\{\hat{\mathbf{u}}_1\} \quad , \tag{12.5}$$

where $\mathbf{F}$ is a vector-valued operator, linear in its argument, $\hat{\mathbf{u}}_1(\mathbf{x})$. The linearized equation is an eigenvalue equation with eigenvalue, ω^2. The operator depends on the equilibrium magnetic field $\mathbf{B}_0(\mathbf{x})$ and can be quite complicated.

One set of eigenvalues can, however, be quickly identified. Posing an eigenfunction

$$\hat{\mathbf{u}}_1(\mathbf{x}) \;=\; \hat{\psi}(\mathbf{x})\,\mathbf{B}_0(\mathbf{x}) \quad , \tag{12.6}$$

where $\hat{\psi}(\mathbf{x})$ is an arbitrary scalar function, leads to $\mathbf{F}\{\hat{\mathbf{u}}_1(\mathbf{x})\} = 0$ and thus $\omega = 0$. This infinite variety of zero-frequency eignemodes correspond to modes related to the slow-magnetosonic waves, which at $\beta \to 0$ have zero frequency. They are not plane waves, but do comprise an infinite variety in as much as the function $\hat{\psi}(\mathbf{x})$ is arbitrary.

The Eikenol approximation offers second approach to finding eigenfuncitons of $\mathbf{F}$. We pose eigenfunctions of the form

$$\hat{\mathbf{u}}_1(\mathbf{x}) \;=\; \hat{\mathbf{A}}(\mathbf{x})\,e^{i\phi(x)} \quad , \tag{12.7}$$

and assume $|\nabla\phi| \gg$ than all other inverse length scales. Only spatial derivatives acting on the complex phase factor will be large enough to be significant. We may therefore replace every $\nabla \to \nabla\phi = i\mathbf{k}$. We may also discard the second term in $\mathbf{F}$ since it involves only a single spatial derivative, compared to the two in the first term. The result is

$$\begin{aligned}
\mathbf{F}(\hat{\mathbf{u}}_1) &= -\frac{\mathbf{B}_0}{4\pi\rho_0} \times \Big\{\mathbf{k}\times\Big[\mathbf{k}\times(\hat{\mathbf{u}}_1\times\mathbf{B}_0)\Big]\Big\} \\
&= -\frac{\mathbf{B}_0}{4\pi\rho_0} \times \Big\{\mathbf{k}\times\Big[(\mathbf{k}\cdot\mathbf{B}_0)\hat{\mathbf{u}}_1 - (\mathbf{k}\cdot\hat{\mathbf{u}}_1)\mathbf{B}_0\Big]\Big\} \\
&= -\frac{1}{4\pi\rho_0}\Big\{[\mathbf{B}_0\times(\mathbf{k}\times\hat{\mathbf{u}}_1)](\mathbf{k}\cdot\mathbf{B}_0) - [\mathbf{B}_0\times(\mathbf{k}\times\mathbf{B}_0)](\mathbf{k}\cdot\hat{\mathbf{u}}_1)\Big\} \\
&= \frac{1}{4\pi\rho_0}\Big\{(\mathbf{k}\cdot\mathbf{B}_0)^2\hat{\mathbf{u}}_1 - (\mathbf{k}\cdot\mathbf{B}_0)(\mathbf{B}_0\cdot\hat{\mathbf{u}}_1)\mathbf{k} + B_0^2(\mathbf{k}\cdot\hat{\mathbf{u}}_1)\mathbf{k} - (\mathbf{k}\cdot\mathbf{B}_0)(\mathbf{k}\cdot\hat{\mathbf{u}}_1)\mathbf{B}_0\Big\} \\
&= \frac{1}{4\pi\rho_0}\Big\{(\mathbf{k}\cdot\mathbf{B}_0)^2\hat{\mathbf{u}}_1 + B_0^2(\mathbf{k}\cdot\hat{\mathbf{u}}_1)\mathbf{k} - (\mathbf{k}\cdot\mathbf{B}_0)(\mathbf{k}\cdot\hat{\mathbf{u}}_1)\mathbf{B}_0\Big\} \quad ,
\end{aligned} \tag{12.8}$$

after using the fact that $\mathbf{B}_0\cdot\hat{\mathbf{u}}_1 = 0$ once we discard the slow magnetosonic modes. This can be contracted with $\mathbf{k}$ to yield the equation

$$\mathbf{k}\cdot\mathbf{F}(\hat{\mathbf{u}}_1) \;=\; \frac{B_0^2k^2}{4\pi\rho_0}(\mathbf{k}\cdot\hat{\mathbf{u}}_1) \;=\; \omega^2(\mathbf{k}\cdot\hat{\mathbf{u}}_1) \quad , \tag{12.9}$$

for the compressive wave mode. This leads to the eigenfrequencies

$$\omega(\mathbf{k}) \;=\; \pm\, v_{\rm A}k \quad , \tag{12.10}$$

which are fast magnetosonic waves in the $\beta = 0$ limit; the dispersion relationship is isotropic.

Contracting (12.8) with the orthogonal vector, $\mathbf{k}\times\mathbf{B}_0$ leads to

$$(\mathbf{k}\times\mathbf{B}_0)\cdot\mathbf{F}(\hat{\mathbf{u}}_1) \;=\; \frac{1}{4\pi\rho_0}(\mathbf{k}\cdot\mathbf{B}_0)^2(\mathbf{k}\times\mathbf{B}_0\cdot\hat{\mathbf{u}}_1) \;=\; \omega^2(\mathbf{k}\times\mathbf{B}_0\cdot\hat{\mathbf{u}}_1) \quad . \tag{12.11}$$

The eigenvalues from this

$$\omega(\mathbf{k}) \;=\; \pm\,\frac{\mathbf{k}\cdot\mathbf{B}_0}{\sqrt{4\pi\rho_0}} \;=\; \pm\, v_{\rm A}\hat{\mathbf{b}}_0\cdot\mathbf{k} \quad , \tag{12.12}$$

are shear Alfvén waves; $\hat{\mathbf{b}}_0 = \mathbf{B}_0/B_0$ is the unit vector along the local magnetic field.

This illustrates several solutions to the eigenvalue equation (12.5). They are related to the modes of a homogeneous equilibrium. That relation is inevitable in the Eikonal limit, since it demands modes whose scales are so small than any equilibrium "appears" homogeneous. The more interesting cases are those where the eigenmode, $\hat{\mathbf{u}}_1(\mathbf{x})$ varies on spatial scales comparable to those in the equilibrium $\mathbf{B}_0(\mathbf{x})$. In this case things are far more complicated.

12B Properties of the linear operator

According to eq. (12.5) the square of the eigenfrequency, ω^2, is an eigenvalue of the vector-valued linear operator $\mathbf{F}$. This operator is fairly complicated, especially since it depends on the equilibrium magnetic field $\mathbf{B}_0(\mathbf{x})$, which could be very complicated. The operator is *self-adjoint* with respect to a particular inner product for complex vector-valued functions, $\hat{\mathbf{u}}(\mathbf{x})$ and $\hat{\mathbf{w}}(\mathbf{x})$,

$$\langle\hat{\mathbf{w}}|\hat{\mathbf{u}}\rangle \;=\; \int\limits_{\mathcal{V}} \rho_0(\mathbf{x})\,\hat{\mathbf{w}}^*(\mathbf{x})\cdot\hat{\mathbf{u}}(\mathbf{x})\,d^3x \quad , \tag{12.13}$$

where the integral is over the volume $\mathcal{V}$ occupied by the equilibrium. This inner product is a scalar, linear in each of its arguments and satisfies requirements of an inner product

$$\langle\hat{\mathbf{w}}|\hat{\mathbf{u}}\rangle \;=\; \langle\hat{\mathbf{u}}|\hat{\mathbf{w}}\rangle^* \quad , \quad \langle\hat{\mathbf{u}}|\hat{\mathbf{u}}\rangle \;\geq\; 0 \quad .$$

It depends on the plasma density, $\rho_0(\mathbf{x})$, which we must assume to be finite and non-zero everywhere within the domain $\mathcal{V}$.

The property of self-adjointness depends on the inner product and also on boundary conditions satisfied by the functions. We will assume that the plasma velocity vanishes at the boundary of this volume: $\hat{\mathbf{u}} = \hat{\mathbf{w}} = 0$ on $\partial\mathcal{V}$. The inner product of one function $\hat{\mathbf{w}}(\mathbf{x})$, with the linear operator from eq. (12.5) operating on a second, $\hat{\mathbf{u}}(\mathbf{x})$, can be rewritten

$$\begin{aligned}
\Big\langle\hat{\mathbf{w}}\Big|\mathbf{F}(\hat{\mathbf{u}})\Big\rangle \;&=\; \frac{1}{4\pi}\int\limits_{\mathcal{V}}\hat{\mathbf{w}}^*\cdot\Big\{\,\mathbf{B}_0\times\Big[\,\nabla\times\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\;-\;\alpha\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\,\Big]\,\Big\}\,d^3x \\
&=\; \frac{1}{4\pi}\int\limits_{\mathcal{V}}(\hat{\mathbf{w}}^*\times\mathbf{B}_0)\cdot\nabla\times\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\,d^3x \\
&\quad -\frac{1}{4\pi}\int\limits_{\mathcal{V}}\alpha(\mathbf{x})\,(\hat{\mathbf{w}}^*\times\mathbf{B}_0)\cdot\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\,d^3x \\
&=\; -\frac{1}{4\pi}\oint\limits_{\partial\mathcal{V}}(\hat{\mathbf{w}}^*\times\mathbf{B}_0)\times\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\cdot\,d\mathbf{a} \\
&\quad +\frac{1}{4\pi}\int\limits_{\mathcal{V}}\nabla\times(\hat{\mathbf{w}}^*\times\mathbf{B}_0)\cdot\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\,d^3x
\end{aligned}$$

$$
\begin{aligned}
& -\frac{1}{4\pi}\int_{\mathcal{V}} \alpha(\mathbf{x})\,(\hat{\mathbf{w}}^*\times\mathbf{B}_0)\cdot\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\,d^3x \\
= \;& \frac{1}{4\pi}\int_{\mathcal{V}} \nabla\times(\hat{\mathbf{w}}\times\mathbf{B}_0)^*\cdot\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\,d^3x \\
& -\frac{1}{4\pi}\int_{\mathcal{V}} \alpha(\mathbf{x})\,(\hat{\mathbf{w}}^*\times\mathbf{B}_0)\cdot\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\,d^3x
\end{aligned}
$$

after invoking $\hat{\mathbf{w}} = 0$ on $\partial\mathcal{V}$ to eliminate surface terms.[1] We have also used the fact that the equilibrium magnetic field $\mathbf{B}_0(\mathbf{x})$ is real. The second term can be rewritten using an integration by parts

$$
\begin{aligned}
& \int_{\mathcal{V}} \alpha\,(\hat{\mathbf{w}}^*\times\mathbf{B}_0)\cdot\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\,d^3x \;=\; \tfrac{1}{2}\int_{\mathcal{V}} \alpha\,(\hat{\mathbf{w}}^*\times\mathbf{B}_0)\cdot\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\,d^3x \\
& -\tfrac{1}{2}\oint_{\partial\mathcal{V}} \alpha\Big[\,(\hat{\mathbf{w}}^*\times\mathbf{B}_0)\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\Big]\cdot d\mathbf{a} \;+\; \tfrac{1}{2}\int_{\mathcal{V}} \nabla\times\big[\,\alpha\,(\hat{\mathbf{w}}^*\times\mathbf{B}_0)\big]\cdot(\hat{\mathbf{u}}\times\mathbf{B}_0)\,d^3x \\
& = \tfrac{1}{2}\int_{\mathcal{V}} \alpha\Big[\,(\hat{\mathbf{w}}\times\mathbf{B}_0)\cdot\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0)^* + \nabla\times(\hat{\mathbf{w}}\times\mathbf{B}_0)^*\cdot(\hat{\mathbf{u}}\times\mathbf{B}_0)\Big]\,d^3x \\
& -\tfrac{1}{2}\int_{\mathcal{V}} \nabla\alpha\cdot(\hat{\mathbf{w}}\times\mathbf{B}_0)^*\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\,d^3x \quad ,
\end{aligned}
$$

after the dropping surface term once again. The inner product is therefore

$$
\begin{aligned}
\Big\langle\hat{\mathbf{w}}\Big|\mathbf{F}(\hat{\mathbf{u}})\Big\rangle \;=\; & \frac{1}{4\pi}\int_{\mathcal{V}}\Big[\,\nabla\times(\hat{\mathbf{w}}\times\mathbf{B}_0)^*\cdot\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0) + \tfrac{1}{2}\nabla\alpha\cdot(\hat{\mathbf{w}}\times\mathbf{B}_0)^*\times(\hat{\mathbf{u}}\times\mathbf{B}_0)\Big]\,d^3x \\
& -\frac{1}{8\pi}\int_{\mathcal{V}}\alpha(\mathbf{x})\Big[\,(\hat{\mathbf{w}}\times\mathbf{B}_0)^*\cdot\nabla\times(\hat{\mathbf{u}}\times\mathbf{B}_0) + \nabla\times(\hat{\mathbf{w}}\times\mathbf{B}_0)^*\cdot(\hat{\mathbf{u}}\times\mathbf{B}_0)\Big]\,d^3x \qquad (12.14)
\end{aligned}
$$

The steps above show that the inner product $\langle\hat{\mathbf{w}}|\mathbf{F}(\hat{\mathbf{u}})\rangle$ is symmetric in the functions $\hat{\mathbf{u}}(\mathbf{x})$ and $\hat{\mathbf{w}}^*(\mathbf{x})$. This leads immediately to the conclusion that $\mathbf{F}$ is a self-adjoint operator

$$
\Big\langle\hat{\mathbf{w}}\Big|\mathbf{F}(\hat{\mathbf{u}})\Big\rangle \;=\; \Big\langle\mathbf{F}(\hat{\mathbf{w}})\Big|\hat{\mathbf{u}}\Big\rangle \;=\; \Big\langle\hat{\mathbf{u}}\Big|\mathbf{F}(\hat{\mathbf{w}})\Big\rangle^* \quad , \qquad (12.15)
$$

which is the defining property of a Stürm-Liouville equation; evidently (12.5) is a Stürm-Liouville equation, albeit a vector-valued version. Numerous consequences follow from this fact. The first and most significant is that eigenvalues of $\mathbf{F}$, namely ω^2, must be real. Taking an inner product of (12.5), shows that it corresponds to the ratio

$$
\omega^2 \;=\; \frac{\Big\langle\hat{\mathbf{u}}\Big|\mathbf{F}(\hat{\mathbf{u}})\Big\rangle}{\langle\mathbf{u}|\mathbf{u}\rangle} \quad , \qquad (12.16)
$$

[1] In fact only the normal component of the displacement need vanish, $\hat{\mathbf{n}}\cdot\hat{\mathbf{w}} = 0$ as long as $\hat{\mathbf{n}}\cdot\mathbf{B}_0 = 0$ on the surface.

which can be seen to be real, owing to the self-adjointness of $\mathbf{F}$. This does not, however, mean that the eigenfrequency ω is real, since

$$\Big\langle \hat{\mathbf{u}} \Big| \mathbf{F}(\hat{\mathbf{u}}) \Big\rangle \;=\; \frac{1}{4\pi} \int\limits_{\mathcal{V}} \Big\{ \Big| \nabla \times (\hat{\mathbf{u}} \times \mathbf{B}_0) \Big|^2 \;-\; \alpha(\mathbf{x}) \, (\hat{\mathbf{u}} \times \mathbf{B}_0)^* \cdot \nabla \times (\hat{\mathbf{u}} \times \mathbf{B}_0) \Big\} \, d^3x \quad . \tag{12.17}$$

could be negative. Since its square is real any eigenvalue ω will be either purely real or purely imaginary.

Expression (12.17), related to the eigenfrequency, is a quadratic form which we designate

$$W\{\hat{\mathbf{u}}\} \;=\; \tfrac{1}{2} \Big\langle \hat{\mathbf{u}} \Big| \mathbf{F}(\hat{\mathbf{u}}) \Big\rangle \;=\; \frac{1}{8\pi} \int\limits_{\mathcal{V}} \Big\{ \Big| \nabla \times (\hat{\mathbf{u}} \times \mathbf{B}_0) \Big|^2 - \alpha(\mathbf{x}) \, (\hat{\mathbf{u}} \times \mathbf{B}_0)^* \cdot \nabla \times (\hat{\mathbf{u}} \times \mathbf{B}_0) \Big\} \, d^3x \quad . \tag{12.18}$$

Consider the variation of this function subject to the normalization constraint, $\langle \hat{\mathbf{u}} | \hat{\mathbf{u}} \rangle = 1$, applied by a Lagrange undetermined multiplier

$$\begin{aligned} & W\{\hat{\mathbf{u}} + \delta\hat{\mathbf{u}}\} + \lambda \Big[\, \langle \hat{\mathbf{u}} + \delta\hat{\mathbf{u}} | \hat{\mathbf{u}} + \delta\hat{\mathbf{u}} \rangle \;-\; 1 \, \Big] \\ &= \; W\{\hat{\mathbf{u}}\} + \tfrac{1}{2} \Big\langle \delta\hat{\mathbf{u}} \Big| \mathbf{F}(\hat{\mathbf{u}}) \Big\rangle + \tfrac{1}{2} \Big\langle \hat{\mathbf{u}} \Big| \mathbf{F}(\delta\hat{\mathbf{u}}) \Big\rangle + \; \lambda \langle \delta\hat{\mathbf{u}} | \hat{\mathbf{u}} \rangle + \lambda \langle \hat{\mathbf{u}} | \delta\hat{\mathbf{u}} \rangle + \cdots \\ &= \; W\{\hat{\mathbf{u}}\} + \underbrace{\tfrac{1}{2} \Big\langle \delta\hat{\mathbf{u}} \, \Big| \, \mathbf{F}(\hat{\mathbf{u}}) + 2\lambda \, \hat{\mathbf{u}} \Big\rangle + \tfrac{1}{2} \Big\langle \mathbf{F}(\hat{\mathbf{u}}) + 2\lambda \, \hat{\mathbf{u}} \, \Big| \, \delta\hat{\mathbf{u}} \Big\rangle}_{\delta W} + \cdots \end{aligned} \tag{12.19}$$

For $\hat{\mathbf{u}}(\mathbf{x})$ to be an extremum of W it is necessary that δW vanish for every variation $\delta\hat{\mathbf{u}}(\mathbf{x})$. For both the real and imaginary part of δW to vanish it is necessary that

$$\Big\langle \delta\hat{\mathbf{u}} \, \Big| \, \mathbf{F}(\hat{\mathbf{u}}) + 2\lambda \, \hat{\mathbf{u}} \Big\rangle \;=\; \int\limits_{\mathcal{V}} \delta\hat{\mathbf{u}}^*(\mathbf{x}) \cdot \Big[\, \mathbf{F}(\hat{\mathbf{u}}) + 2\lambda \, \hat{\mathbf{u}} \, \Big] \, d^3x \;=\; 0 \quad , \tag{12.20}$$

for all possible vector-valued functions $\delta\hat{\mathbf{u}}^*(\mathbf{x})$. This means that to be an extremum $\hat{\mathbf{u}}(\mathbf{x})$ must be an eigenfunction of the operator $\mathbf{F}(\mathbf{x})$. The minimum of the ratio

$$R\{\hat{\mathbf{u}}(\mathbf{x})\} \;=\; \frac{\Big\langle \hat{\mathbf{u}} \Big| \mathbf{F}(\hat{\mathbf{u}}) \Big\rangle}{\langle \hat{\mathbf{u}} | \hat{\mathbf{u}} \rangle} \quad , \tag{12.21}$$

is thus the minimum eigenvalue — the minimum value of ω^2. This is a standard method to analyze Stürm-Lioville equations.

The foregoing provides a way to demonstrate instability without solving the full eigenvalue eq. (12.5). If any function $\hat{\mathbf{w}}(\mathbf{x})$ can be found for which

$$W\{\hat{\mathbf{w}}\} \;=\; \frac{1}{8\pi} \int\limits_{\mathcal{V}} \Big\{ \Big| \nabla \times (\hat{\mathbf{w}} \times \mathbf{B}_0) \Big|^2 \;-\; \alpha(\mathbf{x}) \, (\hat{\mathbf{w}} \times \mathbf{B}_0)^* \cdot \nabla \times (\hat{\mathbf{w}} \times \mathbf{B}_0) \Big\} \, d^3x \quad , \tag{12.22}$$

is negative, then there must be at least one negative eigenvalue ω^2. In particular the minimum eigenvalue, $\omega_0^2 < 0$. Corresponding to this will be a pair of imaginary eigenvalues

$$\omega_0 \;=\; \pm \, i\Omega_0 \quad , \qquad \Omega_0 \;\geq\; \Big| W\{\hat{\mathbf{w}}\} \Big| \quad . \tag{12.23}$$

The eigenmode with the upper sign will grow exponentially in time, $e^{-i\omega_0 t} = e^{\Omega_0 t}$. This mode is unstable. Its existence is demonstrated by finding a single negative value of W. This is far easier to do than it would be to find actual eigenvalues of eq. (12.5) and verify that one is negative.

We show below that W is the second variation (i.e. second derivative) of the total magnetic energy. It is the direct analog of the spring constant for this continuum system. A negative spring constant occurs for an equilibrium at a maximum or saddle of of its potential energy functions, instead of a minimum. The same is true here: instability occurs when there is a way to perturb the system from equilibrium which *lowers* its magnetic (potential) energy.

It is far more difficult to establish *stability*, since one must demonstrate than *no function* exists capable of making W negative. Such a demonstration is, however, straight-forward for a potential field, $\nabla \times \mathbf{B} = 0$, regardless of how complicated it might be. In this case $\alpha(\mathbf{x}) = 0$ and the second term eq. (12.22) vanishes leaving

$$W\{\hat{\mathbf{w}}\} \;=\; \frac{1}{8\pi}\int_{\mathcal{V}} \Big|\nabla\times(\hat{\mathbf{w}}\times\mathbf{B}_0)\Big|^2 d^3x \quad . \tag{12.24}$$

Its form makes it clear that W can never be negative. This demonstrates that, at low plasma-β, *any potential magnetic field is a stable equilibrium.*

The first term in the integral eq. (12.22) will never be negative; it is referred to as a *stabilizing* contribution. Instability therefore requires that the second term, the *destabilizing term*, be negative and larger than the fist for some choice of function $\hat{\mathbf{w}}(\mathbf{x})$. This second term is proportional the equilibrium magnetic twist, $\alpha(\mathbf{x})$, so evidently equilibrium twist of sufficient magnitude is necessary for instability. It multiplies a factor

$$(\hat{\mathbf{w}}\times\mathbf{B}_0)^* \cdot \nabla\times(\hat{\mathbf{w}}\times\mathbf{B}_0) \quad ,$$

which will change sign if the factor $\hat{\mathbf{w}}(\mathbf{x}) \times \mathbf{B}_0(\mathbf{x})$ undergoes a mirror reflection: this would change the sign of the curl. This suggests that the equilibrium will be unstable to perturbations of a particular *handedness* and stable to perturbations of the other handedness. The unstable handedness will depend on the sign of α, since changing that along with the mirror reflection will leave the destabilizing term unchanged.

The destabilizing term involves a single derivative of $\hat{\mathbf{w}}(\mathbf{x})$ while the stabilizing term involves two. It is for this reason that eigenfrequencies in the Eikenol limit are purely real. Unstable modes must have spatial variations comparable to, or longer than, the equilibrium field.

12C Instability of a cylindrical constant-α field — the kink mode

We illustrate the stability analysis using a constant-α field in cylindrical geometry

$$\mathbf{B}_0 \;=\; B_{00}\Big[\, J_0(\alpha r)\,\hat{\mathbf{z}} \;+\; J_1(\alpha r)\,\hat{\boldsymbol{\phi}}\,\Big] \quad , \tag{12.25}$$

where J_m is the Bessel function. We consider this field inside a cylindrical domain bounded by a rigid conductor at $r = a$ and periodic in z with period L. This places no restriction

on α but will play a role in retricting allowed perturbations. Because α is constant it is possible to evaluate W for perturbations which take the form

$$\hat{\mathbf{w}} \times \mathbf{B}_0 \;=\; \mathbf{Q}(\mathbf{x}) \;+\; \nabla\psi \quad , \tag{12.26}$$

where $\mathbf{Q}(\mathbf{x})$ has the property

$$\nabla \times \mathbf{Q} \;=\; \mu\,\mathbf{Q} \quad , \tag{12.27}$$

for some constant μ. Putting this into eq. (12.22) gives

$$\begin{aligned} W\{\hat{\mathbf{w}}\} &= \frac{1}{8\pi}\mu(\mu-\alpha)\int_{\mathcal{V}} |\mathbf{Q}|^2 d^3x \;-\; \frac{\alpha\mu}{8\pi}\int_{\mathcal{V}} \nabla\psi^*\cdot\mathbf{Q}\, d^3x \\ &= \frac{1}{8\pi}\mu(\mu-\alpha)\int_{\mathcal{V}} |\mathbf{Q}|^2 d^3x \;-\; \frac{\alpha\mu}{8\pi}\oint_{\partial\mathcal{V}} \psi^*\,\mathbf{Q}\cdot\, d\mathbf{a} \;+\; \frac{\alpha\mu}{8\pi}\int_{\mathcal{V}} \psi^*\,\underbrace{(\nabla\cdot\mathbf{Q})}_{=0}\, d^3x \\ &= \frac{1}{8\pi}\mu(\mu-\alpha)\int_{\mathcal{V}} |\mathbf{Q}|^2 d^3x \end{aligned} \tag{12.28}$$

after dropping the surface integral because $Q_r = 0$ as explained below. The integral of $|\mathbf{Q}|^2$ is clearly positive, but if the coefficient $\mu(\mu-\alpha)$ is negative the equilibrium is unstable. This will happen for $0<\mu<\alpha$, or $\alpha<\mu<0$ depending on the sign of α. We thus see that the sign of μ, which specifies the handedness of $\nabla\times(\hat{\mathbf{w}}\times\mathbf{B}_0)$, must match that of α, as predicted above.

The value of μ is fixed by the requirement that $\mathbf{Q}(\mathbf{x})$ satisfy (12.27) within the $a\times L$ cylinder. Taking a second curl yields the vector Helmholtz equation

$$\nabla^2\mathbf{Q} \;=\; -\,\mu^2\mathbf{Q} \quad . \tag{12.29}$$

Because the domain is invariant under translation in z and rotation of ϕ we may propose solutions $\sim e^{ikz+im\phi}$. The periodic boundary conditions demand that

$$k \;=\; k_n \;=\; \frac{2\pi n}{L} \qquad , \qquad n = 0,\,\pm1,\,\pm2,\,\cdots \quad . \tag{12.30}$$

Solutions of the vector Helmholtz equation with this symmetry, sometimes called *Chandrasehkar-Kendall functions* are

$$\mathbf{Q}(r,\phi,z) \;=\; \left[J_m\,\hat{\mathbf{z}} \;-\; \frac{k_n J_{m+1} + (\mu+k_n)J'_m}{\sqrt{\mu^2-k_n^2}}\,\hat{\boldsymbol{\phi}} \right. \tag{12.31}$$

$$\left. +\; i\frac{\mu J_{m+1} + (k_n+\mu)J'_m}{\sqrt{\mu^2-k_n^2}}\,\hat{\mathbf{r}} \right] e^{ik_n z+im\phi} \tag{12.32}$$

where all Bessel functions and their derivatives are evaluated at $r\sqrt{\mu^2-k_n^2}$. The $\hat{\mathbf{z}}$ component is a solution of the scalar Helmholtz equation because the vector Laplacian is the same as the scalar Laplacian of any cartesian component; the other components are more complex.

The spectrum of μ is determined by the boundary conditions on $\mathbf{Q}$ which is related to $\hat{\mathbf{w}}$ through eq. (12.26). This involves $\nabla\psi$ about which we do not know. That factor is essential in order that

$$\mathbf{B}_0 \cdot (\hat{\mathbf{w}} \times \mathbf{B}_0) \;=\; \mathbf{B}_0 \cdot \mathbf{Q} \;+\; \mathbf{B}_0 \cdot \nabla\psi \;=\; 0 \quad . \tag{12.33}$$

As a result of this condition ψ will also $\sim e^{ik_n z + im\phi}$ and

$$\nabla\psi \;=\; \frac{\partial\psi}{\partial r}\hat{\mathbf{r}} \;+\; i\left[\frac{m}{r}\,\hat{\boldsymbol{\phi}} \;+\; k_n\hat{\mathbf{z}}\right]\psi \;=\; \frac{\partial\psi}{\partial r}\hat{\mathbf{r}} \;+\; i\boldsymbol{\kappa}\,\psi \quad . \tag{12.34}$$

A notable exception occurs were we to choose $m = n = 0$ so ψ depends on r alone. According to eq. (12.33), there is no displacement capable of producing this change unless $\mathbf{B}_0 \cdot \mathbf{Q} = 0$ everywhere, which will not occur in general. We henceforth exclude the particular case $m = n = 0$.

The radial displacement can be isolated using the components

$$(\hat{\mathbf{r}} \times \boldsymbol{\kappa}) \cdot (\hat{\mathbf{w}} \times \mathbf{B}_0) \;=\; \hat{w}_r(\boldsymbol{\kappa} \cdot \mathbf{B}_0) \;=\; (\hat{\mathbf{r}} \times \boldsymbol{\kappa}) \cdot \mathbf{Q} \quad . \tag{12.35}$$

The outer boundary, $r = a$, is rigid and so $\hat{w}_r$ must vanish there, meaning

$$\begin{aligned}
(\hat{\mathbf{r}} \times \boldsymbol{\kappa}) \cdot \mathbf{Q} \;&=\; \frac{m}{a}J_m \;+\; \frac{k_n^2 J_{m+1} + k_n(\mu + k_n)J'_m}{\sqrt{\mu^2 - k_n^2}} \;=\; 0 \\
&=\; \frac{\mu}{\sqrt{\mu^2 - k_n^2}}\Big[\,\mu J_{m+1} + (k_n + \mu)J'_m\,\Big] \\
&=\; \frac{\mu}{2\sqrt{\mu^2 - k_n^2}}\Big[\,(\mu - k_n)J_{m+1} + (k_n + \mu)J_{m-1}\,\Big]
\end{aligned} \tag{12.36}$$

where Bessel functions are evaluated at $a\sqrt{\mu^2 - k_n^2}$. It is evident from the second line that Q_r also vanishes at the outer boundary (not by coincidence) which justifies dropping the surface term in eq. (12.28).

The outer boundary condition leads to expression (12.36) which is a transcendental equaition for μ of the form

$$\begin{aligned}
S_{m,n}(\mu a) \;&=\; \sqrt{\frac{\mu a - nk_1 a}{\mu a + nk_1 a}}\,J_{m+1}\left(\sqrt{\mu^2 a^2 - n^2 k_1^2 a^2}\right) \\
&\qquad + \sqrt{\frac{\mu a + nk_1 a}{\mu a - nk_1 a}}\,J_{m-1}\left(\sqrt{\mu^2 a^2 - n^2 k_1^2 a^2}\right) \;=\; 0
\end{aligned} \tag{12.37}$$

The function $S_{m,n}$ depends only on the the aspect ratio of the cylindrical domain: $k_1 a = 2\pi a/L$. For a given aspect ratio the functions of μa can be plotted for each choice of integers m and n to find the spectrum of values $\mu_{m,n}(a/L)$. This is done in fig. 12.1 for the case $a/L = 1/4$.

Instability of the equilibrium will occur if $0 < \mu_{m,n} < \alpha$ for any pair of indices m and n. It is evident from the figure that the smallest such value occurs for $m = n = 1$ at $\mu a = 3.156$. This normal mode is called the *kink mode* due to its helical perturbation to the current channel. It turns out to be true for all aspect ratios that the kink mode has

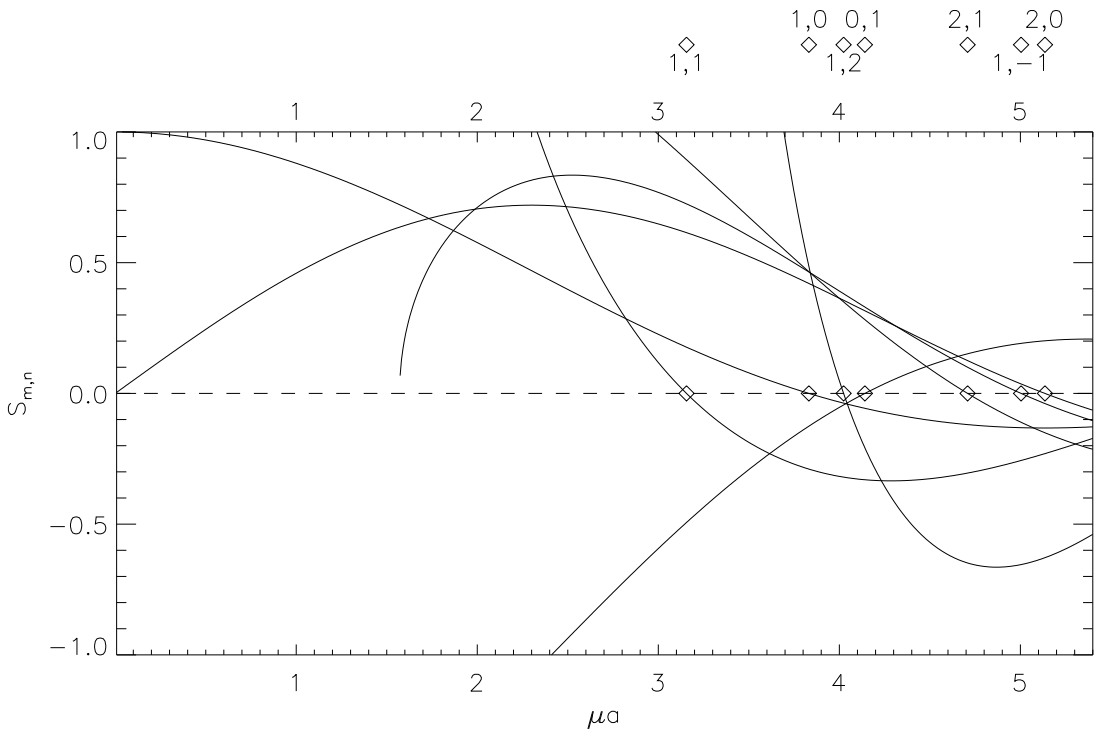

Figure 12.1: Plots of the curves $S_{m,n}(\mu a)$ in a cylinder with aspect ratio $a/L = 1/4$. Curves shown are for $(m,n) = (1,1)$, $(1,0)$, $(1,2)$, $(0,1)$, $(2,1)$, $(1,-1)$, $(2,0)$. Their intersection with zero is markes with a diamond and this location is reproduces, accompanied by a label, above the graphs. The left-most intersection is $\mu a = 3.156$ for $(m,n) = (1,1)$.

the lowest value of μ and is therefore the mode to which the equilibrium is first unstable. Instability will occur when $\alpha > \mu_{1,1}$ meaning

$$\alpha \;\geq\; \frac{3.156}{a} \quad , \tag{12.38}$$

for the $1:4$ aspect ratio cylinder. As predicted above, instability requires the twist α to be larger than the inverse of some characteristic length. The length is evidently the radius of the cylinder and when αa exceeds a critical value the equilibrium becomes unstable.

12D Energy and stability

The total energy of a magnetic field $\mathbf{B}(\mathbf{x})$ within a region $\mathcal{V}$ is

$$W_M\{\mathbf{B}(\mathbf{x})\} \;=\; \frac{1}{8\pi}\int_{\mathcal{V}} |\mathbf{B}(\mathbf{x})|^2\, d^3x \quad . \tag{12.39}$$

To find the field with the minimum energy we demand that this energy be stationary under variation $\mathbf{B}(\mathbf{x}) \to \mathbf{B}(\mathbf{x}) + \delta\mathbf{B}(\mathbf{x})$. This variation is performed so as to preserve the normal magnetic field, $\hat{\mathbf{n}}\cdot\mathbf{B}$ at the boundary $\partial\mathcal{V}$: we require $\hat{\mathbf{n}}\cdot\delta\mathbf{B} = 0$ there. We must also

perform the variation subject to the point-wise constraint $\nabla \cdot \mathbf{B} = 0$. We do this using an undetermined multiplier, $\lambda(\mathbf{x})$, which varies in space. We therefore vary

$$W_M\{\mathbf{B}(\mathbf{x})\} \;=\; \frac{1}{8\pi}\int\limits_{\mathcal{V}} \Big[\, |\mathbf{B}(\mathbf{x})|^2 + \lambda(\mathbf{x})\nabla\cdot\mathbf{B}\,\Big]\, d^3x \quad . \tag{12.40}$$

The first variation of this is

$$\begin{aligned}
\delta W_M &= \frac{1}{8\pi}\int\limits_{\mathcal{V}} \Big[\, 2\mathbf{B}\cdot\delta\mathbf{B} + \lambda(\mathbf{x})\nabla\cdot\delta\mathbf{B}\,\Big]\, d^3x \\
&= \frac{1}{8\pi}\int\limits_{\mathcal{V}} \Big[\, 2\mathbf{B} - \nabla\lambda\,\Big]\cdot\delta\mathbf{B}\, d^3x \;+\; \underbrace{\frac{1}{8\pi}\oint\limits_{\partial\mathcal{V}} \lambda\,\delta\mathbf{B}\cdot d\mathbf{a}}_{=0} \quad ,
\end{aligned} \tag{12.41}$$

since $\hat{\mathbf{n}}\cdot\delta\mathbf{B} = 0$ on $\partial\mathcal{V}$. The perturbation $\delta\mathbf{B}(\mathbf{x})$ is otherwise unconstrained, owing to the undetermined multiplier. Thus the only way for δW_M to vanish is for the factor in square brackets to vanish everywhere. This means

$$\mathbf{B} \;=\; 2\nabla\lambda \quad\Longrightarrow\quad \nabla\times\mathbf{B} \;=\; 0 \quad . \tag{12.42}$$

This is the equation for a potential field, whose solution is unique. The extremum must therefore be a minimum. *A potential magnetic field is the field with minimum magnetic energy.* The scalar potential is related to the undetermined Lagrange multiplier which imposes the constraint $\nabla \cdot \mathbf{B} = 0$.

It is not possible to produce every possible perturbation $\delta\mathbf{B}(\mathbf{x})$ within a highly conducting plasma. This is because the field must be changed by a velocity field $\mathbf{u}(\mathbf{x})$ acting through the ideal induction equation

$$\frac{\partial\mathbf{B}}{\partial t} \;=\; \nabla\times(\mathbf{u}\times\mathbf{B}) \quad . \tag{12.43}$$

This equation permits changes to the magnetic field which preserve every single fleld line in tact — this is the frozen field line thereom which follows from eq. (12.43). This motion can achieve a small displacement $\mathbf{x}\to\mathbf{x}+\delta\boldsymbol{\xi}(\mathbf{x})$ over a time δt with a steady flow

$$\mathbf{u}(\mathbf{x}) \;=\; \frac{1}{\delta t}\delta\boldsymbol{\xi}(\mathbf{x}) \quad . \tag{12.44}$$

This will change the magnetic field to

$$\begin{aligned}
\mathbf{B}(\mathbf{x}) &= \mathbf{B}_0(\mathbf{x}) \;+\; \delta t\,\frac{\partial\mathbf{B}}{\partial t}\bigg|_{t=0} \;+\; \frac{\delta t^2}{2}\,\frac{\partial^2\mathbf{B}}{\partial t^2}\bigg|_{t=0} \;+\; \cdots \\
&= \mathbf{B}_0 + \nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0) + \tfrac{1}{2}\nabla\times\Big[\,\delta\boldsymbol{\xi}\times\nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\,\Big] + \cdots \quad .
\end{aligned} \tag{12.45}$$

The second term, linear in $\delta\boldsymbol{\xi}$,

$$\delta\mathbf{B}(\mathbf{x}) \;=\; \nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0) \quad , \tag{12.46}$$

is the first order the perturbation to the magnetic field created by displacement $\delta\boldsymbol{\xi}(\mathbf{x})$. It automatically satisfies the constraint $\nabla\cdot\delta\mathbf{B}=0$, so the Lagrange multiplier is unnecessary. For any choice of the vector field, $\delta\boldsymbol{\xi}(\mathbf{x})$, this change in magnetic field will preserve all field lines. It is therefore a subset of the arbitrary perturbations used in eq. (12.41).

The first variation of the magnetic field with this more limited perturbation is

$$\begin{aligned}
\delta W_M &= \frac{1}{8\pi}\int_{\mathcal{V}} 2\,\mathbf{B}_0\cdot\delta\mathbf{B}\,d^3x \;=\; \frac{1}{4\pi}\int_{\mathcal{V}}\mathbf{B}_0\cdot\nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\,d^3x \\
&= -\frac{1}{4\pi}\oint_{\partial\mathcal{V}}\mathbf{B}_0\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\cdot\,d\mathbf{a}+\frac{1}{4\pi}\int_{\mathcal{V}}(\nabla\times\mathbf{B}_0)\cdot(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\,d^3x \\
&= \frac{1}{4\pi}\int_{\mathcal{V}}\Big[\,\mathbf{B}_0\times(\nabla\times\mathbf{B}_0)\,\Big]\cdot\delta\boldsymbol{\xi}(\mathbf{x})\;d^3x\quad, \qquad (12.47)
\end{aligned}$$

after discarding surface terms since $\delta\boldsymbol{\xi}=0$ there. For this to vanish under all possible displacements, $\delta\boldsymbol{\xi}(\mathbf{x})$, it is necessary that the factor in square brackets vanish. This requires that

$$(\nabla\times\mathbf{B}_0)\times\mathbf{B}_0 \;=\; 0\quad; \qquad (12.48)$$

the magnetic field must be force free. A force-free magnetic field is an extremum of the magnetic energy under arbitrary perturbations, (12.46), which preserve magnetic field lines: so-called *ideal perturbations*. Since there are a more limited set of changes than the arbitrary variations in eq. (12.41), the force-free field will have higher magnetic energy than the potential field.

The force-free field is an extremum of the energy. Whether it is a minimum, maximum or saddle depends on the second variation — the expansion of W_M to second order,

$$\delta^2 W_M \;=\; \frac{1}{8\pi}\int_{\mathcal{V}}\Big[\,\delta\mathbf{B}\cdot\delta\mathbf{B}+2\mathbf{B}_0\cdot\delta^2\mathbf{B}_0\Big]\,d^3x\quad, \qquad (12.49)$$

where $\delta^2\mathbf{B}$ is the contribution quaradic in $\delta\boldsymbol{\xi}$ in eq. (12.45). Inserting this and $\delta\mathbf{B}$ gives

$$\begin{aligned}
\delta^2 W_M &= \frac{1}{8\pi}\int_{\mathcal{V}}\Big\{\,\Big|\nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\Big|^2+\;\mathbf{B}_0\cdot\nabla\times\Big[\delta\boldsymbol{\xi}\times\nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\Big]\,\Big\}\,d^3x \\
&= \frac{1}{8\pi}\int_{\mathcal{V}}\Big|\nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\Big|^2 d^3x-\frac{1}{8\pi}\oint_{\partial\mathcal{V}}\mathbf{B}\times\Big[\delta\boldsymbol{\xi}\times\nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\Big]\cdot\,d\mathbf{a} \\
&\quad+\frac{1}{8\pi}\int_{\mathcal{V}}(\nabla\times\mathbf{B}_0)\cdot\Big[\delta\boldsymbol{\xi}\times\nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\Big]\,d^3x \\
&= \frac{1}{8\pi}\int_{\mathcal{V}}\Big\{\,\Big|\nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\Big|^2-\Big[\delta\boldsymbol{\xi}\times(\nabla\times\mathbf{B}_0)\Big]\cdot\nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\,\Big\}\,d^3x \\
&= \frac{1}{8\pi}\int_{\mathcal{V}}\Big\{\,\Big|\nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\Big|^2-\alpha\Big(\delta\boldsymbol{\xi}\times\mathbf{B}_0\Big)\cdot\nabla\times(\delta\boldsymbol{\xi}\times\mathbf{B}_0)\,\Big\}\,d^3x\quad. \qquad (12.50)
\end{aligned}$$

This is the same quadratic form as $W\{\delta\boldsymbol{\xi}\}$ in eq. (12.22). In this case the displacement, $\delta\boldsymbol{\xi}(\mathbf{x})$, is a real function, so complex conjugation is unnecessary. If there is a displacement

which makes this second variation negative then the equilibrium is not a minimum: it has at least one direction along which it is maximum.

It is for this reason that such an equilibrium is unstable. We also see now that the linear operator $\mathbf{F}$ is self-adjoint because it arose from the second variation of a scalar potential: it is a second derivative. It is directly analogous to the second derivative of any potential energy which, if it is negative, produces an exponentially growing solution in the linearized dynamics. The exponentially growing kinetic energy is a reflection of a decrease in the magnetic energy (i.e. potential energy). The magnetic perturbation, $\mathbf{B}_1(\mathbf{x}, t)$, grows exponentially in time, but is structured so as to *decrease* the magnetic energy.

Chapter 13

MHD shocks

13A Conservation laws

The MHD equations in conservative form are

$$\partial_t(\rho) = -\nabla\cdot(\rho\mathbf{u}) \tag{13.1}$$

$$\partial_t(\rho\mathbf{u}) = -\nabla\cdot\left(\rho\,\mathbf{uu} + p\,\underline{\underline{\mathrm{I}}} + \frac{B^2}{8\pi}\,\underline{\underline{\mathrm{I}}} - \frac{\mathbf{BB}}{4\pi}\right) + \nabla\cdot\underline{\underline{\sigma}} \tag{13.2}$$

$$\begin{aligned}\partial_t e &= -\nabla\cdot\left[\tfrac{1}{2}\rho\,u^2\mathbf{u} + \frac{\gamma\,p\mathbf{u}}{\gamma-1} - \frac{1}{4\pi}(\mathbf{u}\times\mathbf{B})\times\mathbf{B}\right] \\ &\quad + \nabla\cdot\left[\underline{\underline{\sigma}}\cdot\mathbf{u} - \frac{\eta}{4\pi}(\nabla\times\mathbf{B})\times\mathbf{B}\right]\end{aligned} \tag{13.3}$$

$$\partial_t\mathbf{B} = \nabla\times(\mathbf{u}\times\mathbf{B}) - \nabla\times(\eta\nabla\times\mathbf{B}) \ , \tag{13.4}$$

where the energy density

$$e = \tfrac{1}{2}\rho u^2 + \frac{p}{\gamma-1} + \frac{B^2}{8\pi} \ , \tag{13.5}$$

includes kinetic, thermal and magnetic energy, and the viscous stress tensor $\underline{\underline{\sigma}}$ is proportional to derivatives, $\partial u_i/\partial x_j$. We transform to a reference frame moving normal to the shock surface and assume that within that frame the solution appears stationary: $\partial_t = 0$. The first three equations then take the form

$$0 = -\nabla\cdot\mathbf{\Gamma} + S(\mathbf{x}) \ , \tag{13.6}$$

where $\mathbf{\Gamma}$ is a flux and S is a source. In each case the source term is a divergence of a related flux term, but one depending on diffusive effects and spatial derivatives which will vanish where the solution is uniform. We integrate the equations over a pillbox, $\mathcal{P}$, straddling the shock transition, with two faces located in uniform regions 1 and 2. These faces have area A and surface normals parallel and anti-parallel to the shock normal $\hat{\mathbf{n}}$, defined to face into the flow — directed toward pre-shock region 1.[1] Due to the assumed one-dimensional

[1] This makes $u_n = \hat{\mathbf{n}}\cdot\mathbf{u} < 0$ as it was for the hydrodynamic case. This definition is motivated by the direction of the *shock* rather than the direction of the flows relative to the shock.

symmetry of the shock the contributions of the pillbox's sides will vanish so that all three conservation laws (this excludes the induction eq. [13.4], discussed further on) take the form

$$\frac{1}{A}\oint_{\partial\mathcal{P}} \mathbf{\Gamma}\cdot d\mathbf{a} \;=\; \mathbf{\Gamma}_1\cdot\hat{\mathbf{n}} \;-\; \mathbf{\Gamma}_2\cdot\hat{\mathbf{n}} \;=\; \Gamma_{n,1}-\Gamma_{n,2} \;=\; [\![\,\Gamma_n\,]\!] \;=\; 0 \quad , \tag{13.7}$$

since the source terms all integrate to zero. In other words the normal component of each flux must be preserved through the shock.

We must augment the conservation laws with the explicit constraint $\nabla\cdot\mathbf{B}=0$. This is typically applied only as an initial condition, but we have neglected time evolution, so it assumes the same importance as the other equations. Integrating this over the pillbox $\mathcal{P}$ results in

$$[\![\,B_n\,]\!] \;=\; 0 \quad , \tag{13.8}$$

demanding that the normal component of the magnetic field be the same on both sides of the shock; we hereafter refer to B_n without subscript. Integrating the continuity equation, eq. (13.1), over the pillbox yields the conservation of mass

$$[\![\,\rho\,u_n\,]\!] \;=\; 0 \quad , \tag{13.9}$$

which is identical to the hydrodynamic version. The momentum flux whose divergence appears on the right of eq. (13.2), is a second-rank tensor $\underline{\underline{\Gamma}}_p$. Contracting this with $\hat{\mathbf{n}}$, as in the generic example (13.7), will lead to a jump in the vector $\left[\!\left[\,\hat{\mathbf{n}}\cdot\underline{\underline{\Gamma}}_p\,\right]\!\right] = 0$, which is actually three conditions. It is useful to break this, and all other vectors, into components tangential and normal to the shock surface: $\mathbf{B}=\mathbf{B}_t+B_n\hat{\mathbf{n}}$, etc. Doing so yields separate conservation laws for tangential and normal momentum

$$\left[\!\left[\,\rho\,u_n\mathbf{u}_t-\frac{B_n}{4\pi}\mathbf{B}_t\,\right]\!\right] \;=\; (\rho\,u_n)\,[\![\,\mathbf{u}_t\,]\!]-\frac{B_n}{4\pi}[\![\,\mathbf{B}_t\,]\!] \;=\; 0 \tag{13.10}$$

$$\left[\!\left[\,\rho\,u_n^2+p+\frac{|\mathbf{B}_t|^2}{8\pi}-\frac{B_n^2}{8\pi}\,\right]\!\right] \;=\; \left[\!\left[\,\rho\,u_n^2+p+\frac{|\mathbf{B}_t|^2}{8\pi}\,\right]\!\right] \;=\; 0 \quad . \tag{13.11}$$

Factors which are themselves conserved, such as B_n or (ρu_n), have been taken outside the jump.

The special case $B_n = 0$, known as a perpendicular shock, is very similar to a hydrodynamic shock with additional pressure contributions due to $|\mathbf{B}_t|^2/8\pi$. We leave this simple extension to the reader and pursue the more challenging case of oblique MHD shocks, $B_n \neq 0$.

13B The de Hoffman-Teller reference frame

In the absence of a magnetic field, or in a perpendicular shock ($B_n = 0$), tangential momentum conservation, eq. (13.10), requires $[\![\,\mathbf{u}_t\,]\!]=0$, so the tangential velocity is the same on both sides of the shock. It is then possible to perform a second transformation to a reference frame where $\mathbf{u}_t'=0$ and the flow is purely normal to the shock. This makes hydrodynamic shocks (and perpendicular shocks) essentially one dimensional. In the presence

of magnetic field the tangential velocity may differ across the shock and cannot, in general, be transformed away.

What is done instead is to transform to the so-called *de Hoffman-Teller* reference frame

$$\mathbf{u}' \;=\; \mathbf{u} \;-\; \left(\mathbf{u}_{t,1} \;-\; \frac{u_{n,1}}{B_n}\,\mathbf{B}_{t,1}\right) \quad . \tag{13.12}$$

In this frame the upstream velocity

$$\mathbf{u}'_1 \;=\; u_{n,1}\hat{\mathbf{n}} \;+\; \frac{u_{n,1}}{B_n}\,\mathbf{B}_{t,1} \;=\; \frac{u_{n,1}}{B_n}\,\mathbf{B}_1 \quad , \tag{13.13}$$

is perfectly aligned with the upstream magnetic field. The upstream electric field in this reference frame is then

$$\mathbf{E}'_1 \;=\; -\frac{1}{c}\mathbf{u}'_1 \times \mathbf{B}_1 \;=\; 0 \quad ; \tag{13.14}$$

which is an alternative definition of the de Hoffman-Teller reference frame.

The induction equation does not take the form a traditional conservation law. It is integrated over an Amperian loop straddling the shock. Since the legs within regions 1 and 2 can be oriented in any tangential direction it is necessary that

$$[\![\, (\mathbf{u} \times \mathbf{B})_t \,]\!] \;=\; -\,c[\![\, \mathbf{E}_t \,]\!] \;=\; 0 \quad . \tag{13.15}$$

In the de Hoffman-Teller frame, $\mathbf{E}'_{t.1} = 0$, so it follows that

$$\mathbf{E}'_{t,2} \;=\; 0 \;=\; (\mathbf{u}'_2 \times \mathbf{B}_2)_t \;=\; B_n\left(\mathbf{u}'_{t,2} - \frac{u_{n,2}}{B_n}\mathbf{B}_{t,2}\right) \times \hat{\mathbf{n}} \quad . \tag{13.16}$$

From this it follows that $\mathbf{u}'_2$ is parallel to $\mathbf{B}_2$, so the flow is parallel to the field everywhere. In the de Hoffman-Teller frame the shock consists of flow following field lines as they bend. The angles of the field relative to the shock normal $\hat{\mathbf{n}}$ are denoted θ_1 and θ_2.

13C Co-planarity

The Alfvén mach number defined in the de Hoffman-Teller frame matches the value defined with respect to the normal components of both flow and magnetic field

$$M_{\rm A} \;=\; \frac{|\mathbf{u}'|}{v_{\rm A}} \;=\; \frac{u_n}{B_n/\sqrt{4\pi\rho}} \quad . \tag{13.17}$$

This is the only Mach number to which we need refer and we henceforth drop the subscript A. Squaring the expression leads to the related expression

$$\rho\, u_n^2 \;=\; \frac{B_n^2}{4\pi}\, M^2 \quad . \tag{13.18}$$

Using this in eq. (13.10) gives a relation

$$\left[\!\!\left[\, \rho u_n^2 \frac{\mathbf{u}'_t}{u_n} - \frac{B_n}{4\pi}\mathbf{B}_t \,\right]\!\!\right] \;=\; \frac{B_n^2}{4\pi}\left[\!\!\left[\, M^2\,\frac{\mathbf{u}'_t}{u_n} - \frac{\mathbf{B}_t}{B_n} \,\right]\!\!\right] \;=\; \frac{B_n}{4\pi}\left[\!\!\left[\, (M^2-1)\mathbf{B}_t \,\right]\!\!\right] \;=\; 0 \quad . \tag{13.19}$$

Equation (13.19) requires that

$$(M_1^2 - 1)\mathbf{B}_{t,1} \;=\; (M_2^2 - 2)\mathbf{B}_{t,2} \quad , \tag{13.20}$$

so, with one possible exception (to which we return below) the vectors $\mathbf{B}_{t,1}$ and $\mathbf{B}_{t,2}$ lie within the same line within the shock plane. This means in turn that the three vectors $\mathbf{B}_1$, $\mathbf{B}_2$, and $\hat{\mathbf{n}}$ all lie in a single plane. Adding eq. (13.10) shows that $\Delta\mathbf{u} = \mathbf{u}_1 - \mathbf{u}_2$ also lies in this same plane.

That these four vectors lie in a single plane, a requirement known as *co-planarity*, means that MHD shocks all have a particular structure. A plane is naturally defined by the upstream magnetic field $\mathbf{B}_1$ and the shock normal, $\hat{\mathbf{n}}$. It seems that the shock must bend the magnetic field *within this plane*, to its new direction $\mathbf{B}_2$. This bending creates a surface current density within the shock (distributed over the shock's finite thickness)

$$\mathbf{K} \;=\; -\frac{c}{4\pi}[\![\,\mathbf{B}\,]\!] \times \hat{\mathbf{n}} \;=\; -\frac{c}{4\pi}[\![\,\mathbf{B}_t\,]\!] \times \hat{\mathbf{n}} \quad , \tag{13.21}$$

perpendicular to the plane. The magnetic field crossing the shock produces a tangential Lorentz force density

$$\left(\frac{1}{c}\mathbf{K} \times \mathbf{B}\right)_t \;=\; \frac{B_n}{c}\mathbf{K} \times \hat{\mathbf{n}} \;=\; \frac{B_n}{4\pi}[\![\,\mathbf{B}_t\,]\!] \quad , \tag{13.22}$$

in the same plane. This force changes the tangential momentum in accordance with eq. (13.10). The tangential velocity shear across the shock, $[\![\,\mathbf{u}_t\,]\!]$, is what bends the magnetic field in the first place. Thus the entire process must occur within a plane.

The notable exception to co-planarity occurs when either $M_1^2 = 1$ or $M_2^2 = 1$. We show later that $M_2^2 = 1$ requires that $M_1^2 = 1$ as well. In this case, known as a *rotational discontinuity*, $\mathbf{B}_{t,2}$ may be in a direction different from $\mathbf{B}_{t,1}$. In the second exception, $M_1^2 = 1$ and $M_2^2 < 1$, known as a *switch-off shock*, eq. (13.20) demands $\mathbf{B}_{t,2} = 0$ — the tangential field is "switched off" by the shock. Since $\mathbf{B}_2$ is then parallel to $\hat{\mathbf{n}}$, the switch-off shock is not actually an exception to co-planarity. The rotational discontinuity is an exception to which we return below. In the mean time we will consider only the more general case which, due to co-planarity, can be treated as a two-dimensional problem.

13D Jump conditions

Under the condition of co-planarity, $\mathbf{B}_1$, $\hat{\mathbf{n}}$ and $\mathbf{B}_2$ all lie in a single plane. The magnetic vector components in that plane are B_t and B_n. In the de Hoffman-Teller frame the magnetic field and velocity make the same angle with the normal. We will call this angle θ_1 on the upstream side and and θ_2 on the downstream side. The angle is

$$\frac{B_t}{B_n} \;=\; \frac{u'_t}{u_n} \;=\; \tan\theta \quad . \tag{13.23}$$

The only other dimensionless parameter is the plasma β

$$\beta \;=\; \frac{8\pi p}{|\mathbf{B}|^2} \;=\; \frac{8\pi p}{B_n^2}\frac{B_n^2}{|\mathbf{B}|^2} \;=\; \frac{8\pi p}{B_n^2}\cos^2\theta \quad . \tag{13.24}$$

This takes on different values, β_1 and β_2 on the two sides of the shock. The jump conditions can be recast as relations between dimensionless upstream quantities, M_1, θ_1 and β_1 and dimensionless downstream quantities, M_2, θ_2 and β_2. As in the hydrodynamic case, we seek to determine the latter three given the former three.

In terms of these dimensionless quantities eqs. (13.10) and (13.11) are

$$\begin{aligned} 0 &= \left[\!\!\left[\rho u_n^2 \frac{u_t'}{u_n} - \frac{B_n}{4\pi} B_t \right]\!\!\right] = \frac{B_n^2}{4\pi} \left[\!\!\left[M^2 \frac{u_t'}{u_n} - \frac{B_t}{B_n} \right]\!\!\right] \\ &= \frac{B_n^2}{4\pi} \left[\!\!\left[(M^2 - 1) \tan\theta \right]\!\!\right] \qquad (13.25) \\ 0 &= \left[\!\!\left[\frac{B_n^2}{4\pi} M^2 + p + \frac{B_n^2}{8\pi} \frac{B_t^2}{B_n^2} \right]\!\!\right] = \frac{B_n^2}{4\pi} \left[\!\!\left[M^2 + \frac{\beta}{2\cos^2\theta} + \tfrac{1}{2}\tan^2\theta \right]\!\!\right] . \qquad (13.26) \end{aligned}$$

The equation for energy conservation is particularly simple in the de Hoffman-Teller frame since $\mathbf{u}' \times \mathbf{B} = 0$ leaving

$$\left[\!\!\left[\tfrac{1}{2}\rho (u_n^2 + u_t'^2) u_n + \frac{\gamma p u_n}{\gamma - 1} \right]\!\!\right] = 0 . \qquad (13.27)$$

Multiplying this by the constant quantity ρu_n we obtain the relation

$$\begin{aligned} 0 &= \left[\!\!\left[\tfrac{1}{2}\rho^2 u_n^2 (u_n^2 + u_t'^2) + \rho u_n^2 \frac{\gamma p}{\gamma - 1} \right]\!\!\right] = \left[\!\!\left[\tfrac{1}{2}\rho^2 u_n^4 \left(1 + \frac{u_t'^2}{u_n^2}\right) + \frac{B_n^2}{4\pi} M^2 \frac{\gamma p}{\gamma - 1} \right]\!\!\right] \\ &= \frac{B_n^4}{2(4\pi)^2} \left[\!\!\left[M^4 (1 + \tan^2\theta) \; + \; M^2 \frac{\gamma}{\gamma - 1} \frac{\beta}{\cos^2\theta} \right]\!\!\right] \qquad (13.28) \end{aligned}$$

Together the three relations, eqs. (13.25), (13.26) and (13.28) are three constraints on the three dimensionless quantities, M, θ, and β, across the shock. Equation (13.25) can be used to eliminate θ_2

$$\tan\theta_2 = \frac{M_1^2 - 1}{M_2^2 - 1} \tan\theta_1 . \qquad (13.29)$$

This informative relation shows how the magnetic field vector is deflected by the shock. We return to explore its consequences below. First we must use one of the two remaining jump conditions, either (13.26) or (13.28), to express β_2 in terms of M_1, M_2 and θ_1. We then substitute this into the final jump conditions to obtain the explicit relation between M_1 and M_2, in terms only of θ_1 and β_1. The algebra is straightforward, but tedious and unenlightening. We emerge from it at expression (13.34).

Using eq. (13.29) in eq. (13.26) gives

$$\begin{aligned} \frac{\beta_2}{\cos^2\theta_2} &= \frac{\beta_1}{\cos^2\theta_1} + 2(M_1^2 - M_2^2) + \left[1 - \frac{(M_1^2 - 1)^2}{(M_2^2 - 1)^2}\right] \tan^2\theta_1 \\ &= \frac{\beta_1}{\cos^2\theta_1} + 2(M_1^2 - M_2^2) + (M_2^2 - M_1^2) \frac{M_2^2 + M_1^2 - 2}{(M_2^2 - 1)^2} \tan^2\theta_1 \\ &= \frac{\beta_1}{\cos^2\theta_1} + (M_1^2 - M_2^2) \left[2 - \frac{M_2^2 + M_1^2 - 2}{(M_2^2 - 1)^2} \tan^2\theta_1\right] \qquad (13.30) \end{aligned}$$

These may be inserted into (13.28) to obtain an equation for M_2^2. The first term becomes

$$\begin{aligned}\left[\!\left[M^4(1+\tan^2\theta) \right]\!\right] &= M_1^4 - M_2^4 + \left[M_1^4 - M_2^4 \frac{(M_1^2-1)^2}{(M_2^2-1)^2} \right] \tan^2\theta_1 \\ &= (M_1^2 - M_2^2)(M_1^2+M_2^2) + \frac{M_1^4 - 2M_1^4M_2^2 - M_2^4 + 2M_2^4M_2^2}{(M_2^2-1)^2}\tan^2\theta_1 \\ &= (M_1^2-M_2^2)\left[M_1^2 + M_2^2 + \frac{M_1^2+M_2^2-2M_1^2M_2^2}{(M_2^2-1)^2}\tan^2\theta_1 \right] \end{aligned} \tag{13.31}$$

The second term is $\gamma/(\gamma-1)$ times

$$\begin{aligned}\left[\!\left[\frac{\beta M^2}{\cos^2\theta} \right]\!\right] &= (M_1^2-M_2^2)\frac{\beta_1}{\cos^2\theta_1} \\ &\quad - M_2^2(M_1^2-M_2^2)\left[2 - \frac{M_2^2+M_1^2-2}{(M_2^2-1)^2}\tan^2\theta_1 \right] \end{aligned} \tag{13.32}$$

Both (13.31) and (13.32) are proportional to $(M_1^2 - M_2^2)$, so this will be a factor in the sum, (13.28). This means one method of satisfying all conservation laws is for $M_1 = M_2$, meaning no change occurs across the shock – this is the trivial solution. This possibility will be disregarded by dividing through by $(M_1^2 - M_2^2) \neq 0$ leaving an expression for eq. (13.28)

$$\begin{aligned} 0 &= M_1^2 + M_2^2 + \frac{M_1^2+M_2^2-2M_1^2M_2^2}{(M_2^2-1)^2}\tan^2\theta_1 \\ &\quad + \frac{\gamma}{\gamma-1}\left[\frac{\beta_1}{\cos^2\theta_1} - 2M_2^2 + M_2^2\frac{M_2^2+M_1^2-2}{(M_2^2-1)^2}\tan^2\theta_1 \right] \end{aligned} \tag{13.33}$$

Multiplying this by $(\gamma-1)(M_2^2-1)^2$ gives a single expression for the upstream-to-downstream shock relation

$$\begin{aligned} 0 &= F(M_1,M_2) = (M_2^2-1)^2\Big[(\gamma-1)(M_1^2+M_2^2) - 2\gamma M_2^2 \Big] + \frac{\gamma\beta_1}{\cos^2\theta_1}(M_2^2-1)^2 \\ &\quad + \Big[(\gamma-1)(M_1^2+M_2^2-2M_1^2M_2^2) + \gamma M_2^2(M_2^2+M_1^2-2) \Big]\tan^2\theta_1 \\ &= (M_2^2-1)^2\left[-(\gamma+1)M_2^2 + (\gamma-1)M_1^2 + \frac{\gamma\beta_1}{\cos^2\theta_1} \right] \\ &\quad + \Big[\gamma M_2^4 - (\gamma-2)M_2^2M_1^2 - (\gamma+1)M_2^2 + (\gamma-1)M_1^2 \Big]\tan^2\theta_1 \quad . \end{aligned} \tag{13.34}$$

The downstream properties can be determined by solving this equation for M_2.

Equation (13.34) is a relation by which the downstream Mach number, M_2 may be found from upstream quantities M_1, θ_1 and β_1. It is cubic in M_2^2

$$\begin{aligned} F(M_1,M_2) &= -(\gamma+1)M_2^6 + \left[(\gamma-1)M_1^2 + 2(\gamma+1) + \frac{\gamma\beta_1}{\cos^2\theta_1} + \gamma\tan^2\theta_1 \right] M_2^4 \\ &\quad - \left\{ 2(\gamma-1)M_1^2 + \frac{(\gamma+1)+2\gamma\beta_1}{\cos^2\theta_1} + (\gamma-2)M_1^2\tan^2\theta_1 \right\} M_2^2 \\ &\quad + \frac{(\gamma-1)M_1^2 + \gamma\beta_1}{\cos^2\theta_1} = 0 \end{aligned} \tag{13.35}$$

so the solution may not be unique: either one or three real solutions M_2^2 could exist for a single upstream Mach number M_1. It is linear in M_1^2, so M_1^2 can be expressed as a simple, single-valued function (a rational polynomial) of M_2^2. This is the simplest way to graph the relation, as done in fig. 13.1.

From relation (13.18) we find that

$$\frac{M_1^2}{M_2^2} \;=\; \frac{\rho_1 u_{n,1}^2}{\rho_2 u_{n,2}^2} \;=\; \frac{u_{n,1}}{u_{n,2}} \;=\; \frac{\rho_2}{\rho_1} \quad , \tag{13.36}$$

so only cases with $M_1^2 > M_2^2$ will be compressive. It can also be shown (but will not be) that only in those cases does the entropy increase across the shock, $s_2 > s_1$. Only these are allowed shocks. This still leaves values of M_1 for which two or three acceptable downstream Mach numbers are possible ($1 < M_1 < 1.2$ in fig. 13.1).

13E Three MHD shocks

The region of allowed shocks is bounded by the diagonal, $M_1 = M_2$. The boundary points can be found by setting $M_1 = M_2$ in eq. (13.34)

$$\begin{aligned}
0 &= (M_1^2-1)^2\left[-2M_1^2 + \frac{\gamma\beta_1}{\cos^2\theta_1}\right] + 2M_1^2(M_1^2-1)\tan^2\theta_1 \\
&= (M_1^2-1)\left[-2M_1^2(M_1^2-1) + \frac{\gamma\beta}{\cos^2\theta_1}(M_1^2-1) + 2M_1^2\tan^2\theta_1\right] \\
&= \frac{(M_1^2-1)}{\cos^2\theta_1}\left[-2M_1^4\cos^2\theta_1 + \gamma\beta_1\,(M_1^2-1) + 2M_1^2\right] \\
&= -2\frac{(M_1^2-1)}{\cos^4\theta_1}\left[M_1^4\cos^4\theta_1 - \left(\frac{c_{\rm s}^2}{v_{\rm A}^2}+1\right)M_1^2\cos^2\theta_1 + \frac{c_{\rm s}^2}{v_{\rm A}^2}\cos^2\theta_1\right] \quad .
\end{aligned} \tag{13.37}$$

Using $M_1 \simeq M_2$ in eqs. (13.29) and (13.30), shows that, except in the special case $M_1 \simeq M_2 \simeq 1$ (discussed below) the upstream and downstream angle and pressure match across the discontinuity. These intersections represent trivial solutions with identical states on either side of the "discontinuity". The neighborhoods of the diagonal intersections, however, represent small perturbations to this state, which are by definition linear waves — these are *weak shocks*.

Equation (13.37) is cubic in M_1^2, and has three solutions

$$M_1^2 \;=\; 1 \;\Longrightarrow\; M_1^2\cos^2\theta \;=\; \cos^2\theta \tag{13.38}$$

$$M_1^2\cos^2\theta_1 \;=\; \frac{1}{2v_{\rm A}^2}\left[(c_s^2+v_{\rm A}^2) + \sqrt{(c_s^2+v_{\rm A}^2)^2 - 4c_s^2v_{\rm A}^2\cos^2\theta_1}\right] = \frac{v_{\rm fms}^2}{v_{\rm A}^2} \quad , \tag{13.39}$$

$$M_1^2\cos^2\theta_1 \;=\; \frac{1}{2v_{\rm A}^2}\left[(c_s^2+v_{\rm A}^2) - \sqrt{(c_s^2+v_{\rm A}^2)^2 - 4c_s^2v_{\rm A}^2\cos^2\theta_1}\right] = \frac{v_{\rm sms}^2}{v_{\rm A}^2} \quad . \tag{13.40}$$

readily identified with the phase speeds of the three MHD waves: shear Alfvén wave, fast and slow magnetosonic waves respectively. The identification use the phase speed

$$\frac{\omega}{k} \;=\; v_{\rm A}\,M_1\,\cos\theta \;=\; u_1\cos\theta \;=\; u'_{n,1} \quad , \tag{13.41}$$

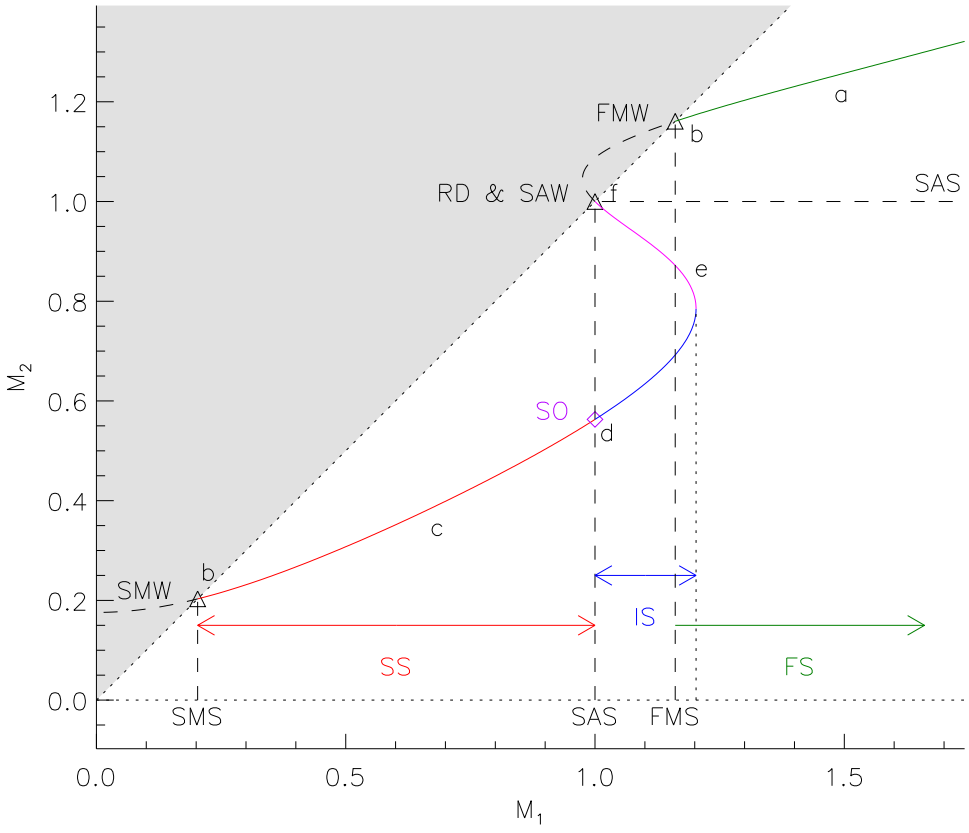

Figure 13.1: Plot of the post-shock Alfvén Mach number, M_2 *vs.* M_1 for the case $\beta_1 = 0.05$, $\theta_1 = 30°$. The grey region, above the diagonal is rarefactory ($\rho_2 < \rho_1$) and thus unphysical. The solid colored curves beneath this are allowed shocks. Triangles mark the linear waves: SMW, slow magnetosonic wave; SAW, shear Alfvén wave; and FMW, fast magnetosonic wave. Horizontal and vertical dashed lines originating from these show the corresponding waves speeds (SMS, SAS and FMS respectively) used to divide the shock curve into its three parts. Arrows show the range of upstream Mach numbers corresponding to the slow shock (SS, red), intermediate shocks (IS, blue and magenta) and fast shock (FS, blue). A diamond marks the so-called switch-off shock (SO), separating the slow shock from the intermediate shock. Letters a–f correspond to the configurations shown in fig. 13.2.

because θ is the angle between the magnetic field vector and the wave vector — the shock surface is a phase front. If we transform into a reference frame where the upstream region is at rest, the shock front moves at speed ω/k: it is a wave. Thus the neighborhoods of the diagonal intersections are linear MHD waves.

The shock curve, $M_2(M_1)$, shown in fig. 13.1, crosses the diagonal ($M_1 = M_2$) at three points corresponding to the three linear waves. It passes from unphysical into the allowed region, $M_2 < M_1$, between the slow mode and shear Alfvén mode, and once again above the fast mode (see fig. 13.1). This means there are two distinct portions of the permitted shock relation. The first of these is traditionally broken at the $M_1 = 1$ point yielding a total of three branches of acceptable shock relation:

$$\begin{aligned} \text{slow shock} &: M_2 < M_1 < 1 \\ \text{intermediate shock} &: M_2 < 1 < M_1 \\ \text{fast shock} &: 1 < M_2 < M_1 \quad . \end{aligned}$$

These portions of the curve, the three different types of MHD shocks, are designated with different colors in fig. 13.1 (two different colors are used to plot the curve through the intermediate shock).

Properties of the three shocks: The ordering of Mach numbers leads, after using eq. (13.29), to an ordering of field angles and field strengths. If we take $\theta_1 > 0$ for concreteness, and use (13.29) to write

$$\frac{B_{t,2}}{B_{t,1}} = \frac{\tan\theta_2}{\tan\theta_1} = \frac{M_1^2 - 1}{M_2^2 - 1} \quad . \tag{13.42}$$

We may translate the Mach number orderings into the relations

$$\begin{aligned} \text{fast shock} &: 0 < \theta_1 < \theta_2 \quad \text{and} \quad |\mathbf{B}_2| > |\mathbf{B}_1| \\ \text{slow shock} &: 0 < \theta_2 < \theta_1 \quad \text{and} \quad |\mathbf{B}_2| < |\mathbf{B}_1| \\ \text{intermediate shock} &: \theta_2 < 0 < \theta_1 \quad . \end{aligned}$$

The three shocks therefore differ in at least one simple characteristic, evident from the top row of fig. 13.2. In an intermediate shock the tangential field reverses direction, while in fast and slow shocks its direction is maintained. A fast shock *strengthens* the magnetic field while a slow shock *weakens* it. Equivalently, the fast shock bends the field *away* from the normal while a slow shock bends it *toward* the normal.

The limiting case separating slow from intermediate shocks is called a switch-off shock, denoted SO in fig. 13.1 (about which, more shortly). This is the point where θ_2 changes sign. The name derives from the fact that the transverse field is switched off ($B_{t,2} = 0$) by bending the downstream field parallel to $\hat{\mathbf{n}}$.

It is evident from fig. 13.1 that within the parameter regime (i.e. M_1 values) admitting an intermediate shock two or three downstream magnetic numbers, M_2, are equally able to satisfy the conservation laws for the same upstream conditions. The multivalued nature is due to the intermediate shock. This ambiguity has made the intermediate shock the subject of considerable uncertainty in the study of MHD shocks; the unambiguous nature of fast and slow shocks make them far more straightforward.

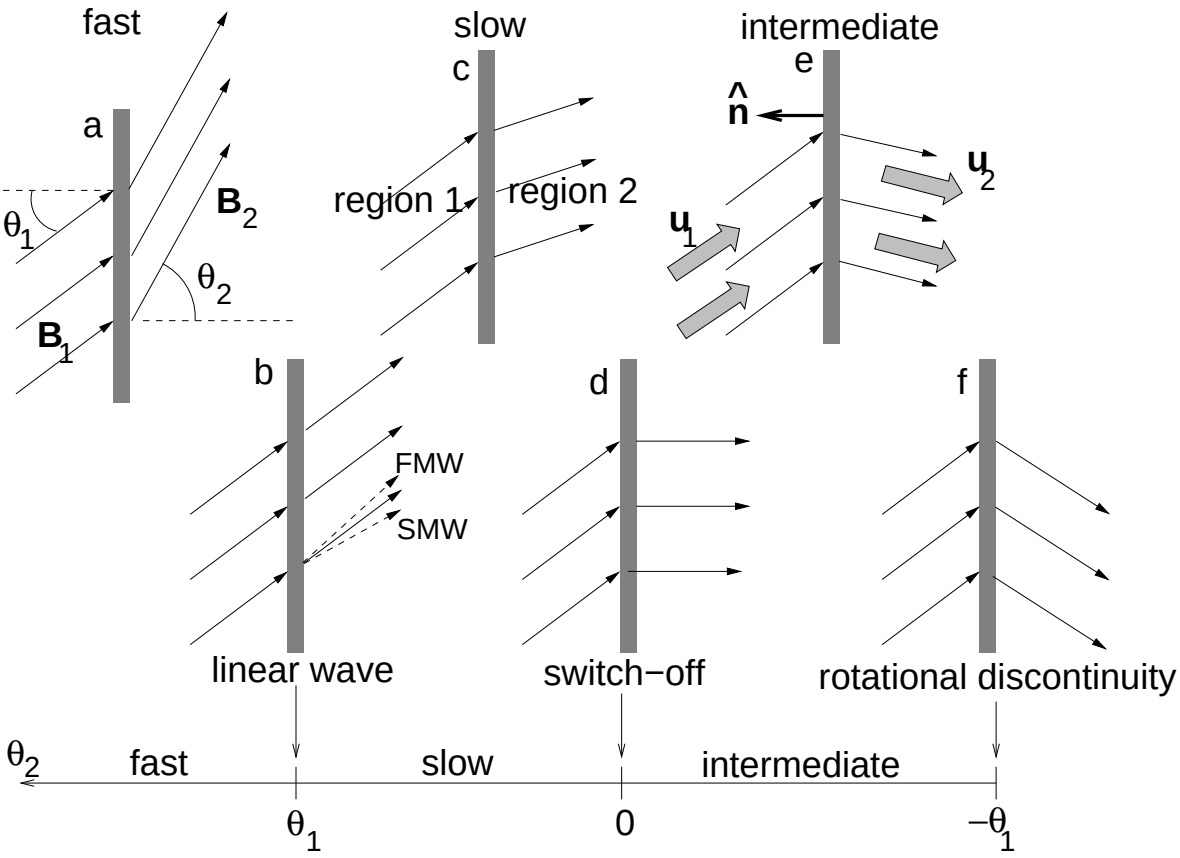

Figure 13.2: The MHD shocks arranged left to right (a–f) in order of decreasing θ_2, for fixed θ_1. The letters correspond to the portions of the shock curve labeled in fig. 13.1 All have a leftward shock normal, $\hat{\mathbf{n}}$, meaning rightward flow, $\mathbf{u}_1$ and $\mathbf{u}_2$. Representative velocity vectors, in the de Hoffman-Teller frame, are shown with shock e. Magnetic field vectors are shown in all cases with pre-shock, $\mathbf{B}_1$, being the same in every case. Values of θ_2 are graphed along the bottom (increasing leftward). The bottom row of cases (b, d and f), are endpoints of the ranges. Case b., labeled "linear wave", represents both the fast magnetosonic wave (FMW, $\theta_2 > \theta_1$) and the slow magnetosonic wave (SMW, $\theta_2 < \theta_1$) at once, with different $\mathbf{B}_2$ vectors suggested by dashed arrows.

We have made use of the fact that $\rho_2 > \rho_1$ for each shock. It also turns out that $p_2 > p_1$, so the plasma pressure is increased as the fluid crosses the shock.[2] The magnetic pressure, however, *increases* across a fast shock but *decreases* across a slow shock. This is consistent with the behavior of fast and slow magnetosonic waves. In a fast wave plasma pressure and magnetic pressure are *in phase*, while in a slow mode they are *out of phase* (one increases when the other decreases). The magnetic pressure also decreases across an intermediate shock, but as shown below this is not directly related to any linear wave.

The three different shocks are defined so that each shock occurs where the upstream flow speed exceeds its relevant linear wave speed. The Mach number used here refers to the Alfvén wave, so the slow shock exceeds the slow magnetosonic wave speed even when $M_1 < 1$. While it is significantly more difficult to show, the downstream flow in each case is *slower* than the corresponding wave speed in that medium: over both the red and blue portions of the curve the downstream flow is slower than the local slow magnetosonic wave speed. The only obvious transition is the intermediate shock from super-Alfvénic ($M_1 > 1$) to sub-Alfvénic ($M_2 < 1$) flow; in this sense it is a shock corresponding to the shear Alfvén mode.

Switch-off Shock: The line $M_1 = 1$ plays the important role of dividing the different kinds of shocks. Setting $M_1 = 1$ in expression (13.34), yields,

$$\begin{aligned} F(M_2,1) &= (M_2^2-1)^2\left[-(\gamma+1)M_2^2 +\gamma-1+\frac{\gamma\beta_1}{\cos^2\theta_1}\right] \\ &\quad +\left[\gamma M_2^4-(2\gamma-1)M_2^2+(\gamma-1)\right]\tan^2\theta_1 \qquad (13.43)\\ &= (M_2^2-1)\Big\{(M_2^2-1)\left[-(\gamma+1)M_2^2+\gamma-1+\gamma\beta_1\sec^2\theta_1\right] \\ &\quad +\left[\gamma M_2^2-(\gamma-1)\right]\tan^2\theta_1\Big\} \\ &= (M_2^2-1)\Big\{-(\gamma+1)M_2^4+\gamma(3+\beta_1\sec^2\theta_1)M_2^2-(\gamma-1+\gamma\beta_1)\sec^2\theta_1\Big\}\ . \end{aligned}$$

This clearly has one solution, $F = 0$, when $M_2^2 = 1$, discussed below. The factor in curly braces is quadratic in M_2^2 and has two real roots provided the discriminant

$$\begin{aligned} D &= \gamma^2(3+\beta_1\sec^2\theta_1)^2-4(\gamma+1)(\gamma-1+\gamma\beta_1)\sec^2\theta_1 \\ &= 5\gamma^2-4+\gamma(2\gamma-4)\beta_1\sec^2\theta_1+\gamma^2\beta_1^2\sec^4\theta_1\ , \qquad (13.44) \end{aligned}$$

is positive. This has a minimum

$$\min(D) = \gamma^2+\frac{\gamma(2-\gamma)}{10\gamma^2-8} > 0 \quad \text{at} \quad \beta_1\sec^2\theta_1 = \frac{\gamma(2-\gamma)}{5\gamma^2-4}\ , \qquad (13.45)$$

so there are always two real roots. The smaller root,

$$M_2^{(so)2} = \frac{\gamma(3+\beta_1\sec^2\theta_1)-\sqrt{D(\beta_1\sec^2\theta_1)}}{2(\gamma+1)}\ , \qquad (13.46)$$

is the switch-off shock. It is also possible to show, with a bit more effort, that $0 < M_2^{(so)2} < 1$, so there is a switch-off shock separating the slow from the intermediate shocks in all possible circumstances.

[2] This is related to the fact that $p_2^\gamma/\rho_2 > p_1^\gamma/\rho_1$ in order that the entropy increase.

Rotational Discontinuity: The solution $M_2 = M_1 = 1$, to the shock equation (13.35), is far richer than it first appears. As noted above, jump condition (13.25) is satisfied automatically, and places no constraint on θ_2. The other jump conditions, eqs. (13.26) and (13.28), become in this limit

$$0 \;=\; \left[\!\!\left[\frac{\beta}{\cos^2\theta} \right]\!\!\right] + \left[\!\!\left[\tan^2\theta \right]\!\!\right] \;=\; \left[\!\!\left[\frac{\beta}{\cos^2\theta} \right]\!\!\right] + \left[\!\!\left[\frac{1}{\cos^2\theta} \right]\!\!\right] \tag{13.47}$$

$$0 \;=\; \left[\!\!\left[\tan^2\theta \right]\!\!\right] + \frac{\gamma}{\gamma-1}\left[\!\!\left[\frac{\beta}{\cos^2\theta} \right]\!\!\right] \;=\; \frac{\gamma}{\gamma-1}\left[\!\!\left[\frac{\beta}{\cos^2\theta} \right]\!\!\right] + \left[\!\!\left[\frac{1}{\cos^2\theta} \right]\!\!\right] \quad . \tag{13.48}$$

Unless $\gamma/(\gamma-1) = 1$ (i.e. $\gamma \to \infty$), these two independent equations can only be satisfied when $\cos^2\theta$ and β are both continuous across the shock. The latter condition, coupled with the fact that

$$\frac{\rho_2}{\rho_1} \;=\; \frac{M_1^2}{M_2^2} \;=\; 1 \quad ,$$

in this case, shows that the plasma's thermodynamic properties, density, pressure, and temperature, remain unchanged across the shock.

It would seem at first that $[\![\cos^2\theta]\!] = 0$ can be satisfied in only two ways:[3] either $\theta_2 = \theta_1$ or $\theta_2 = -\theta_1$. The first is consistent with the other two points where the shock curve intersects the diagonal: the points labeled FMW and SMW in fig. 13.1. In those cases the point of intersection represents a uniform equilibrium and the neighborhood of that intersection consists of small perturbations to that equilibrium: linear MHD waves. Thus we associate the shear Alfvén wave (SAW) with the solution at $M_1 = M_2 = 1$ where $\theta_1 = \theta_2$. In the other case, $\theta_2 = -\theta_1$, the tangential field component reverses direction across the shock. Such a reversal is a general property of intermediate shocks and this state is the limit point of that curve (see fig. 13.2f).

Recall, however, that the case $M_1^2 = M_2^2 = 1$ was the only exception to the earlier derivation of co-planarity using eq. (13.20). This is the only case where $\mathbf{B}_2$ can lie off the plane spanned by $\mathbf{B}_1$ and $\hat{\mathbf{n}}$. The angles between $\hat{\mathbf{n}}$ and the vectors $\mathbf{B}_1$ and $\mathbf{B}_2$ must still be equal, i.e. $\theta_1 = \theta_2$. In this case, however, $\mathbf{B}_1$ and $\mathbf{B}_2$ may be related through a *rotation* by some angle ϕ about $\hat{\mathbf{n}}$. Only in the cases $\phi = 0$ and $\phi = 180°$ will the rotation bring $\mathbf{B}_2$ back to the same plane as $\mathbf{B}_1$ and $\hat{\mathbf{n}}$ — these are the SAW and intermediate shocks discussed above. All other angles, known collectively as *rotational discontinuities*, are assigned to the single point $M_1 = M_2 = 1$ on the shock curve of fig. 13.1.

From the above reasoning (and the hypersonic limit described below) we find that for a given θ_1, taken to be positive for concreteness, the value of $\tan\theta_2$ must fall within the range

$$-\tan\theta_1 \;\le\; \tan\theta_2 \;<\; 4\tan\theta_1 \quad . \tag{13.49}$$

The axis along the bottom of fig. 13.2 shows how the various shocks fall within the range. At the lower end is the rotational discontinuity (13.2f). At the upper end is the hypersonic fast shock.

Variation with parameters: The curve of allowed shocks depends on two parameters: β_1 and the shock angle θ_1. We saw in Chapter 11, that the speeds of the three linear waves

[3] Since B_n must be continuous it is not possible for $\cos\theta_1$ and $\cos\theta_2$ to have different signs.

depend on β and the angle of propagation as well. These three wave speeds determine the intersection of the shock curve with the diagonal, so the two systems, linear and non-linear, will have related dependences on the parameters. In all cases, however, three linear waves exist and have consistently ordered wave speeds. Likewise the three shock branches, with the same basic properties, exist for all parameters. We showed above, in particular, that the switch-off shock exists for all parameters, so there will always be a case separating slow shocks from intermediate shocks.

Figure 13.3 illustrates the nature of this inter-related parameter dependence. A case with high β (specifically $\beta_1 > 1.2$ and therefore $c_s > v_A$) is shown along the top row. The Friedrich's diagram on the left shows the FMW (green curve) consistently faster than (outside) the other two waves. As a result the fast shock is consistently separated from the others when β_1 is large. The SAW and SMW curves (magenta and red) are close together, so the slow and intermediate shock branches are confined to a narrow range around $M_A = 1$.

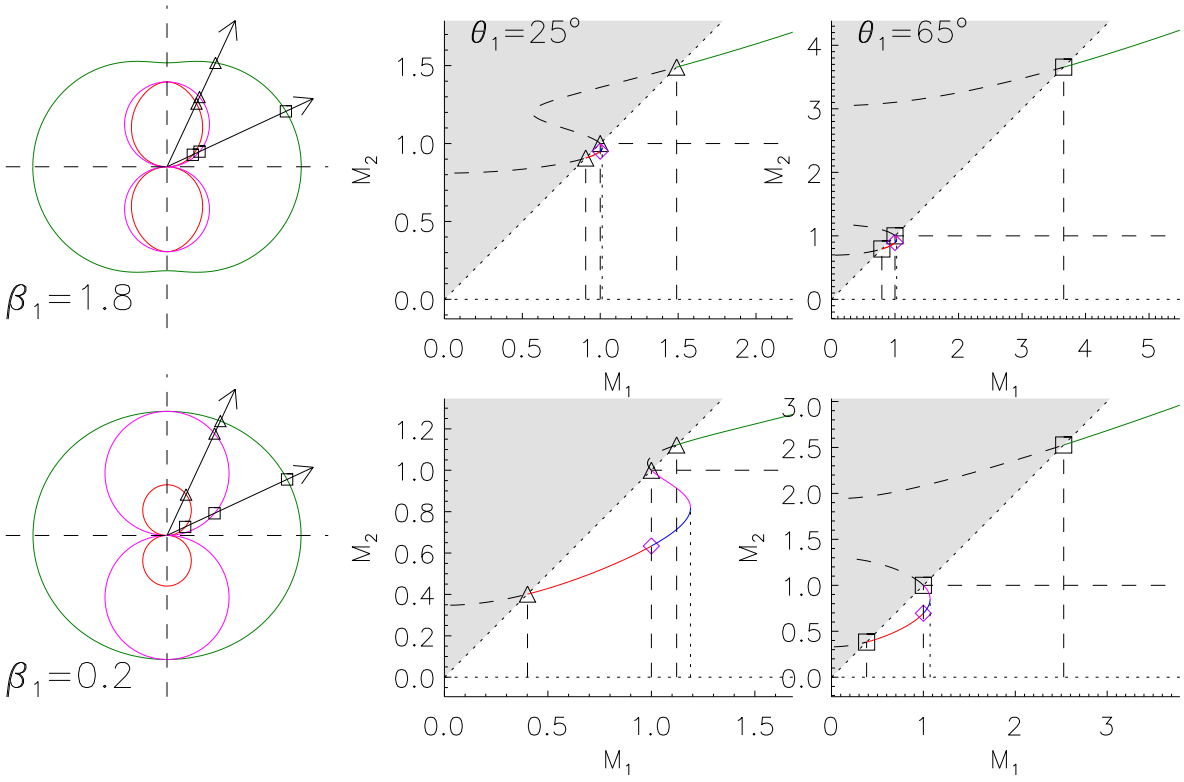

Figure 13.3: The curve of allowed shocks for different parameters plotted with the same format as fig. 13.1. The center and right columns are for shock angles $\theta_1 = 25°$ and $\theta_1 = 65°$ respectively. The top and bottom rows are for $\beta_1 = 1.8$ and $\beta_1 = 0.2$ respectively. A Friedrich's diagram, line those in fig. 11.1, are shown to the left, with an arrow showing the angles $\theta = 25°$ and $\theta = 65°$. The intersections with the three wave curves, giving the linear waves speeds, are marked with the symbol corresponding to the shock diagram to its right.

In a case with low β, such as the bottom row, the FMW and SAW curves become tangent

for parallel propagation: $\theta = 0$. This brings the fast and intermediate shocks close together when θ_1 is small (e.g. the bottom middle panel). The speeds of both the SAW and SMS vanish for perpendicular propagation, $\theta_1 \to 90^\circ$, which brings their curves together. This makes the perpendicular shocks more similar across β, than are the parallel cases (compare the right column to the center column).

13E.1 Limiting cases

Certain aspects of shocks in the limits $\beta \to \infty$ or $\beta \to 0$ bear some resemblance to that of linear waves in those same limits. When $\beta \to \infty$ the magnetic field becomes irrelevant and the MHD waves go over to hydrodynamic waves. It is the fast mode that becomes the sound wave. We show below that the fast shock becomes the simple hydrodynamic shock in the same limit. In the limit $\beta \to 0$ it is the *slow* magnetosonic wave which becomes the sound wave, albeit confined to propagate only along the field lines. We also find below that it is the slow shock that becomes akin to the hydrodynamic shock in this same limit.

Hydrodynamic limit: The MHD case, with 3 distinct kinds of shock, seems far removed from the hydrodynamic case. We can, however, naturally recover hydrodynamics by simply "turning off" the magnetic field. In other words, by taking the lmit $\mathbf{B} \to 0$, which in this case would take $\beta \to \infty$. It would also take $v_{\rm A} \to 0$ so the Alfvén Mach numbers will diverge. We therefore replace these using the relation

$$M_A^2 \;=\; \frac{4\pi\rho u_n^2}{B_n^2} \;=\; \frac{1}{2}\rho u_n^2 \frac{8\pi}{B^2\cos^2\theta} \;=\; \frac{\gamma}{2}\frac{\beta}{\cos^2\theta}\,M_s^2 \quad , \tag{13.50}$$

where $M_s = u_n/c_s$ is the acoustic Mach number. This makes clear that $M_A^2 \sim \beta$, so both will diverge at the same rate in the hydrodynamic limit.

When (13.50) is introduced into (13.35), the leading terms diverge as β^3. Retaining only these terms leaves

$$\begin{aligned} F \;&=\; -(\gamma+1)\left(\frac{\gamma}{2}\frac{\beta_2}{\cos^2\theta_2}\right)^3 M_{s2}^6 \\ &+\; \left[(\gamma-1)\frac{\gamma}{2}\frac{\beta_1}{\cos^2\theta_1}M_{s1}^2 + \frac{\gamma\beta_1}{\cos^2\theta_1}\right]\left(\frac{\gamma}{2}\frac{\beta_2}{\cos^2\theta_2}\right)^2 M_{s2}^4 \\ &+\; \mathcal{O}(\beta^2) \;=\; 0 \quad . \end{aligned} \tag{13.51}$$

Discarding the solutions where $M_{s2} \sim \beta^{-1/2} \to 0$, leaves the relation

$$M_{s2}^2 \;\simeq\; \Big[\,(\gamma-1)M_{s1}^2 + 2\,\Big]\;\frac{\beta_1^2\cos^2\theta_2}{(\gamma+1)\beta_2^2\cos^2\theta_1} \quad . \tag{13.52}$$

Since it contains β_2 and θ_2, this is no longer an explicit relation. Using

$$\frac{p_2}{p_1} \;=\; \frac{\beta_2}{\beta_1}\,\frac{B_2^2}{B_1^2} \;=\; \frac{\beta_2}{\beta_1}\,\frac{B_{n2}^2/\cos^2\theta_2}{B_{n1}^2/\cos^2\theta_1} \;=\; \frac{\beta_2\cos^2\theta_1}{\beta_1\cos^2\theta_2} \quad , \tag{13.53}$$

we may recast the relation as

$$M_{s2}^2 \;=\; \frac{(\gamma-1)M_{s1}^2+2}{(\gamma+1)p_2/p_1} \quad . \tag{13.54}$$

This is easily seen to be equivalent to a combination of eqs. (7.31) and (7.33). This demonstrates that simple hydrodynamic shocks are recovered from *fast shocks* in the hydrodynamic limit, $\beta \to \infty$. The slow and intermediate shocks turn into stationary features with $M_s \sim \beta^{-1/2} \to 0$.

Low-β_1 limit: The limit far more relevant to the solar corona is that of very small plasma β. In this limit the fast magnetosonic wave becomes isotropic and the slow magnetosonic wave goes to a sound wave confined to propagate along the field. MHD shocks show a related behavior.

Taking $\beta_1 \to 0$ in eq. (13.35) does not reduce the complexity of the polynomial. It is still a cubic, so all three shocks still exist. There is, however, a solution in which both M_1^2 and M_2^2 can go to zero together. Retaining from eq. (13.35) only those terms linear in these factors gives

$$0 \;=\; -\,\frac{\gamma+1}{\cos^2\theta_1}\,M_2^2 \;+\; \frac{\gamma-1}{\cos^2\theta_1}\,M_1^2 \quad . \tag{13.55}$$

This is reminiscent of a hypersonic hydrodynamic shock,

$$\frac{M_2^2}{M_1^2} \;=\; \frac{\gamma-1}{\gamma+1} \;=\; \frac{\rho_1}{\rho_2} \quad . \tag{13.56}$$

Since both M_1^2 and M_2^2 are very small, eq. (13.29) demands that $\theta_2 \simeq \theta_1$, so the magnetic field barely bends at the shock. Using this limit in (13.30) yields

$$\beta_2 \;=\; \frac{4}{\gamma+1}M_1^2 \quad . \tag{13.57}$$

While this is still much smaller than unity, owing to the smallness of M_1^2, it is still much greater than β_1, which has been set to zero. The post shock pressure is

$$p_2 \;=\; \frac{2\rho_1 u_{n,1}^2}{\gamma+1} \quad , \tag{13.58}$$

exactly the same as the hypersonic limit for a hydrodynamic shock. Viewed from the de Hoffman-Teller reference frame, this shock occurs as plasma flows at speeds much greater than $c_{s,1}$ along an unbent magnetic field line. The kinetic energy of the cold, inflowing plasma is converted to heat, and the magnetic field plays no role other that directing the flow.

The behavior of the slow shock will deviate from the hydrodynamic case as M_1^2 becomes larger. The extreme point on the branch is the switch-off shock. The $\beta_1 \to 0$ limit of eq. (13.46)

$$M_2^{(so)2} \;=\; \frac{3\gamma-\sqrt{5\gamma^2-4}}{2(\gamma+1)} \;=\; \frac{\gamma-1}{\gamma+1}+\frac{\gamma+2-\sqrt{5\gamma^2-4}}{2(\gamma+1)} \quad , \tag{13.59}$$

is slightly greater than the $M_1^2 = 1$ extrapolation of eq. (13.56). They do, however, agree to a sufficient degree (0.35 vs. 0.25 for $\gamma = 5/3$), that it is reasonable to picture the slow shock as a field-aligned hydrodynamic shock.

Hypersonic limit: For extremely high Mach numbers expression (13.35) becomes

$$F(M_1, M_2) \simeq -(\gamma+1)M_2^6 + (\gamma-1)M_1^2 M_2^4 + \cdots = 0 \ , \tag{13.60}$$

regardless of the value of β_1. The non-trivial solution is a fast shock for which

$$\frac{M_1^2}{M_2^2} \simeq \frac{\gamma+1}{\gamma-1} \simeq \frac{\rho_2}{\rho_1} \simeq \frac{B_{t,2}}{B_{t,1}} \ , \tag{13.61}$$

where the final two expressions are limits of eqs. (13.36) and (13.42). It is notable that the limiting density ratio is identical with that for hydrodynamic shocks: $\rho_2/\rho_1 = 4$ for $\gamma = 5/3$. Since the tangential magnetic field is compressed by the same flow, it is compressed to the same limiting ratio (the induction equation and continuity equation can be combined to show that B_t/ρ is constant for one-dimensional flow). This means the magnetic pressure can be enhanced by no more than the square of this ratio: $\times 16$. The plasma pressure, on the other hand, will, according to eq. (13.30), be enhanced to

$$p_2 \simeq p_1 + \frac{B_n^2}{4\pi}(M_1^2 - M_2^2) \simeq p_1 + \frac{2}{\gamma+1}\rho_1 u_{n,1}^2 \simeq \frac{2\gamma}{\gamma+1}\frac{u_{n,1}^2}{c_{s,1}^2} p_1 \ . \tag{13.62}$$

This is unbounded and matches the hydrodynamic limit.

It is evident from the arguments above, that the bulk kinetic energy of the upstream plasma is converted mainly to plasma pressure in this extreme case. This is done by the Ohmic and viscous dissipation — essential for a self-consistent shocks. This illustrates a notable difference between magnetic pressure and plasma pressure: dissipative processes increase plasma pressure not magnetic pressure. Indeed, if the induction equation is used to derive an equation for magnetic pressure, Ohmic dissipation appears as a *loss*, reducing magnetic pressure. The same term appears as a source of plasma pressure in order to conserve total energy.

Chapter 14

Kinetic Theory

Up to this point we have treated plasmas as continua, like any other fluid. A continuum is characterized by continuous functions of space and time, such as mass density, $\rho(\mathbf{x}, t)$. These quantities make no reference to the the individual particles we know to be composing the plasma: electrons and ions. Here we show how the continuum treatment is derived from the physics of the individual particles composing it. The major step is to derive continuum fluid equations from the dynamics of a large number of identical particles of one species, either electrons or ions. After that one can consider how two different fluids can be described in terms of a single multi-component fluid.

14A The Distribution function

We consider a gas of identical point particles of mass m and charge q. Each has a position $\mathbf{x}(t)$ and velocity $\mathbf{v}(t)$ which evolve according to laws of classical mechanics

$$\frac{d\mathbf{x}}{dt} \;=\; \mathbf{v} \quad , \quad \frac{d\mathbf{v}}{dt} \;=\; \mathbf{a}(\mathbf{x}, \mathbf{v}, t) \quad , \tag{14.1}$$

where $\mathbf{a} = \mathbf{F}/m$ is the acceleration of the particle. In a typical case the particle will be acted upon by electric, magnetic and gravitational forces giving it an acceleration

$$\mathbf{a}(\mathbf{x}, \mathbf{v}, t) \;=\; \frac{q}{m}\mathbf{E}(\mathbf{x}, t) \;+\; \frac{q}{mc}\mathbf{v} \times \mathbf{B}(\mathbf{x}, t) \;+\; \mathbf{g} \quad . \tag{14.2}$$

The fields, $\mathbf{E}$, $\mathbf{B}$ and $\mathbf{g}$ will depend in various ways on the distribution of particles and the velocities of these particles.

Rather than consider each particle separately we use a *distribution function*, $f(\mathbf{x}, \mathbf{v}, t)$ which specifies the distribution of particles in 6-dimensional *phase space*.[1] It is the phase space density of particles, whose use is illustrated by the following example. Within a region of space $\mathcal{V}$ the number of particles with x-velocity in the rage $v_a < v_x < v_b$ is

$$\text{number of particles} \;=\; \int\limits_{\mathcal{V}} d^3\mathbf{x} \int\limits_{v_a}^{v_b} dv_x \int\limits_{-\infty}^{+\infty} dv_y \int\limits_{-\infty}^{+\infty} dv_z \, f(\mathbf{x}, \mathbf{v}) \quad . \tag{14.3}$$

[1] Strictly speaking phase space uses momenta rather than velocities. In non-relativistic work these differ only by a factor m, so the 6-dimensional space of position and velocity has the same structure as canonical phase space, and we use the term *phase space* to describe it.

Note that we integrate over the entire v_y and v_z axes since these do not appear in our condition. Had we taken $v_a = -\infty$ and $v_b = +\infty$ then every particle would have velocity in the specified range.[2] The integral would be over the entire velocity space in order to determine the total number of particles within the region, independent of their velocity. Based on that we see that the number density of particles (number per unit volume) is given by

$$n(\mathbf{x}) \;=\; \int f(\mathbf{x},\mathbf{v})\, d^3\mathbf{v} \quad , \tag{14.4}$$

where the integral is over the entire velocity space. The dimensions of f are evidently $\sec^3 \mathrm{cm}^{-6}$.

In what follows we do not make much use of the specific forces in eq. (14.2). One property which we will use frequently can be verified from that form,

$$\frac{\partial a_i}{\partial v_i} \;=\; 0 \quad . \tag{14.5}$$

In this expression v_i, the i^{th} coordinate of the particle velocity, is a phase space coordinate separate from position. We also use the Einstein index notation where repeated indices (i in this case) are summed over.

The collection of particles has properties we associate with continua. These can be found by integrating the distribution function over its velocity space coordinates after first multiplying by powers of velocity — these are called *moments* of the distribution function. The number density in (14.4) is often called the "zeroth" moment. First and second moments give the mean velocity (center-of-mass velocity) and mean kinetic energy per particle at a given point

$$\mathbf{u}(\mathbf{x},t) \;=\; n^{-1} \int \mathbf{v} f(\mathbf{x},\mathbf{v},t)\, d^3\mathbf{v} \tag{14.6}$$

$$\epsilon(\mathbf{x},t) \;=\; n^{-1} \int \tfrac{1}{2} m |\mathbf{v}|^2 f(\mathbf{x},\mathbf{v},t)\, d^3\mathbf{v} \quad . \tag{14.7}$$

Here we see the fluid velocity $\mathbf{u}(\mathbf{x},t)$ is a moment of the distribution function and must never be confused with $\mathbf{v}$ which is the velocity of a particular particle and therefore a position in phase space. It is to make this distinction patently clear that we have reserved the variable name $\mathbf{v}$ until now, using $\mathbf{u}$ for the fluid velocity.

The distribution function changes over time due to the evolution of the particles which compose it; those evolve according to eq. (14.1). Assuming particles are neither created nor destroyed their phase-space density, f, must "flow" according to a 6-dimensional conservation law

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial x_i}\,(v_i f) + \frac{\partial}{\partial v_i}\,(a_i f) \;=\; 0 \quad . \tag{14.8}$$

We can take advantage of eq. (14.5), and the fact that $\partial v_i / \partial x_i = 0$ since x_i and v_i are independent coordinates of phase space[3] to rewrite this

$$\frac{\partial f}{\partial t} + v_i \frac{\partial f}{\partial x_i} + a_i \frac{\partial f}{\partial v_i} \;=\; \frac{\partial f}{\partial t} + \mathbf{v}\cdot\frac{\partial f}{\partial \mathbf{x}} + \mathbf{a}\cdot\frac{\partial f}{\partial \mathbf{v}} \;=\; 0 \quad . \tag{14.9}$$

[2] In our non-relativistic model we do not need to worry about super-luminal velocity. A relativistic treatment is typically done using 4-momenta, which are also unbounded.

[3] The definition of the partial derivative $\partial/\partial x_1$ is to vary x_1 while holding fixed all other coordinates, x_2, x_3, v_1, v_2, v_3 and time t.

Known as *Vlasov's equation*, this states that the phase space density is unchanging as one follows a particular particle through phase space. It is related to a result of classical mechanics knows as *Liouville's theorem*.

Vlasov's equation forms the basis of all *Plasma Physics*. The acceleration depends on fields such as $\mathbf{E}$. This depends, through Poisson's equation, on the charge density to which the particles under consideration contribute, $qn(\mathbf{x})$, where n is given by a moment of the distribution function according to (14.4). In solving for the unknown function $f(\mathbf{x}, \mathbf{v}, t)$, we see that the Vlasov equation is actually a *non-linear integro-differential equation* — in seven dimensions! Needless to say its full solution is a formidable task. When only electro-static forces are used, the system is called the *Vlasov-Poisson* equations, and when using both electric and magnetic fields it is called the *Vlasov-Maxwell* equations.

We will not pursue here the full solution of either the Vlasov-Poisson of Vlasov-Maxwell equations; that is the purview of Plasma Physics. Instead we will show how the equations governing the time-evolution of the fluid fields, $n(\mathbf{x}, t)$, $\mathbf{u}(\mathbf{x}, t)$ and $\epsilon(\mathbf{x}, t)$, can be derived from Vlasov's eqauation. These fields are moments of the distribution function, according to eqs. (14.4)–(14.7), so their evolution is naturally given by moments of the Vlasov equation.

14B Collisions

In principle Vlasov's equation (14.9) contains all of classical mechanics and is complete as it stands. In order to be completely accurate $f(\mathbf{x}, \mathbf{v}, t)$ would be a terribly singular function, with δ-functions at the phase-space position of each and every particle. Solving it would be no easier than solving Newton's laws separately for each and every particle. So in practice we consider $f(\mathbf{x}, \mathbf{v}, t)$ to represent a smoothed version which represents a *statistically averaged* density of particles, averaged over an an ensemble of identically prepared gasses of particles. Integrals such as eq. (14.3) yield an *expectation* which will probably not be an integer. More importantly the moments in eqs. (14.4)–(14.7) will be smooth functions of space and time reflecting the average particle density at a particular point. Clearly without this averaging n would be zero almost everywhere, except at the exact position of particles where it would be infinite. That is not the kind of field we use in fluid mechanics.

It would be ideal if averaging the Vlasov equation yielded an identical equation where all factors were now the averages. This is in fact the case for terms linear in f since the average of a sum is the sum of the average. As discussed above, however, the final term, $a_i \partial f / \partial v_i$ is actually non-linear in f. The average of a product does not equal the product of the averages.

This discrepancy can be understood by considering two different charged particles both at the same point[4] in space, $\mathbf{x}$, but with different velocities $\mathbf{v}$ and $\mathbf{v}'$. The electric field at that point, $\mathbf{E}(\mathbf{x})$, reflects the charge density of all particles, including the two under consideration. It is, however, a "smoothed" electric field and the two particles experience from it the same acceleration since they occupy the same position. This clearly neglects the local electric field which the particles would exert on each other. That would produce equal and opposite contributions to their respective accelerations — repelling them from one another. This pairwise repulsion is basically what we call a *collision* between the two

[4]In truth they cannot be at the same point. We assume they are separated by a distance much smaller than the scale on which our averaging procedure smooths fields like $n(\mathbf{x})$.

particles. It occurs though small scale fluctuations in the electric field which we have agreed to average away. Since it lacks these kinds of pairwise forces the Vlasov equation, (14.9), is said to govern the *collisionless* evolution of particles.

To account for collisions we add one term to Vlasov's equation, yielding an equation called either Boltzmann's equation or the Fokker-Planck equation,

$$\mathcal{L}f \;=\; \frac{\partial f}{\partial t} + \mathbf{v}\cdot\frac{\partial f}{\partial \mathbf{x}} + \mathbf{a}\cdot\frac{\partial f}{\partial \mathbf{v}} \;=\; \left(\frac{\partial f}{\partial t}\right)_{\text{col}} \;=\; C\{f\} \quad . \tag{14.10}$$

The rhs is a collision term $C\{f\}$ expressing the change in the distribution function due to inter-particle collisions. Since it arises from *local* collisions of the kind described above, its value at one point, $\mathbf{x}$, depends only on the distribution function at that same point. Its value at one velocity, $\mathbf{v}$, represents changes to the distribution function at that velocity, and will in general depend on the function at other velocities. One general form, proposed by Boltzmann, is the nonlinear functional

$$C\{f\} \;=\; \int \mathcal{C}(\mathbf{v},\mathbf{v}',\mathbf{v}'')f(\mathbf{v}')f(\mathbf{v}'')\,d^3v'\,d^3v'' \;-\; f(\mathbf{v})\int \mathcal{D}(\mathbf{v},\mathbf{v}')f(\mathbf{v}')\,d^3v' \quad , \tag{14.11}$$

where the function $\mathcal{C}$ characterizes the rate at which collisions between particles with velocities $\mathbf{v}'$ and $\mathbf{v}''$ produce[5] particles of velocity $\mathbf{v}$ and $\mathcal{D}$ the rate at which collisions with particles of velocity $\mathbf{v}'$ remove particles of velocity $\mathbf{v}$.

The Fokker-Planck and Boltzmann's equations have the same basic form given in eq. (14.10), but with different collision operators, $C\{f\}$. The Fokker-Plank collision operator is a kind of velocity-space diffusion resulting from many small electric-field interactions described above. Boltzmann's equation, using the integral operator in eq. (14.11), is better suited to collisions between neutral particles such as solid spheres. We will leave the collision operator unspecified and refer to the generic form as "Boltzmann's equation".

More important than the detailed form of C are some of its properties. The collisions captured in $C\{f\}$ are assumed to arise from pairwise force interactions, just like *elastic collisions* in classical mechanics. Each collision therefore conserves the number of particles, momentum and kinetic energy. As a result the collision operator must conserve the associated moments of the distribution function given by eqs. (14.4)–(14.7). The changes to each moment due to collisions alone can be expressed as

$$\left(\frac{\partial n}{\partial t}\right)_{\text{col}} \;=\; \int C\{f\}\,d^3\mathbf{v} \;=\; 0 \tag{14.12}$$

$$\left(\frac{\partial \mathbf{u}}{\partial t}\right)_{\text{col}} \;=\; \frac{1}{n}\int \mathbf{v}\,C\{f\}\,d^3\mathbf{v} \;=\; 0 \tag{14.13}$$

$$\left(\frac{\partial \epsilon}{\partial t}\right)_{\text{col}} \;=\; \frac{m}{2n}\int |\mathbf{v}|^2\,C\{f\}\,d^3\mathbf{v} \;=\; 0 \quad . \tag{14.14}$$

Each of these must vanish for every choice of distribution function f.

[5]Particles are not really created out of the vacuum or annihilated. What is meant by these dramatic terms is that the velocity of one of the particles is changed through collision. If its final value is $\mathbf{v}$, we say a particle of this velocity has been "produced", if its initial velocity was $\mathbf{v}$, a particle of that velocity was "destroyed".

14C The fluid equations

Using only the conservation properties of the collision we can derive fluid equations from time derivatives of the moments. Differentiating the first moment, eq. (14.4), and using eq. (14.10)

$$\frac{\partial n}{\partial t} \;=\; \int \frac{\partial f}{\partial t} d^3\mathbf{v} \;=\; -\int v_i \frac{\partial f}{\partial x_i} d^3\mathbf{v} \;-\; \int a_i \frac{\partial f}{\partial v_i}\, d^3\mathbf{v} \quad . \tag{14.15}$$

The collision term is absent from the rhs because of eq. (14.12). We now make use of eq. (14.5) to recast the final term as the integral of divergence

$$\frac{\partial n}{\partial t} \;=\; -\frac{\partial}{\partial x_i}\int v_i f\, d^3\mathbf{v} \;-\; \underbrace{\int \frac{\partial}{\partial v_i}(a_i f) d^3\mathbf{v}}_{=0} \;=\; -\nabla\cdot(n\mathbf{u}) \tag{14.16}$$

where we have assumed that $f \to 0$ as $|\mathbf{v}| \to \infty$ sufficiently rapidly for the surface term to vanish. We hereafter assume this is always the case for integrals of total divergences in velocity space.

14C.1 The momentum equation

Taking the derivative of the next moment

$$\frac{\partial(nu_i)}{\partial t} \;=\; \int v_i \frac{\partial f}{\partial t} d^3\mathbf{v} \;=\; -\int v_i v_j \frac{\partial f}{\partial x_j} d^3\mathbf{v} \;-\; \int v_i \frac{\partial(a_j f)}{\partial v_j}\, d^3\mathbf{v} \tag{14.17}$$

where the collision term is absent because of eq. (14.13). Dropping the surface term for the integration by parts we obtain

$$\frac{\partial(nu_i)}{\partial t} \;=\; -\frac{\partial}{\partial x_j}\int v_i v_j f\, d^3\mathbf{v} \;+\; \int \underbrace{\frac{\partial v_i}{\partial v_j}}_{\delta_{ij}} a_j f\, d^3\mathbf{v} \tag{14.18}$$

We now introduce the relative velocity, $\delta\mathbf{v} = \mathbf{v} - \mathbf{u}$. The first moment of f w.r.t. this will vanish

$$\int (\mathbf{v}-\mathbf{u}) f\, d^3\mathbf{v} \;=\; \int \mathbf{v} f\, d^3\mathbf{v} - \mathbf{u}\int f\, d^3\mathbf{v} \;=\; n\mathbf{u} - \mathbf{u}n \;=\; 0 \quad .$$

In terms of this we can express the second moment

$$\int v_i v_j f\, d^3\mathbf{v} \;=\; \int (u_i+\delta v_i)(u_j+\delta v_j) f\, d^3\mathbf{v} \;=\; nu_i u_j + \int \delta v_i\, \delta v_j\, f\, d^3 v \quad . \tag{14.19}$$

The last term is a second moment of the distribution relative to the center of mass; we define this to be the *pressure tensor*

$$P_{ij} \;=\; m\int \delta v_i\, \delta v_j\, f(\mathbf{x},\mathbf{v})\, d^3 v \quad , \tag{14.20}$$

and then write the conservative form of the momentum equation

$$\frac{\partial(nu_i)}{\partial t} \;+\; \frac{\partial(nu_i u_j)}{\partial x_j} \;=\; -\frac{1}{m}\frac{\partial P_{ij}}{\partial x_j} \;+\; n\bar{a}_i \quad . \tag{14.21}$$

Here we have introduced the acceleration vector $\bar{\mathbf{a}}(\mathbf{x},t)$, to denote the average acceleration of a particle at position $\mathbf{x}$. If $\mathbf{a}(\mathbf{x},\mathbf{v})$ is given by eq. (14.2) then

$$\bar{\mathbf{a}}(\mathbf{x},t) \;=\; \frac{q}{m}\mathbf{E}(\mathbf{x},t) \;+\; \frac{q}{mc}\mathbf{u}(\mathbf{x},t)\times\mathbf{B}(\mathbf{x},t) \;+\; \mathbf{g} \quad . \tag{14.22}$$

The only significant difference is that $\mathbf{v}$ has been replaced by the mean velocity $\mathbf{u}$, which is itself a function of $\mathbf{x}$.

14C.2 The pressure tensor

The pressure tensor P_{ij} is central to fluid mechanics. It is evident from its definition in eq. (14.20) that it is symmetric, $P_{ji} = P_{ij}$, and therefore has purely real eigenvalues. Its eigenvalues are also all non-negative. To see this we contract the tensor with an arbitrary unit vector $\hat{\mathbf{e}}$ to get

$$P_{ij}\,\hat{e}_i\hat{e}_j \;=\; m\int |\hat{\mathbf{e}}\cdot\delta\mathbf{v}|^2\, f(\mathbf{x},\mathbf{v})\, d^3v \;\geq\; 0 \quad . \tag{14.23}$$

In the case that all its eigenvalues are equal, call that value p, the tensor is isotropic

$$P_{ij} \;=\; p\,\delta_{ij} \quad , \tag{14.24}$$

where δ_{ij} is the Kronecker delta. In this case the divergence of the pressure tensor is a gradient,

$$-\frac{\partial P_{ij}}{\partial x_j} \;=\; -\,\frac{\partial p}{\partial x_i} \quad , \tag{14.25}$$

which is the force we usually associate with pressure. The isotropic form will occur when the distribution function $f(\mathbf{v})$ is unchanged under rotation of the velocity vector $\mathbf{v}$ about the mean velocity $\mathbf{u}$ — that is when there is no preferred direction in the velocity distribution.

Even when the tensor is not isotropic we can define a scalar pressure in terms of its trace

$$p \;\equiv\; \tfrac{1}{3}P_{ii} \;=\; \tfrac{1}{3}\mathrm{Tr(P)} \quad . \tag{14.26}$$

The full pressure tensor can then be written

$$P_{ij} \;=\; p\,\delta_{ij} \;+\; \pi_{ij} \quad , \tag{14.27}$$

where π_{ij} is a traceless matrix ($\pi_{ii} = 0$) containing all anisotropic aspects of the pressure tensor. The so-called viscous stress tensor is $\sigma_{ij} = -\pi_{ij}$.

The volumetric energy density of the particles is

$$\begin{aligned} n\epsilon \;&=\; \int \tfrac{1}{2}m|\mathbf{v}|^2 f(\mathbf{x},\mathbf{v})\, d^3v \;=\; \tfrac{1}{2}m\int |\mathbf{u}+\delta\mathbf{v}|^2 f(\mathbf{x},\mathbf{v}) d^3v \\ &=\; \tfrac{1}{2}mn|\mathbf{u}|^2 + \tfrac{1}{2}m\int |\delta\mathbf{v}|^2 f(\mathbf{x},\mathbf{v}) d^3v \;=\; \tfrac{1}{2}mn|\mathbf{u}|^2 + \tfrac{1}{2}P_{ii} \\ &=\; \tfrac{1}{2}mn|\mathbf{u}|^2 + \tfrac{3}{2}p \quad . \end{aligned} \tag{14.28}$$

Thus the volumetric internal energy of the gas is $(3/2)p$. We have not yet invoked the idea of temperature, nor assumed anything about the form of the distribution function. Pressure

is simply a measure of the kinetic energy of a collection of particles — in particular the energy in the center-of-mass frame. It is also noteworthy that pressure has nothing to do with collisions.

Using this notation the momentum equation becomes

$$\frac{\partial(mnu_i)}{\partial t} \; + \; \frac{\partial(mnu_iu_j)}{\partial x_j} \;=\; -\frac{\partial p}{\partial x_i} - \frac{\partial \pi_{ij}}{\partial x_j} \; + \; n\bar{F}_i \quad , \tag{14.29}$$

where $\bar{\mathbf{F}} = m\bar{\mathbf{a}}$ is the average force acting on a particle – called the body force.

14C.3 The energy equation

The final fluid equation governs the evolution of the scalar pressure p. We find this from the derivative of the second moment

$$\begin{aligned} \frac{\partial(n\epsilon)}{\partial t} &= \tfrac{1}{2}m\int |\mathbf{v}|^2 \frac{\partial f}{\partial t} d^3\mathbf{v} \;=\; -\tfrac{1}{2}m\int |\mathbf{v}|^2 v_i \frac{\partial f}{\partial x_i} d^3\mathbf{v} \;-\; \tfrac{1}{2}m\int |\mathbf{v}|^2 \frac{\partial (a_i f)}{\partial v_i}\, d^3\mathbf{v} \\ &= -\tfrac{1}{2}m\frac{\partial}{\partial x_i}\int |\mathbf{v}|^2 v_i f\, d^3\mathbf{v} \;+\; m\int v_i m_i\, f\; d^3\mathbf{v} \end{aligned} \tag{14.30}$$

after using the elasticity of the collisions through eq. (14.14). Using expression (14.2) for a_i we see that v_ia_i is linear in velocity — the quadratic term vanishes in the dot product. This means that

$$\int v_i a_i\, f\; d^3\mathbf{v} \;=\; n\,\mathbf{u}\cdot\bar{\mathbf{a}}(\mathbf{x},t) \quad ,$$

since the Lorentz force in $\bar{\mathbf{a}}$ vanishes upon contraction with $\mathbf{u}$.

The other term in the expression is re-written using $\mathbf{v} = \mathbf{u} + \delta\mathbf{v}$. The fact that any moment with only a single $\delta\mathbf{v}$ will vanish means

$$\begin{aligned} \int |\mathbf{v}|^2 v_i f\, d^3\mathbf{v} &= n|\mathbf{u}|^2 u_i + \frac{1}{m}\left(P_{jj}u_i + 2P_{ij}u_j\right) + \int \delta v_i\, |\delta\mathbf{v}|^2 f\, d^3\mathbf{v} \\ &= n|\mathbf{u}|^2 u_i + 5\frac{p}{m}u_i + 2\frac{\pi_{ij}u_j}{m} + \frac{2q_i}{m} \end{aligned} \tag{14.31}$$

where we have replaced P_{ij} using eq. (14.27) and introduced the third moment

$$\mathbf{q} \;=\; \tfrac{1}{2}m\int \delta\mathbf{v}|\delta\mathbf{v}|^2 f(\mathbf{x},\mathbf{v})\, d^3v \tag{14.32}$$

known as the *heat flux*. Using this in the energy equation gives

$$\frac{\partial(n\epsilon)}{\partial t} + \frac{\partial(\frac{1}{2}mn|\mathbf{u}|^2u_i)}{\partial x_i} \;=\; -\tfrac{5}{2}\frac{\partial(pu_i)}{\partial x_i} - \frac{\partial(\pi_{ij}u_j)}{\partial x_i} - \frac{\partial q_i}{\partial x_i} + nu_i\bar{F}_i \quad . \tag{14.33}$$

Taking the dot product of eq. (14.29) with u_i, and making use of the continuity equation gives a conservation law for kinetic energy density alone

$$\frac{\partial(\frac{1}{2}mn|\mathbf{u}|^2)}{\partial t} \; + \; \frac{\partial(\frac{1}{2}mnu_j|\mathbf{u}|^2)}{\partial x_j} \;=\; -u_i\frac{\partial p}{\partial x_i} - u_i\frac{\partial \pi_{ij}}{\partial x_j} \; + \; nu_i\bar{F}_i \tag{14.34}$$

This can be subtracted from eq. (14.33), the conservation of total energy, to yield an equation for the evolution of the *internal* energy density, $(3/2)p$, after using eq. (14.28). The internal energy evolves according to

$$\frac{\partial(\frac{3}{2}p)}{\partial t} + \mathbf{u}\cdot\nabla(\tfrac{3}{2}p) \;=\; -\tfrac{5}{2}p\nabla\cdot\mathbf{v} - \pi_{ij}\frac{\partial u_i}{x_j} - \nabla\cdot\mathbf{q} \quad . \tag{14.35}$$

This is the usual energy equation with $\gamma = 5/3$ — the value appropriate for point particles.

14C.4 The unclosed system of fluid equations

Equations (14.16), (14.29) and (14.35) are the fluid equations we are familiar with. Collected together, and in conventional notation, they are

$$\frac{\partial\rho}{\partial t} + \nabla\cdot(\rho\mathbf{u}) = 0 \tag{14.36}$$

$$\rho\frac{D\mathbf{u}}{Dt} = -\nabla p - \nabla\cdot\underline{\underline{\pi}} + \rho\bar{\mathbf{F}} \tag{14.37}$$

$$\frac{Dp}{Dt} = -\tfrac{5}{3}p\nabla\cdot\mathbf{u} - \tfrac{2}{3}\pi_{ij}\frac{\partial u_i}{\partial x_j} - \tfrac{2}{3}\nabla\cdot\mathbf{q} \tag{14.38}$$

where $D/Dt = \partial_t + \mathbf{u}\cdot\nabla$ is the advective deriviative and $\rho = mn$ is the mass density. These follow directly from moments of Boltzmann's equation, eq. (14.10). We have made no assumptions about the distribution function, $f(\mathbf{x},\mathbf{v})$, other than that its evolution is governed by eq. (14.10). Nor have we assumed anything about collisions other than that they conserve momentum and energy according to (14.12)–(14.14). Thus the fluid equations are extremely general.

They are not, however, a closed system since they involve fields $\pi_{ij}(\mathbf{x},t)$ and $\mathbf{q}(\mathbf{x},t)$, whose evolution we have not determined. These are moments of the distribution function different from the ones defining n, $\mathbf{u}$ and p. We could, in principle, take further moments Boltzmann's equation to derive equations for the evolution of π_{ij} and $\mathbf{q}$. It is evident from the procedure so far that this will not be the final step in closing the system. For example, the heat flux, $\mathbf{q}$ is a third moment of the distribution function and its time-evolution will depend on the fourth moment. This process will regress to infinity, never closing. Rather than solve for an infinite number of moments we may as well solve eq. (14.10) for the full distribution function $f(\mathbf{x},\mathbf{v})$.

The alternative used in all fluid equations is to find expressions for the viscous stress tensor, $-\pi_{ij}$ and the heat flux $\mathbf{q}$, in terms of the fields we already know: n, $\mathbf{u}$ and p. This prescription requires use of the collision operaor $C\{f\}$ in eq. (14.10).

14D The attraction of the Maxwellian

A gas of particles has an entropy whose value at any point is determined by the distribution function f at that point. Entropy is a kind of mean information content[6] and is therefore

[6] In Shannon's information theory the "information" assigned to an event which can occur with probability p is defined to be $-\log_2 p$. Since the base-2 logarithm is used information is measured in "bits". An event which is certain to occur contains no information, one that occurs half the time ($p = 1/2$), like a 1 in a binary file, contains one bit of information.

proportional to the integral

$$s(\mathbf{x}) \;=\; -\int f\ln(f)\,d^3\mathbf{v} \quad . \tag{14.39}$$

This is not, strictly speaking, a moment since it is non-linear in f. To conform to the second law of thermodynamics we expect collisions to induce only positive changes to entropy

$$\left(\frac{\partial s}{\partial t}\right)_{\rm col} \;=\; -\int[1+\ln(f)]\,C\{f\}\,d^3\mathbf{v} \;\geq 0 \quad . \tag{14.40}$$

Once again this inequality should hold for all possible functions $f(\mathbf{x},\mathbf{v})$.

Left to itself the collision operator should drive the distribution function to ever larger entropy until it reaches a maximum. This maximum must be achieved while preserving moments (14.4)–(14.7). The function maximizing eq. (14.39) subject to these constraints is

$$f(\mathbf{x},\mathbf{v}) \;=\; n\left(\frac{\beta}{\pi}\right)^{3/2}\exp\left[-\beta|\mathbf{v}-\mathbf{u}|^2\right] \quad , \tag{14.41}$$

where β is a Lagrange undetermined multiplier from constraining the energy. All other Lagrange multipliers have been eliminated by re-applying the constraints on moments n and $\mathbf{u}$. This can be seen to be traditional Maxwellian after making the association $\beta = m/2k_{\rm B}T$

$$f_0(\mathbf{x},\mathbf{v}) \;=\; \frac{n}{(2\pi k_{\rm B}T/m)^{3/2}}\exp\left[-\frac{|\mathbf{v}-\mathbf{u}|^2}{2k_{\rm B}T/m}\right] \;=\; \frac{n}{(2\pi)^{3/2}v_{\rm th}^3}\exp\left[-\frac{|\delta\mathbf{v}|^2}{2v_{\rm th}^2}\right] \quad , \tag{14.42}$$

where the thermal velocity is defined by $v_{\rm th}^2 = k_{\rm B}T/m$. The temperature is essentially defined by its appearance in a Maxwellian distribution; particles with any other form of distribution cannot, strictly speaking, be assigned a temperature. We have seen above, however, such cases *can* be characterized by density and pressure.

If the distribution function has achieved its entropy maximum the collisions should change it no further. This means the operator must also satisfy the condition

$$C\{f_0\} \;=\; 0 \quad . \tag{14.43}$$

for every choice of moments, n, $\mathbf{u}$ and T.

Using the Maxwellian distribution in the definition of the pressure tensor, eq. (14.20), and heat flux, (14.32), gives

$$P_{ij} \;=\; nk_{\rm B}T\,\delta_{ij} \quad , \quad \mathbf{q} \;=\; 0 \quad . \tag{14.44}$$

The first equation gives us the so-called ideal gas law, $p = nk_{\rm B}T = nmv_{\rm th}^2$, relating moments n and p of the distribution function, f, with the temperature in the corresponding Maxwellian distribution f_0. The procedure also shows there is no viscous stress $\pi_{ij} = 0$. The absence of viscous stress is a natural consequence of the isotropy of a Maxwellian — it has no preferred direction.

Assuming the distribution function to be a Maxwellian distribution leads to $\pi_{ij} = 0$ and $\mathbf{q} = 0$ which effectively closes the fluid system — there is nothing left unknown. Using these values in eqs. (14.36)—(14.38) results in fluid equations known as *the ideal fluid equations*

$$\frac{Dn}{Dt} \;=\; -n(\nabla\cdot\mathbf{u}) \quad , \tag{14.45}$$

$$\frac{D\mathbf{u}}{Dt} = -\frac{\nabla p}{mn} + \bar{\mathbf{a}} \ , \tag{14.46}$$

$$\frac{Dp}{Dt} = -\tfrac{5}{3}p(\nabla \cdot \mathbf{u}) \ , \tag{14.47}$$

These equations describe the evolution of fluid moments when collisions are so effective they can maintain the distribution function in the form of a perfect Maxwellian. This limit may appear perverse since viscosity and thermal conductivity are typically associated with collisions (we will confirm this fact below), but their absence occurs not from an absence of collisions, but rather from an overwhelming dominance of collisions.

Non-ideal fluid equations contain viscosity and thermal conduction which must arise from a distribution function different from a Maxwellian. The departure from Maxwellian occurs because the fluid properties themselves break isotropy. For example the temperature gradient ∇T defines a direction along which heat flux can occur. We seek a relation between temperature gradient ∇T and heat flux $\mathbf{q}$. Such a relation, involving a thermal conductivity κ, is what we seek using the collision operator. In a similar vein we seek a relation between velocity gradients and the stress tensor $-\pi_{ij}$ involving a viscosity μ. These relationships, along with the transport coefficients they involve, depend on the collision operator $C\{f\}$.

Chapter 15

Collisions and transport

15A The Chapman-Enskog assumption

The collision operator vanishes for a Maxwellian, $C\{f_0\} = 0$. If $\mathcal{L}f_0 = 0$ then the Maxwellian distribution f_0 would completely satisfy Boltzmann's equation, eq. (14.10), and would be its solution. As we will see below, however, $\mathcal{L}f_0 \neq 0$. The Boltzmann equation is therefore a competition between $\mathcal{L}$ driving the distribution away from a Maxwellian and collisions driving it toward Maxwellian. Collisions drive the distribution function toward a Maxwellian on some time-scale $\tau_{\rm c}$. They will have the upper hand in the competition provided $\mathcal{L}$ works on a time-scale longer than τ_c; this is what we will hereafter assume.

The Chapman-Enskog approach is to assume collision time-scale $\tau_{\rm c}$ is sufficiently small that the balance between these two tendencies occurs at a distribution function *close* to a Maxwellian. We write the total distribution as $f = f_0 + f_1$, and assume f_1 to be, in some sense, small compared to f_0. The method is therefore a perturbations about a Maxwellian distribution.

Due to eq. (14.43) we see that $C\{f_0 + f_1\}$ will scale, to leading order, with f_1. The ratio of this small quantity to the small collision time will be comparable to f_0 and will balances $\mathcal{L}f_0$. This kind of balance is described by the equation

$$\mathcal{L}f_0 \;=\; C\{f_0 + f_1\} \quad , \tag{15.1}$$

which is to be solved for the unknown perturbation $f_1(\mathbf{x}, \mathbf{v})$. Equation (15.1) neglects $\mathcal{L}f_1$ since it is assumed to be smaller than $\mathcal{L}f_0$ on the basis that $f_0 \gg f_1$. To be more accurate we should demand $\mathcal{L}f_1 \ll \mathcal{L}f_0$ which is somewhat different. We are safe making the former assumption when the frequencies in $\mathcal{L} \ll 1/\tau_{\rm c}$, since then $\mathcal{L}f_1 \ll C\{f_0 + f_1\}$. This will not be the case for strongly magnetized particles, $\Omega\tau_{\rm c} \gg 1$; we return to that case below, and use the simpler case, eq. (15.1) for the time being.

Using $\mathbf{v} = \mathbf{u} + \delta\mathbf{v}$ and the advective derivative $D/Dt = \partial_t + \mathbf{u}\cdot\nabla$ we can re-write the Vlasov operator

$$\mathcal{L} \;=\; \frac{D}{Dt} + \delta\mathbf{v}\cdot\frac{\partial}{\partial\mathbf{x}} + \mathbf{a}\cdot\frac{\partial}{\partial\mathbf{v}} \tag{15.2}$$

It turns out to be convenient to divide eq. (15.1) by f_0 to get an equation

$$\mathcal{L}\ln f_0 \;=\; \frac{C\{f_0 + f_1\}}{f_0} \quad . \tag{15.3}$$

Up to constant terms, whose derivatives vanish, the logarithm of f_0 is

$$\ln f_0 \;=\; \ln(p/T^{5/2}) - \frac{|\delta\mathbf{v}|^2}{2v_{\rm th}^2} \;=\; \ln(n/T^{3/2}) - \frac{|\delta\mathbf{v}|^2}{2v_{\rm th}^2} \tag{15.4}$$

This shows the function in terms of its fluid moments. These moments will satisfy fluid equations which include two other moments: the pressure tensor, given by eq. (14.20), and the heat flux given by eq. (14.32). Due to our scaling, $f_1 \ll f_0$ we assume that the moments satisfy *ideal* fluid equations, at least to leading order. For convenience we use the energy equation and the ideal fluid equation to form a equation for temperature

$$\frac{DT}{Dt} \;=\; -\tfrac{2}{3}T(\nabla\cdot\mathbf{u}) \quad , \tag{15.5}$$

The advective time derivatives of $\ln f_0$ is easily found

$$\begin{aligned}\frac{D\ln f_0}{Dt} &= \overbrace{\frac{1}{n}\frac{Dn}{Dt} - \tfrac{3}{2}\frac{1}{T}\frac{DT}{Dt}}^{=0} - \frac{\delta\mathbf{v}}{v_{\rm th}^2}\cdot\frac{D\delta\mathbf{v}}{Dt} + \frac{|\delta\mathbf{v}|^2}{2v_{\rm th}^2}\frac{1}{T}\frac{DT}{Dt} \\ &= \frac{\delta\mathbf{v}}{v_{\rm th}^2}\cdot\frac{D\mathbf{u}}{Dt} \;-\; \frac{|\delta\mathbf{v}|^2}{3v_{\rm th}^2}\nabla\cdot\mathbf{u}\end{aligned}$$

after using the continuity equation and the ideal energy equation, eq. (15.5). Its other derivatives are

$$\begin{aligned}\delta\mathbf{v}\cdot\frac{\partial\ln f_0}{\partial\mathbf{x}} &= \frac{\delta\mathbf{v}\cdot\nabla p}{p} - \tfrac{5}{2}\frac{\delta\mathbf{v}\cdot\nabla T}{T} + \frac{|\delta\mathbf{v}|^2\delta\mathbf{v}\cdot\nabla T}{2v_{\rm th}^2 T} - \frac{\delta\mathbf{v}\,\delta\mathbf{v}:\nabla\delta\mathbf{v}}{v_{\rm th}^2} \\ &= \frac{\delta\mathbf{v}\cdot\nabla p}{v_{\rm th}^2 mn} \;+\; \left[\frac{|\delta\mathbf{v}|^2}{v_{\rm th}^2} - 5\right]\frac{\delta\mathbf{v}\cdot\nabla T}{2T} \;+\; \frac{\delta\mathbf{v}\,\delta\mathbf{v}:\nabla\mathbf{u}}{v_{\rm th}^2}\end{aligned}$$

and

$$\mathbf{a}\cdot\frac{\partial\ln f_0}{\partial\mathbf{v}} \;=\; -\frac{\mathbf{a}\cdot\delta\mathbf{v}}{v_{\rm th}^2} \;=\; -\frac{\bar{\mathbf{a}}\cdot\delta\mathbf{v}}{v_{\rm th}^2} \quad .$$

The final expression follows from the fact that $\mathbf{a}-\bar{\mathbf{a}} = (q/mc)\delta\mathbf{v}\times\mathbf{B}$.

Assembling these derivatives in Boltzmann's equation we find

$$\begin{aligned}\mathcal{L}\ln f_0 &= \frac{\delta\mathbf{v}}{v_{\rm th}^2}\cdot\overbrace{\left[\frac{D\mathbf{u}}{Dt} + \frac{\nabla p}{mn} - \bar{\mathbf{a}}\right]}^{=0} + \left[\frac{|\delta\mathbf{v}|^2}{v_{\rm th}^2} - 5\right]\frac{\delta\mathbf{v}\cdot\nabla T}{2T} \\ &+ \frac{1}{v_{\rm th}^2}\left[\delta\mathbf{v}\,\delta\mathbf{v}:\nabla\mathbf{u} - \tfrac{1}{3}|\delta\mathbf{v}|^2\nabla\cdot\mathbf{u}\right] \\ &= \left[\frac{|\delta\mathbf{v}|^2}{v_{\rm th}^2} - 5\right]\frac{\delta\mathbf{v}\cdot\nabla T}{2T} \;+\; \frac{\delta\mathbf{v}\,\delta\mathbf{v}}{v_{\rm th}^2} : \left[\nabla\mathbf{u} - \tfrac{1}{3}(\nabla\cdot\mathbf{u})\,I\right]\end{aligned}$$

after using the ideal momentum equation, (14.46).

Introducing the normalized velocity difference, $\boldsymbol{\xi} = \delta\mathbf{v}/v_{\rm th}$ we can write the collision term

$$\mathcal{L}f_0 \;=\; f_0(\xi)\Big[\xi^2-5\Big]\boldsymbol{\xi}\cdot\frac{v_{\rm th}\nabla T}{2T} \;+\; \tfrac{1}{2}f_0(\xi)\boldsymbol{\xi}\boldsymbol{\xi}:\underbrace{\Big[\nabla\mathbf{u}+(\nabla\mathbf{u})^T-\tfrac{2}{3}(\nabla\cdot\mathbf{u})\,I\Big]}_{\underline{\underline{\rm V}}} \tag{15.6}$$

introducing the symmetric form of the rate of strain tensor, $\underline{\underline{\rm V}}$. Only the symmetric part contributes to any contraction with a symmetric tensor such as $\boldsymbol{\xi}\boldsymbol{\xi}$. Evidently a Maxwellian distribution is inconsistent with the Vlasov equation (collisionless Boltzmann) in the presence of either temperature gradients or velocity shear. Those properties drive $f(\mathbf{x},\mathbf{v})$ away from Maxwellian and it is the collisions which must restore it.

A notable feature of eq. (15.6) is that $\mathbf{u}$ appears only in the traceless shear tensor $\underline{\underline{\rm V}}$. This vanishes for homologous expansion, $\mathbf{u}(\mathbf{x}) = \alpha(\mathbf{x}-\mathbf{x}_0)$, although $\nabla\cdot\mathbf{u} = 3\alpha$ does not. The viscous stress tensor, arising from the non-Maxwellian contribution, f_1, will therefore vanish under homologous expansion. This means that a fluid satisfying Boltzmann's equation will have zero *bulk viscosity* — only shear viscosity. This property is independent of the the form of the collision term; it follows from $\mathcal{L}f_0$ alone.

The moments of the Maxwellian are worked out in an appendix. With these in hand it can be verified that the lhs of (15.1) satisfies all the same conservation laws as the collision operator on the rhs

$$\begin{aligned}
\int(\mathcal{L}f_0)d^3v \;&=\; \tfrac{1}{2}\left[\int\boldsymbol{\xi}\boldsymbol{\xi}f_0\,d^3v\right]:\underline{\underline{\rm V}} \;=\; \tfrac{1}{2}n{\rm Tr}(\underline{\underline{\rm V}}) \;=\; 0\\
\int(\mathcal{L}f_0)\boldsymbol{\xi}\,d^3v \;&=\; \left[\int(\xi^2-5)\boldsymbol{\xi}\boldsymbol{\xi}f_0\,d^3v\right]\cdot\frac{v_{\rm th}\nabla T}{2T} \;=\; 0\\
\int(\mathcal{L}f_0)\xi^2\,d^3v \;&=\; \tfrac{1}{2}\left[\int\xi^2\boldsymbol{\xi}\boldsymbol{\xi}f_0\,d^3v\right]:\underline{\underline{\rm V}} \;=\; \tfrac{5}{2}n{\rm Tr}(\underline{\underline{\rm V}}) \;=\; 0
\end{aligned}$$

after dropping all odd moments.

15B The BGK collision operator

A collision operator far simpler than eq. (14.11), but with all the necessary properties enumerated above, is the the so-called BGK operator

$$C_{\rm BGK}\{f\} \;=\; -\tau_{\rm c}^{-1}(f-f_0)\quad, \tag{15.7}$$

where $f_0(\mathbf{x},\mathbf{v})$ is the Maxwellian with the same moments as $f(\mathbf{x},\mathbf{v})$. This dependence on moments means that (15.7) is actually a functional of f, and not a simple function as it appears. The relaxation times $\tau_{\rm c}$ may itself depend on moments of f.

The operator (15.7) does exactly what we want the collision operator to do: it drives the distribution function toward a Maxwellian. The effects of collisions alone are described by the equation

$$\left(\frac{\partial f}{\partial t}\right)_{\rm col} \;\simeq\; \left(\frac{\partial}{\partial t}\right)_{\rm col}(f-f_0) \;=\; -\tau_{\rm c}^{-1}(f-f_0)\quad. \tag{15.8}$$

The distribution function relaxes, under the effects of collisions, exponentially to a Maxwellian on time scale τ_c. What's more, it is evident from its form that $C_{\rm BGK}\{f\}$ satisfies (14.12)–(14.14) as well as (14.43). The final criterion involves the entropy change

$$\left(\frac{\partial s}{\partial t}\right)_{\rm col} \;=\; \tau_{\rm c}\int [1+\ln(f)]\,(f-f_0)\,d^3\mathbf{v} \;=\; \tau_{\rm c}\int \ln(1+f_1/f_0)\,f_1\,d^3\mathbf{v} \tag{15.9}$$

after using the fact that the lowest moments of $f_1 = f - f_0$ all vanish and using (15.4) for $\ln f_0$. Expanding the logarithm in f_1/f_0 gives

$$\left(\frac{\partial s}{\partial t}\right)_{\rm col} \;\simeq\; \tau_{\rm c}\int \frac{f_1^2(\mathbf{v})}{f_0(\mathbf{v})}\,d^3\mathbf{v} \;\geq\; 0, \tag{15.10}$$

where equality clearly occurs when $f_1(\mathbf{v}) \equiv 0$. Thus the BGK operator satisfies the second law of thermodynamics, at least when the distribution function is nearly Maxwellian as we have assumed.

Using (15.7) in eq. (15.1) we can quickly find an explicit solution for f_1:

$$f_1(\mathbf{x},\mathbf{v}) = -\tau_{\rm c}\mathcal{L}f_0 \;=\; -\tfrac{1}{2}\tau_{\rm c}v_{\rm th}f_0(\xi)\Big(\xi^2-5\Big)\boldsymbol{\xi}\cdot\nabla\ln T \;-\; \tfrac{1}{2}\tau_{\rm c}f_0(\xi)\boldsymbol{\xi}\boldsymbol{\xi}:\underline{\underline{\rm V}}\quad. \tag{15.11}$$

This shows how a temperature gradient or velocity shear create a departure from Maxwellian. We have assumed, at the outset, that this perturbation would be small: $f_1 \ll f_0$. Denoting the mean-free-path, $\ell_{\rm mfp} = \tau_{\rm c}v_{\rm th}$, the thermal gradient scale $\lambda_T = |\nabla\ln T|^{-1}$ and the hydrodynamic time as the inverse matrix norm of the shear tensor, $\tau_h = ||\underline{\underline{\rm V}}||^{-1}$, our assumption appears to require

$$\lambda_T \gg \ell_{\rm mfp} \quad\text{ and }\quad \tau_h \gg \tau_{\rm c}\quad. \tag{15.12}$$

According to eq. (15.11), however, the ratio f_1/f_0 increases without bound as $\xi\to\infty$ so the requirement $f_1 \ll f_0$ can never be truly satisfied. Under a temperature gradient alone (i.e. $\underline{\underline{\rm V}} = 0$) the two are equal when ξ satisfies the equation

$$\xi_{\rm cr}^3 - 5\xi_{\rm cr} = 2\frac{\lambda_T}{\ell_{\rm mfp}}\quad. \tag{15.13}$$

For large values of the rhs this will occur at approximately $\xi_{\rm cr} \simeq (2\lambda_T/\ell_{\rm mfp})^{1/3}$. In the direction of the temperature gradient ($\boldsymbol{\xi}\cdot\nabla T > 0$) $f_1 < 0$ and the entire distribution function, $f_0 + f_1$, will be negative for $\xi > \xi_{\rm cr}$. This is clearly unphysical. The best we can hope is that this unphysical portion of $f(\boldsymbol{\xi})$ affects very few particles. If, for example, $\xi_{\rm cr} > 3$, then less than 1% of the particles will be affected. Demanding this we find we need

$$\lambda_T \;>\; \tfrac{1}{2}3^3\ell_{\rm mfp} \;=\; 13\,\ell_{\rm mfp}\quad, \tag{15.14}$$

$$\tau_h \;>\; \tfrac{1}{2}3^2\tau_{\rm c} \;=\; 4.5\,\tau_{\rm c} \tag{15.15}$$

for the Chapman-Enskog approach to be justified.

It is illustrative to consider the situation with $\nabla p = 0$ and $\nabla\mathbf{u} = 0$. In that case the spatial gradient of the Maxwellian, eq. (14.42), can be expressed

$$\nabla f_0 \;=\; \tfrac{1}{2}(\xi^2-5)f_0\nabla\ln T\quad. \tag{15.16}$$

The perturbation to the distribution function, eq. (15.11), is then

$$f_1 \;=\; -\,\tau_{\rm c}\delta\mathbf{v}\cdot\nabla f_0 \quad . \tag{15.17}$$

The complete distribution function

$$f(\mathbf{x},\mathbf{v}) \;=\; f_0 + f_1 \;=\; (1-\tau_{\rm c}\delta\mathbf{v}\cdot\nabla)f_0 \;\simeq\; f_0(\mathbf{x}-\tau_{\rm c}\delta\mathbf{v},\mathbf{v}) \quad , \tag{15.18}$$

is the Maxwellian evaluated at a position one collision away. The result of mixing together Maxwellians from different temperatures is not itself a Maxwellian. Its departure is the perturbation f_1. The displacement, $\delta\mathbf{x} = \tau_{\rm c}\delta\mathbf{v}$, will be very large for sufficiently large velocities $|\delta\mathbf{v}|$. In this case Taylor expansion is not warranted. If, however, it occurs for $|\delta\mathbf{v}| \gg v_{\rm th}$ the error will not be appreciable since $f_0(\delta\mathbf{v})$ will be very small.

15B.1 Thermal conductivity

Using eq. (15.11) in the heat flux, eq. (14.32), and recalling that f_0 contributes nothing to the third moment, gives

$$\mathbf{q} \;=\; \tfrac{1}{2}mv_{\rm th}^3\int \xi^2\boldsymbol{\xi}\, f\, d^3v \;=\; -\,\tfrac{1}{4}k_{\rm B}v_{\rm th}^2\nabla T\cdot\left[\int \tau_{\rm c}f_0(\xi)\Big(\xi^2-5\Big)\xi^2\boldsymbol{\xi}\boldsymbol{\xi}\, d^3v\right] \tag{15.19}$$

The tensor in square brackets will be isotropic, equal to the identity tensor times a third of its trace. The result is

$$\mathbf{q} \;=\; -\,\kappa\nabla T \quad , \tag{15.20}$$

where

$$\begin{aligned}\kappa \;&=\; \tfrac{1}{12}k_{\rm B}v_{\rm th}^2\tau_{\rm c}\int f_0(\boldsymbol{\xi})\Big(\xi^2-5\Big)\xi^4\, d^3v \\ &=\; \tfrac{1}{12}k_{\rm B}v_{\rm th}^2\tau_{\rm c}n(105-75) \;=\; \tfrac{5}{2}k_{\rm B}nv_{\rm th}^2\tau_{\rm c}\end{aligned} \tag{15.21}$$

The thermal diffusion coefficient is

$$\tilde\kappa \;=\; \tfrac{2}{3}\frac{\kappa}{nk_{\rm B}} \;=\; \tfrac{5}{3}v_{\rm th}^2\tau_{\rm c} \;=\; \tfrac{5}{3}\ell_{\rm mfp}^2/\tau_{\rm c} \quad , \tag{15.22}$$

where $\ell_{\rm mfp} = v_{\rm th}\tau_{\rm c}$ is the mean free path of a thermal particle. This is clearly consistent with the random walk with step size $\ell_{\rm mfp}$.

The lower bound on the gradient scale length λ_T, eq. (15.14), implies an upper bound on $|\nabla T|$ and therefore on $\mathbf{q}$. This is

$$|\mathbf{q}| < \tfrac{5}{2}nk_{\rm B}Tv_{\rm th}\frac{1}{13} = 0.13\, q_{\rm fs} \quad , \tag{15.23}$$

where $q_{\rm fs} = (3/2)nk_{\rm B}Tv_{\rm th}$ is the free-streaming heat flux: as if thermal energy were transported at the thermal velocity.

15B.2 Viscosity

Writing the pressure tensor in the form $P_{ij} = p\,\delta_{ij} + \pi_{ij}$ the first term is from the Maxwellian, f_0, and the second from f_1. Using eq. (15.11) in the moment (14.20) gives the stress tensor modification

$$\pi_{ij} \;=\; m v_{\rm th}^2 \int \xi_i \xi_j \, f_1 \, d^3v \;=\; -\tfrac{1}{2} m v_{\rm th}^2 n \tau_{\rm c} (\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) V_{kl} \quad , \tag{15.24}$$

using the fourth moment from the appendix. The factor involving δ_{kl} will lead to $\mathrm{Tr}(\underline{\underline{V}}) = 0$ since the rate-of-strain tensor is traceless. It is also symmetric so the other two terms are equal and the stress tensor becomes

$$\pi_{ij} \;=\; -\,m v_{\rm th}^2 n \tau_{\rm c}\, V_{ij} \;=\; -\,p \tau_{\rm c}\, V_{ij} \quad . \tag{15.25}$$

We see that π_{ij} is traceless so the BGK collision term does not modify the pressure. The coefficient is the dynamic viscosity, $\mu = p\tau_{\rm c}$. The kinematic viscosity is therefore $\nu = \mu/\rho = v_{\rm th}^2 \tau_{\rm c} = \ell_{\rm mfp}^2/\tau_{\rm c}$, also consistent with a random walk.

15C Strong magnetization

We had neglected the term, $\mathcal{L} f_1$, by assuming that it would be small compared the $\mathcal{L} f_0$. This is not, however, always the case. In fact we should compare $\tau_{\rm c}\mathcal{L} f_1$ to f_1. Several of the derivatives in $\mathcal{L}$ can be quickly dismissed. f_1 will solve an inhomogeneous equation to which $\mathcal{L} f_0$, given in eq. (15.6), is the source. Due to its dependence on moments, $n(\mathbf{x},t)$, $\mathbf{u}(\mathbf{x},t)$ and $T(\mathbf{x},t)$, the source will have spatial and temporal variation on hydrodynamic scales. We therefore expect the spatial and temporal derivatives, ∂_t and $\mathbf{v}\cdot\nabla$, of the solution, f_1, to scale with hydrodynamic frequencies. We take these frequencies to be small compared to the collision frequency, and therefore neglect those derivatives in $\tau_{\rm c}\mathcal{L} f_1$.

Neglecting the spatial and temporal derivatives in the Vlasov operator leaves

$$\tau_{\rm c}\mathcal{L} f_1 \;\simeq\; \frac{\tau_{\rm c} q}{m}\left[\mathbf{E} + \frac{\mathbf{v}}{c}\times\mathbf{B}\right]\cdot\frac{\partial f_1}{\partial \mathbf{v}} \;=\; \frac{\tau_{\rm c} q}{m}\mathbf{E}'\cdot\frac{\partial f_1}{\partial \mathbf{v}} + \frac{\tau_{\rm c} q B}{mc}\,\delta\mathbf{v}\times\hat{\mathbf{b}}\cdot\frac{\partial f_1}{\partial \mathbf{v}} \tag{15.26}$$

where $\mathbf{E}' = \mathbf{E} + (\mathbf{u}\times\mathbf{B})/c$ is the electric field in the frame moving with the particles' center of mass. Defining the Dreicer field to be $E_D = m v_{\rm th}/q\tau_{\rm c}$ the first term in the rightmost expression has magnitude $\sim (|\mathbf{E}|/E_D) f_1 \ll f_1$ provided the electric field is small. The remaining term can be rewritten using the fact that, to be Galilean invariant, f_1 will depend on $\mathbf{v}$ only through $\delta\mathbf{v}$,

$$\tau_{\rm c}\mathcal{L} f_1 \;\simeq\; \tau_{\rm c}\Omega\,\delta\mathbf{v}\times\hat{\mathbf{b}}\cdot\frac{\partial f_1}{\partial(\delta\mathbf{v})} \;=\; -\,\tau_{\rm c}\Omega\,\hat{\mathbf{b}}\times\boldsymbol{\xi}\cdot\frac{\partial f_1}{\partial\boldsymbol{\xi}} \quad , \tag{15.27}$$

where the cyclotron frequency $\Omega = qB/mc$ has the same sign as the charge. For even modestly magnetized plasma, $\tau_{\rm c}|\Omega| \geq 1$, and this term is comparable to f_1 itself.

Including this into the Boltzmann equation,and using the BGK collision operator, gives

$$\underbrace{\left[\,1 - \tau_{\rm c}\Omega\,\hat{\mathbf{b}}\times\boldsymbol{\xi}\cdot\frac{\partial}{\partial\boldsymbol{\xi}}\right]}_{\mathcal{M}} f_1 \;=\; -\,\tau_{\rm c}\mathcal{L} f_0 \quad , \tag{15.28}$$

The full solution requires the inversion of the operator $\mathcal{M}$. This inversion is trivial if the rhs depends only on $\xi = |\boldsymbol{\xi}|$ since $\mathcal{M} \cdot g(\xi) = g(\xi)$. The function $\mathcal{L}f_0$ depends on direction though factors $\boldsymbol{\xi} \cdot \nabla \ln T$ and $\boldsymbol{\xi}\boldsymbol{\xi} : V$. The inversion can therefore be achieved fairly simply by first establishing the relations

$$\begin{aligned}
\left(\hat{\mathbf{b}} \times \boldsymbol{\xi} \cdot \frac{\partial}{\partial \boldsymbol{\xi}}\right) \boldsymbol{\xi} &= \hat{\mathbf{b}} \times \boldsymbol{\xi} \\
\left(\hat{\mathbf{b}} \times \boldsymbol{\xi} \cdot \frac{\partial}{\partial \boldsymbol{\xi}}\right) \hat{\mathbf{b}} \times \boldsymbol{\xi} &= \hat{\mathbf{b}} \times (\hat{\mathbf{b}} \times \boldsymbol{\xi}) \;=\; \hat{\mathbf{b}}(\hat{\mathbf{b}} \cdot \boldsymbol{\xi}) - \boldsymbol{\xi} \\
\left(\hat{\mathbf{b}} \times \boldsymbol{\xi} \cdot \frac{\partial}{\partial \boldsymbol{\xi}}\right) \hat{\mathbf{b}}(\hat{\mathbf{b}} \cdot \boldsymbol{\xi}) &= 0
\end{aligned}$$

From these basic relations we can establish that

$$\mathcal{M} \cdot g(\xi) \left[\boldsymbol{\xi} + \tau_{\rm c}\Omega(\hat{\mathbf{b}} \times \boldsymbol{\xi}) + \tau_{\rm c}^2\Omega^2\, \hat{\mathbf{b}}(\hat{\mathbf{b}} \cdot \boldsymbol{\xi})\right] \;=\; (1 + \tau_{\rm c}^2\Omega^2) g(\xi)\boldsymbol{\xi} \tag{15.29}$$

15C.1 Thermal conductivity

The contribution to f_1 due to the temperature gradient alone is expressed by the equation

$$\mathcal{M} \cdot f_1 \;=\; -\tfrac{1}{2} v_{\rm th}\tau_{\rm c}\, f_0(\xi)\, (\xi^2 - 5)\boldsymbol{\xi} \cdot \nabla \ln T \quad .$$

Inverting this we find

$$\begin{aligned}
f_1(\mathbf{x}) &= -\tfrac{1}{2}\frac{v_{\rm th}\tau_{\rm c}}{1 + \tau_{\rm c}^2\Omega^2}\, f_0(\xi)\, (\xi^2 - 5)\, \left[\boldsymbol{\xi} + \tau_{\rm c}\Omega(\hat{\mathbf{b}} \times \boldsymbol{\xi}) + \tau_{\rm c}^2\Omega^2\, \hat{\mathbf{b}}(\hat{\mathbf{b}} \cdot \boldsymbol{\xi})\right] \cdot \nabla \ln T \\
&= -\tfrac{1}{2}\frac{v_{\rm th}\tau_{\rm c}}{1 + \tau_{\rm c}^2\Omega^2} f_0(\xi)\, (\xi^2 - 5)\boldsymbol{\xi} \cdot [I + \tau_{\rm c}^2\Omega^2 \hat{\mathbf{b}}\hat{\mathbf{b}}] \cdot \nabla \ln T \\
&\quad + \tfrac{1}{2}\frac{v_{\rm th}\tau_{\rm c}}{1 + \tau_{\rm c}^2\Omega^2} f_0(\xi)\, (\xi^2 - 5)\boldsymbol{\xi} \cdot (\tau_{\rm c}\Omega\hat{\mathbf{b}} \times \nabla \ln T)
\end{aligned} \tag{15.30}$$

Placing this into the heat flux gives

$$\mathbf{q} \;=\; -\,\frac{\kappa}{1 + \tau_{\rm c}^2\Omega^2}\left[I + \tau_{\rm c}^2\Omega^2 \hat{\mathbf{b}}\hat{\mathbf{b}}\right] \cdot \nabla T \;+\; \frac{\kappa}{1 + \tau_{\rm c}^2\Omega^2}\, \frac{\tau_{\rm c} q\mathbf{B}}{mc} \times \nabla T \tag{15.31}$$

where κ is the unmagnetized coefficient given in eq. (15.21). Using the projection matrix $\underline{\underline{\Pi}} = I - \hat{\mathbf{b}}\hat{\mathbf{b}}$ we can write this in the form

$$\mathbf{q} \;=\; -\left[\kappa\, \hat{\mathbf{b}}\hat{\mathbf{b}} \;+\; \frac{\kappa}{1 + \tau_{\rm c}^2\Omega^2}\, \underline{\underline{\Pi}}\right] \cdot \nabla T \;+\; \frac{\kappa}{1 + \tau_{\rm c}^2\Omega^2}\, \frac{\tau_{\rm c} q\mathbf{B}}{mc} \times \nabla T \tag{15.32}$$

The thermal diffusivity appears to be unimpeded in the direction parallel to the magnetic field. For very strong fields, $\Omega\tau_{\rm c} \gg 1$, perpendicular diffusion is characterized by a coefficient

$$\tilde{\kappa}_\perp \;=\; \tfrac{2}{3}\frac{\kappa}{n k_{\rm B}\Omega^2\tau_{\rm c}^2} \;=\; \tfrac{5}{3}\frac{v_{\rm th}^2}{\Omega^2\tau_{\rm c}} \;=\; \tfrac{5}{3} r_g^2/\tau_{\rm c} \quad , \tag{15.33}$$

where $r_g = v_{\rm th}/\Omega$ is the Larmour radius of the particle. It seems that perpendicular diffusion occurs as a random walk with step size limited by magnetization to a gyro-radius.

Finally we note that there is a heat flux *perpendicular* to the temperature gradient. For strong fields this becomes

$$\mathbf{q}_\wedge \;\simeq\; \frac{\kappa}{\Omega\tau_{\rm c}}\hat{\mathbf{b}}\times\nabla T \;=\; \tfrac{5}{2}\,\frac{nk_{\rm B}v_{\rm th}^2}{\Omega}\,\hat{\mathbf{b}}\times\nabla T \quad . \tag{15.34}$$

This does not depend on the collision operator or the collision time. It is evident from the final expression that the flux does not depend on the mass of the particles but has the sign of its charge.

The collision-independent term comes from the final contribution in eq. (15.30)

$$f_1 \;=\; \tfrac{1}{2}f_0(\xi)\,(\xi^2-5)\frac{\delta\mathbf{v}\times\hat{\mathbf{b}}}{\Omega}\cdot\nabla\ln T \quad , \tag{15.35}$$

using $\tau_{\rm c}\Omega \gg 1$. In the case with $\nabla p = 0$ and $\nabla\mathbf{u} = 0$ the gradient of the Maxwellian is given by eq. (15.16). In terms of this the complete distribution function is

$$f_0 + f_1 \;\simeq\; \left(1+\frac{\delta\mathbf{v}\times\hat{\mathbf{b}}}{\Omega}\cdot\nabla\right) f_0 \tag{15.36}$$

Under gyromotion a particle has a displacement from its guiding center, $\delta\mathbf{x} = \hat{\mathbf{b}}\times\delta\mathbf{v}/\Omega$. Expression (15.36) is the Maxwellian evaluated at the guiding center of that particle, $f_0(\mathbf{x}-\delta\mathbf{x})$. This is the origin of the finite-Larmour radius contribution to the heat flux $\mathbf{q}_\wedge$.

15C.2 Viscosity

The viscosity is a fourth-rank tensor relating the rate of strain tensor V_{kl} to the stress tensor π_{ij}:

$$\pi_{ij} \;=\; -\,\mu_{ijkl}V_{kl} \quad .$$

Without magnetization μ_{ijkl} had to appear the same under any rotation of coordinates: it was an isotropic tensor. There are only two such tensors, and only one produces a traceless result. Thus the viscosity, in an isotropic situation, is basically a scalar: μ such that $\underline{\underline{\pi}} = -\mu\underline{\underline{\rm V}}$.

The magnetic field breaks the isotropy so μ_{ijkl} will be more complicated. As a fourth rank tensor it contains $3^4 = 81$ coefficients which would, in the worst conceivable case, all be different. Fortunately our system is still invariant under rotations about the magnetic field direction, $\hat{\mathbf{b}}$: it is gyrotropic. This makes it somewhat simpler. Furthermore it couples two tensors which are each symmetric and traceless. There are a limited number of ways in which such a coupling can be done in a gyrotropic way.

For the time being we introduce coordinates $\hat{\mathbf{z}} = \hat{\mathbf{b}}$ so that $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are the directions perpendicular to $\hat{\mathbf{b}}$. A rotation by α about $\hat{\mathbf{z}}$ transforms the unit vectors

$$\hat{\mathbf{x}}\to\cos\alpha\hat{\mathbf{x}}+\sin\alpha\hat{\mathbf{y}} \quad , \quad \hat{\mathbf{y}}\to-\sin\alpha\hat{\mathbf{x}}+\cos\alpha\hat{\mathbf{y}} \quad , \quad \hat{\mathbf{z}}\to\hat{\mathbf{z}} \quad . \tag{15.37}$$

Any traceless, symmetric tensor can be written as a superposition of the five basis tensors

$$\begin{aligned} \underline{\underline{\rm m}}^{zz} \;&=\; \hat{\mathbf{z}}\hat{\mathbf{z}} - (\hat{\mathbf{x}}\hat{\mathbf{x}}+\hat{\mathbf{y}}\hat{\mathbf{y}})/2 \quad . \\ \underline{\underline{\rm m}}^{xz} = \hat{\mathbf{x}}\hat{\mathbf{z}}+\hat{\mathbf{z}}\hat{\mathbf{x}} \quad &, \quad \underline{\underline{\rm m}}^{yz} = \hat{\mathbf{y}}\hat{\mathbf{z}}+\hat{\mathbf{z}}\hat{\mathbf{y}} \quad , \\ \underline{\underline{\rm m}}^{xy} = \hat{\mathbf{x}}\hat{\mathbf{y}}+\hat{\mathbf{y}}\hat{\mathbf{x}} \quad &, \quad \underline{\underline{\rm m}}^{\times} = \hat{\mathbf{x}}\hat{\mathbf{x}}-\hat{\mathbf{y}}\hat{\mathbf{y}} \quad , \end{aligned}$$

For example the full rate-of-strain tensor is decomposed

$$\underline{\underline{\mathrm{V}}} \;=\; V_\times \underline{\underline{\mathrm{m}}}^\times + V_{xy}\underline{\underline{\mathrm{m}}}^{xy} + V_{xz}\underline{\underline{\mathrm{m}}}^{xz} + V_{yz}\underline{\underline{\mathrm{m}}}^{yz} + V_{zz}\underline{\underline{\mathrm{m}}}^{zz} \quad , \tag{15.38}$$

where $V_\times = (V_{xx} - V_{yy})/2$, and all other coefficients are those from the tensor itself. A similar decomposition of $\underline{\underline{\pi}}$ leads to five coefficients of stress. The viscosity is therefore a 5×5 matrix with no more that $5^2 = 25$ terms. Gyrotropy makes it simpler still.

The action of rotation, eq. (15.37), induces the following transformation among the matrices

$$\begin{aligned} \underline{\underline{\mathrm{m}}}^{zz} &\to \underline{\underline{\mathrm{m}}}^{zz} \quad , \\ \underline{\underline{\mathrm{m}}}^{xz} \to \cos\alpha\, \underline{\underline{\mathrm{m}}}^{xz} + \sin\alpha\, \underline{\underline{\mathrm{m}}}^{yz} \quad &, \quad \underline{\underline{\mathrm{m}}}^{yz} \to -\sin\alpha\, \underline{\underline{\mathrm{m}}}^{xz} + \cos\alpha\, \underline{\underline{\mathrm{m}}}^{yz} \quad , \\ \underline{\underline{\mathrm{m}}}^{xy} \to \cos 2\alpha\, \underline{\underline{\mathrm{m}}}^{xy} + \sin 2\alpha\, \underline{\underline{\mathrm{m}}}^{\times} \quad &, \quad \underline{\underline{\mathrm{m}}}^{\times} \to -\sin 2\alpha\, \underline{\underline{\mathrm{m}}}^{xy} + \cos 2\alpha\, \underline{\underline{\mathrm{m}}}^{\times} \quad . \end{aligned}$$

The tensors $\underline{\underline{\mathrm{m}}}^{xz}$ and $\underline{\underline{\mathrm{m}}}^{yz}$ transform among one another. The tensors $\underline{\underline{\mathrm{m}}}^{xy}$ and $\underline{\underline{\mathrm{m}}}^{\times}$ transform among one another, but in a different way. And finally, $\underline{\underline{\mathrm{m}}}^{zz}$, in unchanged.

In order to be invariant under these rotations (i.e. to be gyrotropic) viscosity may only couple coefficients from $\underline{\underline{\mathrm{V}}}$ to similarly-transforming coefficients from $\underline{\underline{\pi}}$. In other words π_{xz} may be proportional to V_{xz} or V_{yz} but not to V_{xy} or V_{zz}. In all there are therefore *five* different viscosities characterizing the five possible couplings. By convention they are designated μ_0, μ_1, $\ldots \mu_4$ and are defined by the relations

$$\pi_{zz} = -\mu_0 V_{zz} \tag{15.39}$$
$$\pi_{xy} = -\mu_1 V_{xy} + \mu_3 V_\times \tag{15.40}$$
$$\pi_\times = -\mu_3 V_{xz} - \mu_1 V_\times \tag{15.41}$$
$$\pi_{xz} = -\mu_2 V_{xz} - \mu_4 V_{yz} \tag{15.42}$$
$$\pi_{yz} = \mu_4 V_{xz} - \mu_2 V_{yz} \tag{15.43}$$

To see that this form of coupling is invariant under rotations, consider how two of the coefficients will transform to accommodate the new basis

$$\begin{bmatrix} V_{xy} \\ V_\times \end{bmatrix} \to \begin{bmatrix} V'_{xy} \\ V'_\times \end{bmatrix} = \begin{bmatrix} \cos 2\alpha & \sin 2\alpha \\ -\sin 2\alpha & \cos 2\alpha \end{bmatrix} \cdot \begin{bmatrix} V_{xy} \\ V_\times \end{bmatrix} \quad . \tag{15.44}$$

The most general possible relation between these two and their corresponding stresses is through a 2×2 matrix $\underline{\underline{\mu}}$

$$\begin{bmatrix} \pi_{xy} \\ \pi_\times \end{bmatrix} = -\underline{\underline{\mu}} \cdot \begin{bmatrix} V_{xy} \\ V_\times \end{bmatrix} \quad . \tag{15.45}$$

After transformation, the relation between the new quantities reads

$$\begin{bmatrix} \pi'_{xy} \\ \pi'_\times \end{bmatrix} = -\underbrace{\begin{bmatrix} \cos 2\alpha & \sin 2\alpha \\ -\sin 2\alpha & \cos 2\alpha \end{bmatrix} \cdot \underline{\underline{\mu}} \cdot \begin{bmatrix} \cos 2\alpha & -\sin 2\alpha \\ \sin 2\alpha & \cos 2\alpha \end{bmatrix}}_{\underline{\underline{\mu'}}} \cdot \begin{bmatrix} V'_{xy} \\ V'_\times \end{bmatrix} \tag{15.46}$$

While properties of the fluid, such as V_{xy}, are changed by the transformation we do not expect the form of the relations between them to change. In other words we expect the

matrices in eqs. (15.45) and (15.46) to be exactly the same: $\underline{\underline{\mu}} = \underline{\underline{\mu}}'$. This is the principle of symmetry.

The general coupling matrix $\underline{\underline{\mu}}$ can be decomposed into a sum over four possible 2×2 matrices.

$$\underline{\underline{\mu}} \;=\; \mu_1 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + \mu_3 \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} + \zeta_1 \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} + \zeta_2 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad . \tag{15.47}$$

The last two matrices are both symmetric and traceless and they transform among one another under the transformation in eq. (15.46). This first matrix, the identity, and the second, a purely anti-symmetric matrix, are unchanged. Thus for $\underline{\underline{\mu}}$ to remain unchanged it seems that $\zeta_1 = \zeta_2 = 0$ and the coupling is exactly as stated in eqs. (15.40) and (15.41).

An analogous argument can be made for the couplings in eqs. (15.42) and (15.43) except that all rotation matrices will have α in place of 2α. A proposed coupling between $[\pi_{xy}, \pi_\times]$ and $[V_{xz}, V_{yz}]$ will have a different kind of transformation with a 2α matrix on one side and a α matrix on the other. It can be shown that no matrix remains unchanged by such a transformation so an invariant coupling of this kind is impossible. This is the reason for stating above that coupling is only possible between tensor components which transform the same way.

If the magnetic field is reduced to zero then $\Omega \to 0$, the operator $\mathcal{M} \to 1$, the system should revert to the isotropic one: $\underline{\underline{\pi}} = -\mu \underline{\underline{\mathrm{V}}}$. This means,

$$\mu_0 \to \mu \quad , \quad \mu_1 \to \mu \quad , \quad \mu_2 \to \mu \quad , \quad \mu_3 \to 0 \quad , \quad \mu_4 \to 0 \quad .$$

Changing the sign of Ω is equivalent to changing either $\mathbf{B} \to -\mathbf{B}$ or $q \to -q$. This will reverse the sense of the particle's gyration, which could be undone by a the spatial reflection, $\hat{\mathbf{x}} \to -\hat{\mathbf{x}}$. The viscosity tensor will therefore be unchanged under this pair of transformations. The spatial reflection will change the matrices according to

$$\underline{\underline{\mathrm{m}}}^{xz} \to -\underline{\underline{\mathrm{m}}}^{xz} \quad , \quad \underline{\underline{\mathrm{m}}}^{xy} \to -\underline{\underline{\mathrm{m}}}^{xy} \quad ,$$

and all other matrices will be unchanged. Since the coefficients μ_3 and μ_4 couple terms which transform differently under reflection, each must change sign with the gyrofrequency. We therefore conclude that μ_3 and μ_4 must be *odd* functions of the gyrofrequency Ω, while all other coefficients will be even functions of Ω.

The same conclusion about the couplings of pairs of matrices could have been reached through more elegant reasoning by constructing the quadratic forms $Q(\xi, \theta, \phi) = \boldsymbol{\xi} \cdot \underline{\underline{\mathrm{m}}} \cdot \boldsymbol{\xi}$ for each of the five basis tensors. Each will naturally be ξ^2 times a function of θ, ϕ. Since $\nabla^2 Q = 2\mathrm{Tr}(\underline{\underline{\mathrm{m}}}) = 0$, the angular dependence must be composed of spherical harmonics $Y_2^m(\theta, \phi)$. The tensor $\underline{\underline{\mathrm{m}}}^{zz}$ is related to $m = 0$; $\underline{\underline{\mathrm{m}}}^{xz}$ and $\underline{\underline{\mathrm{m}}}^{yz}$ are related to $m = \pm 1$; and $\underline{\underline{\mathrm{m}}}^{xy}$ and $\underline{\underline{\mathrm{m}}}^{\times}$ to $m = \pm 2$. These correspondences show the transformation properties of the basis tensors as consequences of the transformation of spherical harmonics under coordinate rotation about the z axis: $\phi \to \phi - \alpha$.

Indeed, the rhs of eq. (15.28)

$$\mathcal{M} \cdot f_1 \;=\; -\tfrac{1}{2} \tau_c f_0(\xi)\, \boldsymbol{\xi} \cdot \underline{\underline{\mathrm{V}}} \cdot \boldsymbol{\xi} \quad , \tag{15.48}$$

includes exactly this quadratic form and will thus consist of terms of the form $f_0(\xi)\xi^2 Y_2^m(\theta,\phi)$. When expressed using spherical coordinates the operator

$$\mathcal{M} \;=\; 1 \;-\; \tau_c\Omega\,\hat{\mathbf{b}}\times\boldsymbol{\xi}\cdot\frac{\partial}{\partial\boldsymbol{\xi}} \;=\; 1 \;-\; \tau_c\Omega\frac{\partial}{\partial\phi} \tag{15.49}$$

clearly couples only terms with the same value of azimuthal mode number m. Applied to the axisymmetric term, $m=0$, i.e. $\underline{\underline{\mathrm{m}}}^{zz}$, the operator $\mathcal{M}=1$, and eq. (15.28) matches the unmagnetized one, eq. (15.11). This shows that $\mu_0=\mu=p\tau_c$, for the unmagnetized case.

The other terms, $m\neq 0$, include pairs of contributions which can be written using real versions of $Y_2^m(\theta,\phi)$,

$$f_1(\xi,\theta,\phi) \;=\; P_2^{(m)}(\cos\theta)\Big[\,a(\xi)\cos(m\phi) \;+\; b(\xi)\sin(m\phi)\,\Big] \quad ,$$

where $P_\ell^{(m)}$ is the associated Legendre function. The operator then becomes a matrix

$$\mathcal{M}\cdot f_1 \;\rightarrow\; \begin{bmatrix} 1 & -m\,\tau_c\Omega \\ m\,\tau_c\Omega & 1 \end{bmatrix}\cdot\begin{bmatrix} a \\ b \end{bmatrix} \quad .$$

which is easily inverted

$$\mathcal{M}^{-1} \;\rightarrow\; \frac{1}{1+m^2\tau_c^2\Omega^2}\begin{bmatrix} 1 & m\,\tau_c\Omega \\ -m\,\tau_c\Omega & 1 \end{bmatrix} \quad . \tag{15.50}$$

Based on this, and the arguments concerning the limits, $\Omega\rightarrow 0$, we can then write all the viscosity coefficients

$$\begin{aligned} \mu_0 \;&=\; \mu \;=\; p\tau_c \\ \mu_1 \;=\; \mu\frac{1}{1+4\tau_c^2\Omega^2} \quad &, \quad \mu_2 \;=\; \mu\frac{1}{1+\tau_c^2\Omega^2} \\ \mu_3 \;=\; \mu\frac{2\tau_c\Omega}{1+\tau_c^2\Omega^2} \quad &, \quad \mu_4 \;=\; \mu\frac{\tau_c\Omega}{1+4\tau_c^2\Omega^2} \end{aligned}$$

The five different magnetized viscosity coefficients behave similarly to the three coefficients of magnetized thermal conductivity. The analogy is particularly evident in the limit of strong magnetization, $\tau_c\Omega\gg 1$. The parallel component remains unchanged from the unmagnetized case, $\mu_0=\mu\sim v_{\rm th}^2\tau_c$. The diagonal perpendicular components, $\mu_2=4\mu_1\sim r_g^2/\tau_c$ actually increase with decreasing collision time. These are smaller than the parallel case by $(\tau_c\Omega)^{-2}\ll 1$. Finally, there are off-diagonal coefficients $\mu_4=2\mu_3\sim v_{\rm th}^2/\Omega$, independent of the collision time. These terms capture the deflection of momentum caused by the magnetic field.

15D Electrical conductivity

Electric conduction requires that both ions and electrons be considered. Collisions amongst electrons will drive them toward a Maxwellian distribution, $f_0(\mathbf{x},\mathbf{v})$, as described above. Electrons will also collide with ions, and these collisions will have a different effect on the

electron population. If we again invoke the second law of thermodynamics, however, inter-species collisions should still increase the entropy and therefore drive the distribution toward a Maxwellian. That Maxwellian $F_0(\mathbf{x},\mathbf{v})$, however, will depend on properties of the ions. For one thing, it will be centered at the ion mean velocity velocity $\mathbf{U}$

$$F_0(\mathbf{x},\mathbf{v}) \;=\; \frac{n(\mathbf{x})}{[2\pi k_{\rm B}T(\mathbf{x})/m_e]^{3/2}}\exp\left[-\frac{|\mathbf{v}-\mathbf{U}(\mathbf{x})|^2}{2v_{\rm th}^2}\right] \quad , \tag{15.51}$$

rather than the electron mean velocity $\mathbf{u}$. The electron-ion collisions will also change the ion distribution function, ultimately aiming to give both species the same temperature and mean velocity — that is the thermodynamic equilibrium for the system.

The two types of collisions will produce different changes in the electron distribution function $f(\mathbf{x},\mathbf{v})$. The changes will occur on different time scales: τ_{ee} for electron-electron collisions and τ_{ei} for ion-electron collisions. These can be captured in a two-term BGK-type collision operator

$$\mathcal{L}f \;=\; C_{\rm BGK}\{f\} \;=\; \frac{1}{\tau_{ei}}(F_0-f)+\frac{1}{\tau_{ee}}(f_0-f) \quad . \tag{15.52}$$

Expanding the solution as $f=f_0+f_1$, and neglecting magnetization for simplicity, gives the solution

$$f_1 \;=\; -\,\frac{\tau_{ei}\tau_{ee}}{\tau_{ei}+\tau_{ee}}\,\mathcal{L}f_0 \;+\; \frac{\tau_{ee}}{\tau_{ei}+\tau_{ee}}(F_0-f_0) \quad . \tag{15.53}$$

The ion-electron collision term does not preserve moments in the same way other $C\{f\}$ were assumed to. For one thing, its first moment

$$\left(\frac{d\mathbf{u}}{dt}\right)_{\rm col} \;=\; \frac{1}{n}\int \mathbf{v}\,C_{\rm BGK}\{f\}\,d^3\mathbf{v} \;=\; \frac{1}{n\tau_{ei}}\int \mathbf{v}[\,F_0(\mathbf{v})-f_0(\mathbf{v})\,]\,d^3\mathbf{v} \;=\; \frac{1}{\tau_{ei}}(\mathbf{U}-\mathbf{u}) \quad ,$$

describes a drag force. This drag seeks to bring the mean electron velocity to the same value as that of the ions. There is naturally a compensating drag on the ions, in order that the inter-species collisions preserve net momentum.

The drag term enters to electron momentum equation, (14.46)

$$\frac{D\mathbf{u}}{Dt} \;=\; -\,\frac{\nabla p}{m_e n} \;+\; \mathbf{a} \;+\; \frac{1}{\tau_{ei}}(\mathbf{U}-\mathbf{u}) \quad . \tag{15.54}$$

When the dominant terms are the drag and electric force, $\mathbf{a}=-e\mathbf{E}/m_e$, a balance will occur when

$$\mathbf{E} \;=\; \frac{m_e}{e\tau_{ei}}(\mathbf{U}-\mathbf{u}) \;=\; \frac{m_e}{e^2 n\tau_{ei}}\mathbf{J} \quad , \tag{15.55}$$

where $\mathbf{J}=ne(\mathbf{U}-\mathbf{u})$ is the electric current. This is Ohm's law with the conductivity from a Drude model

$$\sigma_e \;=\; \frac{e^2 n\tau_{ei}}{m_e} \;=\; \frac{\omega_{pe}^2\tau_{ei}}{4\pi} \quad . \tag{15.56}$$

We had earlier established that electrical conductivity led to magnetic field diffusion with diffusion coefficient

$$\eta \;=\; \frac{c^2}{4\pi\sigma_e} \;=\; \frac{(c/\omega_{pe})^2}{\tau_{ei}} \quad . \tag{15.57}$$

Like the viscous and thermal diffusion coefficients, $\nu = \mu/\rho$ and $\tilde{\kappa}$, the magnetic diffusion arises from a random walk with time-step given by the collision time. In this case the step size corresponds with the collisionless skin depth $d_e = c/\omega_{pe}$ which is independent of the collisions. This contrasts with the unmagnetized values of the other coefficients whose step sizes were a mean-free path and therefore proportional the collision time themselves. For this reason those other diffusion coefficients are proportional to collision time while the magnetic diffusion is *inversely* proportional to it. This latter situation is more analogous to the cross-field diffusion coefficients, whose step size is the gyroradius, and therefore does not depend on collision time.

15E Collision times and mean free paths

The foregoing has shown how fluid equations can be closed using the effect of collisions. To obtain this closure we assumed that collisions drive the distribution toward a Maxwellian on times τ_c much shorter than temperature gradients or velocity shear can drive it away. The result was thermal conduction or viscous stresses proportional to τ_c. We demonstrated this quantitatively using a simplified collision operator, $C_{\rm BGK}$ in eq. (15.7), in which the collision time was a parameter. The coefficients of thermal conductivity and viscosity were each proportional to this parameter.

In order to derive the collision time, τ_c, or equivalently the mean-free path, $\ell_{\rm mfp} = v_{\rm th}\tau_c$, from first principles it is necessary to consider collisions in detail. To do this legitimately we should begin with a collision operator such as the Boltzmann operator, in eq. (14.10), and re-derive the perturbation f_1 as we did with the simplified, BGK, operator. This is extremely cumbersome and we will not attempt it. Instead we will use the dynamics of pairwise collisions to derive a mean-free path. We then use this, or the collision time, $\tau_c = \ell_{\rm mfp}/v_{\rm th}$, in the expressions we derived from the BGK analysis.

The simplest case, resembling a neutral gas, is one where hard spheres of radius R, with a number density n, interact by direct collision. One sphere will collide with another when their centers pass within a distance, $2R$. The mean free path between such collisions is

$$\ell_{\rm mfp} \;=\; \frac{1}{n\,\pi(2R)^2} \quad , \tag{15.58}$$

regardless of how fast the spheres are moving. Using this in eq. (15.25) gives a coefficient of dynamic viscosity

$$\mu \;=\; mn\,v_{\rm th}\,\ell_{\rm mfp} \;=\; \frac{m\,v_{\rm th}}{4\pi R^2} \;=\; \frac{\sqrt{mk_{\rm B}T}}{4\pi R^2} \quad . \tag{15.59}$$

A full solution of Boltzmann's equation, with collision operator eq. (14.10), yields a value different by only a factor of 0.55 (Lifshitz & Pitaevskii, 1981, see also table 1.1). This simple picture does capture the magnitude and temperature dependence of viscosity in a gas of neutral particles like atoms or molcules: raising the temperature by a factor of two increases the viscosity by $\sqrt{2} \simeq 140\%$. Naturally, thermal conductivity has the same temperature dependence.

The charged particles in a plasma, electrons and ions, do not interact by bumping into one another like spheres. Instead they "collide" by electrostatic repulsion, so-called *Coulomb*

collisions, which is a more subtle process than the simple hard sphere collisions. To get a feel for the subtlety consider a particle of charge q_1 moving at speed v interacting with a stationary "target" charge q_2. We can define a "collision" using the impact parameter, b, which would bring the two close enough that the magnitude of their electrostatic potential would exceed their initial kinetic energy:

$$\frac{|q_1 q_2|}{b} \;>\; \tfrac{1}{2} m\, v^2 \quad . \tag{15.60}$$

In the actual trajectory the incident particles would be severely deflected by such an interaction; we define this as a "collision". This defines an effective collision cross section

$$\sigma_{\rm eff} \;\simeq\; \pi b^2 \;=\; \frac{4\pi e^4}{m^2\, v^4} \quad , \tag{15.61}$$

assuming $|q_1| = |q_2| = e$. This shows that Coulomb collisions, unlike hard-spheres, have a cross section which depends on the collision speed. Faster particles have smaller cross sections than slower particles.

We can use this approximate treatment to estimate a mean-free path. Considering the average pair of particles, moving at relative speed roughly, $v_{\rm th} = \sqrt{k_{\rm B}T/m}$, giving

$$\ell_{\rm mfp} \;=\; \frac{1}{n\sigma_{\rm eff}} \;\sim\; \frac{m^2 v_{\rm th}^4}{4\pi\, e^4 n} \;=\; \frac{(k_{\rm B}T)^2}{4\pi e^4 n} \quad . \tag{15.62}$$

Using this in eq. (15.21) gives a coefficient of thermal conductivity

$$\kappa \;=\; \tfrac{5}{2} k_{\rm B} n\, v_{\rm th}\, \ell_{\rm mfp} \;\sim\; \frac{5 k_{\rm B} (k_{\rm B}T)^{5/2}}{8\pi e^4 \sqrt{m}} \quad , \tag{15.63}$$

which depends on temperature alone, and scales as $T^{5/2}$.

The viscosity scales identically with T. The contrast with hard spheres, where $\mu \sim T^{1/2}$ is due to the temperature dependence of the Coulomb cross section in contrast to the fixed cross section of the spheres.

For a more precise treatment we consider a single particle initially moving at speed v_0, and encountering a sea of stationary targets. A single interaction, by Coulomb repulsion from a stationary target with impact parameter b, is the classical problem of Rutherford scattering treated in most undergraduate texts on classical mechanics (see for example Marion & Thornton, 1988). The path of the incident particle will be deflected by an angle

$$\Delta\theta \;=\; \cos^{-1}\left(\frac{m^2 v_0^4 - e^4/b^2}{m^2 v_0^4 + e^4/b^2}\right) \quad . \tag{15.64}$$

A large deflection, say $\Delta\theta > 90^\circ$, will occur only when $b < b_0 = e^2/m v_0^2$, in approximate agreement with eq. (15.60).

If the incident particle passes well outside this, $b \gg b_0$, then its path will be deflected only slightly by the weak electrostatic repulsion. This is called a *small-angle* scattering event, and would seem to be slight enough to be ignorable. There are, however, far more weak scattering events than strong scattering events. So many more, in fact, that their

effect turns out to be the *dominant* effect — collisions in plasmas are dominated by weak scattering.

The deflection of the particle's path by any scattering, eq. (15.64), will decrease the velocity component along the initial direction by

$$\delta v_s \;=\; [\,\cos(\Delta\theta) - 1\,]\, v_0 \;=\; -\,\frac{2e^4}{e^4 + m^2 v_0^4\, b^2}\, v_0 \quad . \tag{15.65}$$

As the particle passes through a slab of thickness Δs, it encounters a number of scatterers

$$N_s \;=\; n\, 2\pi\, b\, db\, \Delta s \quad , \tag{15.66}$$

whose impact parameter falls between b and $b + db$. Each of these will decrease the velocity by δv_s. The total decrease due to all targets out to a maximum impact parameter $b_{\max}$ is

$$\begin{aligned} \Delta v_s \;&=\; -n\, 2\pi \Delta s\, v_0 \int_0^{b_{\max}} \frac{2e^4\, b\, db}{e^4 + m^2 v_0^4\, b^2} \\ &=\; -\Delta s\, \frac{2\pi\, n\, e^4}{m^2 v_0^3}\, \ln(1 + m^2 v_0^4 b_{\max}^2/e^4) \quad . \end{aligned} \tag{15.67}$$

The velocity component thus decreases with distance as

$$\frac{dv_s}{ds} \;=\; -\,\frac{2\pi\, n\, e^4}{m^2 v_s^3}\, \ln(1 + m^2 v_s^4 b_{\max}^2/e^4) \quad . \tag{15.68}$$

This can, in principle, be solved to find $v_s(s)$.

While the solution of eq. (15.68) will never vanish, it will decrease significantly over a distance

$$L_{\rm slow} \;=\; \frac{v_0}{|dv_s/ds|} \;=\; \frac{m^2 v_0^4}{2\pi\, n\, e^4}\, \frac{1}{\ln(1 + m^2 v_0^4 b_{\max}^2/e^4)} \quad . \tag{15.69}$$

Restricting consideration to large-angle scattering, by taking $b_{\max} = b_0 = e^2/mv_0^2$, returns a result very similar to the mean-free path in eq. (15.62). Thus we see that scattering by large-angle Coulomb collisions behaves roughly like those of simple hard sphere collisions, except with a velocity dependent cross section. Including, however, the many small-angle collisions, by taking $b_{\max} \gg b_0$, will reduce $L_{\rm slow}$. Indeed, if we attempt to include *all* Coulomb interactions by taking $b_{\max} \to \infty$, we find $L_{\rm slow} \to 0$. This occurs because, while the Coulomb field decreases with distance, the number of interacting particles *increases* at the same rate. This is the aspect of Coulomb collisions which is very different than simple hard sphere collisions: Coulomb interactions are dominated by many small angle scatterings, not by individual pairwise scatterings. As a result, the Boltzmann operator, eq. (14.10), is not applicable in the case of Coulomb collisions which describe a plasma. One must use a different collision operator, the Fokker-Planck operator, which accounts for many simultaneous weak interactions. We will continue to avoid such rigorous treatment, and derive an approximate expression for $\ell_{\rm mfp}$ using single-particle dynamics.

If Coulomb interactions did occur over arbitrarily large distances, and $b_{\max} \to \infty$, then it would be hard to escape the conclusion that $\ell_{\rm mfp} \to 0$ — there would be too many

interactions. While electrostatic fields extend indefinitely when the universe contains only two particles, this is fortunately not the case in a plasma where numerous particles are all interacting with each other at once. We have already mentioned the fact that a plasma's many charged particles will neutralize any free charge over a time $\omega_p^{-1} = \sqrt{m/4\pi e^2 n}$, called the plasma period. This same effect limits the distance over which the electric field of any single charge can be felt. The sea of other charges, moving at $v_{\rm th} = \sqrt{k_{\rm B}T/m}$, will screen out the field over a distance

$$\lambda_D \;=\; \frac{v_{\rm th}}{\omega_p} \;=\; \sqrt{\frac{k_{\rm B}T}{4\pi e^2 n}} \;=\; 0.2\,{\rm cm}\left(\frac{T}{10^6\,{\rm K}}\right)^{1/2}\left(\frac{n_e}{10^9\,{\rm cm}^{-3}}\right)^{-1/2} \quad , \tag{15.70}$$

known as the Debeye length, (we derive this more rigorously in the next chapter). This is an extremely small distance (millimeters), in astrophysical terms, but is quite substantial compared to the spacing between particles, $n^{-1/3} \sim 10^{-3}$ cm. Indeed, any single particle is electrostatically interacting simultaneously with a very large number of other particles

$$\Lambda \;=\; \frac{4\pi}{3}\lambda_D^3 n \;=\; \frac{(k_{\rm B}T)^{3/2}}{3\sqrt{4\pi}e^3\,n^{1/2}} \;=\; 4.3\times 10^7\left(\frac{T}{10^6\,{\rm K}}\right)^{3/2}\left(\frac{n_e}{10^9\,{\rm cm}^{-3}}\right)^{-1/2} \quad . \tag{15.71}$$

This dimensionless quantity, called the *plasma parameter*, characterizes the extent to which large-angle, pairwise collisions can be ignored: if $\Lambda \sim 1$ they are an important effect, but in all plasmas of astrophysical interest $\Lambda \gg 1$ and they are not.

Following the reasoning above we set we set $b_{\rm max} = \lambda_D$ since that is the approximate limit of Coulomb interactions in a plasma. The logarithmic term in the denominator of eq. (15.69) can then be written

$$\ln(1 + m^2 v_0^4 b_{\rm max}^2/e^4) \;\simeq\; \ln(1 + k_{\rm B}^3 T^3/4\pi e^6\, n) \;=\; \ln(1+9\Lambda^2) \;\simeq\; 2\ln\Lambda \quad , \tag{15.72}$$

after setting $v_0 = v_{\rm th}$, and using the fact that $\Lambda \gg 1$. The plasma parameter, Λ, is always so large for astrophysical plasmas that its logarithm, called the *Coulomb lograithm*, is typically $\ln\Lambda \sim 20$. Indeed, because the logarithm is so insensitive to its argument the value of $\ln\Lambda$ is typically ~ 20 for almost all plasmas. In a more rigorous treatment we should account for the smooth departure from Coulomb's inverse square law as b approaches and exceeds λ_D. This would serve to modify the numerical factor inside the logarithm, leading to a very small correction; we ignore this.

A particle moving significantly faster than the thermal speed, $v_0 \gg v_{\rm th}$, will be slowed by the collisions with the sea of thermal particles, which appear stationary to it. It will have slowed significantly over a distance, $L_{\rm slow}$, where

$$n\,L_{\rm slow} \;=\; \frac{m^2 v_0^4}{4\pi\, e^4 \ln\Lambda} \;=\; \frac{E_0^2}{\pi e^4 \ln\Lambda} \;\simeq\; 6.8\times 10^{17}\,{\rm cm}^{-2}\left(\frac{E_0}{1\,{\rm keV}}\right)^2 \quad , \tag{15.73}$$

is known as the *stopping* column. Electrons are routinely accelerated to energies of 10 keV, or more, in solar flares. Coulomb collisions in the corona ($n \simeq 10^9\,{\rm cm}^{-3}$), can slow them down to $v \sim v_{\rm th}$ (i.e. "thermalize" them), only over distances of $L_{\rm slow} \sim 7\times 10^{10}$ cm or more. In other words, these particles will remain at energies far above $k_{\rm B}T$ throughout the corona: they are *non-thermal particles*. Indeed, because of the strong energy-dependence in $L_{\rm slow}$, many plasmas contain a population of non-thermal, or *super-thermal*, particles for

which Coulomb collisions are too infrequent to permits their relaxation to a Maxwellian.[1] The full distribution function thus consists of a Maxwellian "core" ($v \lesssim 3v_{\rm th}$) plus "tails" extending to very high energies.

The particles within the Maxwellian core, $v_0 \simeq v_{\rm th}$, will travel in one direction for a distance

$$\ell_{\rm mfp} \;=\; \frac{m^2 v_{\rm th}^4}{4\pi\, e^4\, n\, \ln\Lambda} \;=\; \frac{(k_{\rm B}T)^2}{4\pi\, e^4\, n\, \ln\Lambda} \quad , \tag{15.74}$$

before their direction has been significantly changed: this is their mean free path. It is smaller by a factor $1/\ln\Lambda$ than our preliminary estimate, eq. (15.62), due to its inclusion of small angle collisions — the dominant source of scattering. Using this more correct expression in eq. (15.21) gives a coefficient of thermal conductivity

$$\kappa_j \;=\; \tfrac{5}{2} k_{\rm B} n\, v_{\rm th}\, \ell_{\rm mfp} \;=\; \frac{5 k_{\rm B} (k_{\rm B}T)^{5/2}}{8\pi e^4 \sqrt{m_j}\ln\Lambda} \quad , \tag{15.75}$$

for particles of species j with charge e and mass m_j. Mass appears in the denominator so it is evident that electrons conduct heat $\sqrt{1836} = 43$ times effectively than protons.[2] The conductivity of the plasma as a whole is therefore approximately that of its electrons

$$\kappa \;\simeq\; \kappa_e \;=\; \frac{5}{8\pi}\frac{k_{\rm B}^{7/2}}{e^4\sqrt{m_e}\ln\Lambda}\, T^{5/2} \quad . \tag{15.76}$$

A complete solution of the Fokker-Planck equation yields an expression with a numerical factor about five times larger but otherwise identical, $\kappa \simeq 10^{-6} T^{5/2}$, known as the Spitzer conductivity.

The viscosity of from particle species j follows similarly

$$\mu_j \;=\; m_j n\, v_{{\rm th},j}\, \ell_{\rm mfp} \;=\; \frac{\sqrt{m_j}(k_{\rm B}T)^{5/2}}{4\pi\, e^4\, \ln\Lambda} \quad . \tag{15.77}$$

The particle mass is now in the numerator, so it is the viscosity of the ions that transports the most momentum and dominates the viscosity of the plasma as a whole. It has the same strong $T^{5/2}$ temperature dependence as the thermal conductivity.

Electrical conductivity, σ_e, is the final transport coefficient determined by collisions. Considering the collisions between electrons and ions we derived an expression in eq. (15.56). Its inverse gives the magnetic diffusivity

$$\eta \;=\; \frac{c^2}{4\pi\sigma_e} \;=\; \frac{m_e c^2}{4\pi e^2 n_e}\frac{1}{\tau_{ei}} \;=\; \frac{m_e c^2}{4\pi e^2 n_e}\frac{v_{{\rm th},e}}{\ell_{\rm mfp}} \;=\; \frac{\sqrt{m_e} c^2 e^2\, \ln\Lambda}{(k_{\rm B}T)^{3/2}} \quad , \tag{15.78}$$

whose value was presented in eq. (8.13). This diffusion coefficient depends on temperature differently than either the kinematic viscosity, $\nu = \mu/m_p n_e \sim T^{5/2}$ or the thermal diffusion,

[1] One other well-known example are the galactic cosmic rays which fill our galaxy. These particles are accelerated in astrophysical shocks, probably driven by supernova explosions, to such high energies that $L_{\rm slow}$ exceeds the diameter of the Milkyway.

[2] Our derivation used scattering angle (15.64), appropriate for a stationary target. In fact both particles will initially be moving, and in the case of proton-proton scattering, the target proton will be deflected as well. Accounting correctly for these subtleties changes mean-free path slightly.

$\tilde{\kappa} \sim T^{5/2}$. The magnetic diffusivity, or "resistivity", of the plasma *decreases* as the temperature increases. This behavior also contradicts our experience with solid conductors, where the resistivity *increases* when the temperature increases. In both cases the way to decrease the resistivity is to increase the mean-free path of the charge carriers (electrons). Electrons in a solid scatter off fluctuations of the ion lattice, and their mean free path increases as those fluctuations are reduced by decreasing temperature. In a plasma, on the other hand, electrons scatter electrostatically from freely moving ions. As the temperature increases, the scattering cross section decreases, and so the mean-free path increases.

Chapter 16

Two Fluid Theory

The most recent chapters have considered a collection of identical particles with a distribution of positions and velocities described by a single distribution function $f(\mathbf{x}, \mathbf{v}, t)$. This is sufficient to describe a neutral fluid, such as an unionized gas. A plasma, however, is composed of particles of several distinct types, or *species*. In most cases, there are free electrons, bare protons, and perhaps some neutral Hydrogen. Astrophysical plasmas also include many other elements, each at a variety of ionization states.

A particular species, call it s, must be described by its own distribution function, $f_s(\mathbf{x}, \mathbf{v}, t)$, whose moments describe the density, fluid velocity and pressure of that constituent

$$n_s(\mathbf{x}, t) = \int f_s(\mathbf{x}, \mathbf{v}, t)\, d^3v \quad , \tag{16.1}$$

$$\mathbf{u}_s(\mathbf{x}, t) = \frac{1}{n_s(\mathbf{x}, t)} \int \mathbf{v} f_s(\mathbf{x}, \mathbf{v}, t)\, d^3v \quad , \tag{16.2}$$

$$P_{s,ij}(\mathbf{x}, t) = m_s \int (v_i - u_{s,i})(v_j - u_{s,j}) f_s(\mathbf{x}, \mathbf{v}, t)\, d^3v \quad , \tag{16.3}$$

where m_s is the mass of a single particle of species s. If there were no collisions between particles then each distribution function would evolve according to Vlasov's equation, eq. (14.9). The derivation of Chapter 14 could then be used on the Vlasov equation of each species to derive separate fluid equations for each species. The continuity and momentum equation for collisionless species s, for example, would be

$$\frac{\partial n_s}{\partial t} + \nabla \cdot (n_s \mathbf{u}_s) = 0 \quad , \tag{16.4}$$

$$m_s n_s \left[\frac{\partial \mathbf{u}_s}{\partial t} + (\mathbf{u}_s \cdot \nabla) \mathbf{u}_s \right] = -\nabla \cdot \underline{\underline{\mathrm{P}}}_s + \mathbf{f}_s \quad , \tag{16.5}$$

where $\mathbf{f}_s(\mathbf{x}, t)$ is the force density experienced by particles of species s. There would also be an energy equation for each species.

In order to close a system of fluid equations it is necessary to write the pressure tensor, $P_{s,ij}$, and heat flux, $q_{s,i}$, in terms of the scalar pressure, p_s, and other fluid variables. Provided this can be done a plasma with N distinct species would then be described by a *set* of $3N$ equations, called a *multi-fluid* system. If such a closure is not possible then one

must be satisfied with a set of N different Vlasov equations. In some cases it is possible to treat one species as a fluid but not another; models of fluid ion and kinetic (i.e. Vlasov) electrons, is an example often encountered in the literture.

If the collision term for a particular species obeyed the same conservation laws used in Chapter 14, eqs. (14.12)–(14.14), then the fluid equation of each species would be exactly as that of the single-species fluid. This will not, however, be the case in general. Pairwise collisions will still preserve momentum. A collision between particles of different species will, however, change the net momentum of each species. Thus electro-proton collisions can change the momentum of electrons as a whole, and of protons as a whole. Only when the momenta of all species are combined will momentum conservation be recovered.

There is also the possibility for a collision between particles of two species, say an electron and a proton, to result in a particle of an entirely new species: a neutral Hydrogen. Such collisions do not even preserve particle number (i.e. eq. [14.12] is violated). This would lead to a sink or source term in the continuity equation of a particuar species, eq. (16.1). The description of collisions between different species can therefore be a complicated and time-consuming undertaking. We will forego this topic for the time being, interesting as it would be, and return to it after we get a feel for multi-fluid equations.

In order to simplify things still farther we will restrict ourselves to a fully ionized, pure Hydrogen plasma. There are then only two species and s will assume the value e, for electrons, and i for ions (i.e. protons).[1] Since we are neglecting (for the present) collisions between electrons and protons it might seem they would evolve independently of one another. Each is effected by electric and magnetic fields leading to the force density

$$\mathbf{f}_s \;=\; q_s n_s \mathbf{E} \;+\; \frac{q_s n_s \mathbf{u}_s}{c} \times \mathbf{B} \quad . \tag{16.6}$$

The electric and magnetic fields satisfy Gauss's law and Ampère's law whose sources, the charge density and current density, depend on the properties of both fluids

$$\nabla \cdot \mathbf{E} \;=\; 4\pi e (n_i - n_e) \quad , \tag{16.7}$$

$$\nabla \times \mathbf{B} - \frac{1}{c}\frac{\partial \mathbf{E}}{\partial t} \;=\; \frac{4\pi e}{c}(n_i \mathbf{u}_i - n_e \mathbf{u}_e) \quad . \tag{16.8}$$

It is thus that the electrons and ions interact with one another in a collisionless plasma.

16A Linear waves in a two-fluid plasma: general remarks

To obtain a sense of how a two-fluid treatment relates to single fluid hydrodynamics we linearize the fluid equations about a homogeneous, static equilibrium. The velocity of each species vanishes in equilibrium: $\mathbf{u}_i = \mathbf{u}_e = 0$. The pressures, $p_{0,p}$ and $p_{0,e}$, are both uniform but need not be the same. We assume also that there is no electric field in equilibrium. Gauss's law, eq. (16.7), then requires that the equilibrium number density of each species be the same

$$n_{0,i} \;=\; n_{0,e} \;=\; n_0 \quad , \tag{16.9}$$

[1] It would seem more natural to use p to recall that all ions are protons. But it is conventional to use the subscript i even in an electron-proton plasma.

which we call n_0 without ambiguity.

Departure from charge neutrality will be first order in the linearization,

$$\rho_{c,1} \;=\; e\,(n_{1,i} - n_{1,e}) \quad , \tag{16.10}$$

as will be current density,

$$\mathbf{J}_1 \;=\; n_0 e\,(\mathbf{u}_{1,i} - \mathbf{u}_{1,e}) \quad . \tag{16.11}$$

This means that Ampère's law and Faraday's law are each linear at lowest order

$$\partial_t \mathbf{B}_1 \;=\; -c\nabla \times \mathbf{E}_1 \tag{16.12}$$

$$\partial_t \mathbf{E}_1 \;=\; c\nabla \times \mathbf{B}_1 - 4\pi\, n_0\, e(\mathbf{u}_{1,i} - \mathbf{u}_{1,e}) \quad . \tag{16.13}$$

We have retained the displacement current, i.e. $\partial_t \mathbf{E}$, in Ampere's law because we are no longer restricting consideration to low frequencies. To these first order equations we add the linearized fluid equations for each species

$$\partial_t n_{1,s} \;=\; -n_0\, \nabla \cdot \mathbf{u}_{1,s} \tag{16.14}$$

$$n_0 m_s\, \partial_t \mathbf{u}_{1,s} \;=\; -\nabla p_{1,s} + n_0 q_s\, \mathbf{E}_1 \tag{16.15}$$

$$\partial_t p_{1,e} \;=\; -\gamma p_{0,s} \nabla \cdot \mathbf{u}_{1,s} \quad , \tag{16.16}$$

after discarding the Lorentz force as second order ($\sim \mathbf{u}_{1,s} \times \mathbf{B}_1$).

Accounting for all vector components and both species these constitute 16 differential equations, first order in t, so we might expect 16 different wave solutions. Two of these turn out to be unphysical waves. The divergence of eq. (16.12) yields

$$\partial_t(\nabla \cdot \mathbf{B}_1) \;=\; 0 \quad , \tag{16.17}$$

reflecting the fact, remarked upon several times before, that $\nabla \cdot \mathbf{B} = 0$ serves as an initial condition. If we propose a perturbation that violates this condition, that violation will persist. This is an unphysical "wave" with zero frequency.

The divergence of Ampère's law, eq. (16.13), can be rewritten

$$\partial_t \nabla \cdot \mathbf{E}_1 \;=\; -4\pi e n_0 (\nabla \cdot \mathbf{u}_{1,i} - \nabla \cdot \mathbf{u}_{1,e}) \;=\; 4\pi e (\partial_t n_{1,i} - \partial_t n_{1,e}) \quad , \tag{16.18}$$

after using the continuity equation of each species. This leads to

$$\partial_t [\nabla \cdot \mathbf{E}_1 \;-\; 4\pi e (n_{1,i} - n_{1,e})] \;=\; 0 \quad , \tag{16.19}$$

describing a situation is analogous to that of the magnetic divergence. Poisson's equation,

$$\nabla \cdot \mathbf{E}_1 \;=\; 4\pi e (n_{1,i} - n_{1,e}) \quad , \tag{16.20}$$

should be imposed as an initial condition. If our perturbation violates it, this violation will persist. This is a second unphysical wave with zero frequency.

The two unphysical waves are removed by requiring that the initial conditions satisfy the two of Maxwell's equations with divergences. There remain then 14 genuine equations from which we should find 14 linear waves. Recall that the system of single-fluid MHD equations admits 7 different waves: an entropy wave and two each of Alfvén, fast magnetosonic and

slow magnetosonic waves. Neutral fluids have only 5 of these, since in the limit $|\mathbf{B}_0| \to 0$ the Alfvén and slow modes (4 waves) become zero frequency modes: 2 shear modes and 2 "magnetic ghost" modes. It is shown below that the two-fluid plasma includes all of the MHD and neutral fluid waves at its low-frequency, long-wavelength limits. The remaining waves lie outside the MHD approximation, having either very high frequencies or short wavelengths.

16B Linear waves in an unmagnetized two-fluid plasma

Since we chose a homogeneous equilibrium we know in advance that plane waves will satisfy the linear equations

$$n_{1,i}(\mathbf{x},t) \;=\; \hat{n}_{1,i}\, e^{i\mathbf{k}\cdot\mathbf{x}-i\omega t} \quad ,$$

and so forth. With these in place the linearized fluid equations for species s are

$$\omega\, \hat{n}_{1,s} \;=\; n_0\, \mathbf{k}\cdot\hat{\mathbf{u}}_{1,s} \quad , \tag{16.21}$$

$$\omega\, m_s n_0 \hat{\mathbf{u}}_{1,s} \;=\; \mathbf{k}\, \hat{p}_{1,s} \;+\; i\, q_s n_0 \hat{\mathbf{E}}_1 \quad , \tag{16.22}$$

$$\omega\, \hat{p}_{1,s} \;=\; \gamma p_{0,s} \mathbf{k}\cdot\hat{\mathbf{u}}_{1,s} \quad . \tag{16.23}$$

The Fourier transforms of Faraday's and Ampère's laws, eqs. (16.12) and (16.13), yield

$$\omega\hat{\mathbf{B}}_1 \;=\; c\mathbf{k}\times\hat{\mathbf{E}}_1 \tag{16.24}$$

$$\omega\hat{\mathbf{E}}_1 \;=\; -c\mathbf{k}\times\hat{\mathbf{B}}_1 \;-\; 4\pi i\, en_0(\hat{\mathbf{u}}_{1,i}-\hat{\mathbf{u}}_{1,e}) \quad . \tag{16.25}$$

In order to eliminate the unphysical waves we must insist that the perturbation satisfy the time-independent conditions

$$i\mathbf{k}\cdot\hat{\mathbf{B}}_1 \;=\; 0 \tag{16.26}$$

$$i\mathbf{k}\cdot\hat{\mathbf{E}}_1 \;=\; 4\pi e(\hat{n}_{1,i}-\hat{n}_{1,e}) \quad . \tag{16.27}$$

This means that the components of $\hat{\mathbf{E}}_1$ and $\hat{\mathbf{B}}_1$ parallel to the wave vector are fixed by the initial conditions and need not be found from (16.24) or (16.25).

With no velocity or magnetic field the equilibrium lacks a distinct direction — it is *isotropic*. Linear waves in isotropic equilibria must be either *longitudinal* or *transverse*. Longitudinal waves are perturbations to scalar quantities, namely $\hat{n}_{1,i}$, $\hat{n}_{1,e}$, $\hat{p}_{1,i}$ and $\hat{p}_{1,e}$, and to vector components parallel to $\mathbf{k}$, namely $\hat{\mathbf{k}}\cdot\hat{\mathbf{u}}_{1,i}$ and $\hat{\mathbf{k}}\cdot\hat{\mathbf{u}}_{1,e}$ (recall that the parallel components of both $\hat{\mathbf{E}}_1$ and $\hat{\mathbf{B}}_1$ are replaced using eq. [16.26] and [16.27]). The longitudinal modes therefore comprise a 6×6 system. Transverse waves are perturbations to vector quantities perpendicular to $\mathbf{k}$. There are two such components in each of the four vector quantities, $\hat{\mathbf{u}}_{1,i}$, $\hat{\mathbf{u}}_{1,e}$, $\hat{\mathbf{E}}_1$ and $\hat{\mathbf{B}}_1$. Transverse modes therefore comprise an 8×8 system. Inspection of eq. (16.21)– (16.25) confirm that these two systems, the 6×6 longitudinal modes and the 8×8 transverse modes, are uncoupled from one another. We consider them separately below.

16B.1 Transverse waves

To simplify analysis in this section we choose our coordinates so that $\mathbf{k} = k\hat{\mathbf{z}}$. The two transverse directions are therefore $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$. It is evident from the vector equations, (16.22), (16.24) and (16.25), that $\hat{u}_{s,x}$, $\hat{E}_x$ and $\hat{B}_y$ are mutually coupled, and likewise are $\hat{u}_{s,y}$, $\hat{E}_y$ and $\hat{B}_x$. We consider only the first of these, which constitute the 4×4 system

$$\omega \begin{bmatrix} \hat{u}_{i,x} \\ \hat{u}_{e,x} \\ \hat{B}_y \\ \hat{E}_x \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & i\,e/m_p \\ 0 & 0 & 0 & -i\,e/m_e \\ 0 & 0 & 0 & ck \\ -4\pi i e n_0 & 4\pi i e n_0 & ck & 0 \end{bmatrix} \cdot \begin{bmatrix} \hat{u}_{i,x} \\ \hat{u}_{e,x} \\ \hat{B}_y \\ \hat{E}_x \end{bmatrix} \quad . \tag{16.28}$$

Two zero-frequency right-eigenvectors

$$\underline{\hat{\mathrm{U}}}^{(1)} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad , \quad \underline{\hat{\mathrm{U}}}^{(2)} = \begin{bmatrix} 1 \\ -1 \\ 8\pi i e n_0/ck \\ 0 \end{bmatrix} \quad , \tag{16.29}$$

can be found. The first can be identified as a shear mode with no current. The second is a current perpendicular to $\mathbf{k}$. The Lorentz force offers no restoring force since it is second order, $c^{-1}\mathbf{J}_1 \times \mathbf{B}_1$. The corresponding left eigenvectors (i.e. right eigenvectors of the transpose)

$$\underline{\hat{\mathrm{W}}}^{(1)} = \begin{bmatrix} m_p \\ m_e \\ 0 \\ 0 \end{bmatrix} \quad , \quad \underline{\hat{\mathrm{W}}}^{(2)} = \begin{bmatrix} i\,m_p \\ -i\,m_e \\ 2e/ck \\ 0 \end{bmatrix} \quad , \tag{16.30}$$

are orthogonal to the right eignevectors of all other modes.

A general vector which is orthogonal to both of the zero-frequency left eigenvectors at once can be written

$$\begin{bmatrix} \hat{u}_{i,x} \\ \hat{u}_{e,x} \\ \hat{B}_y \\ \hat{E}_x \end{bmatrix} = \begin{bmatrix} i\,e/m_p ck \\ -i\,e/m_e ck \\ 1 \\ 0 \end{bmatrix} \hat{B}_y + \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \hat{E}_x \quad , \tag{16.31}$$

for arbitrary $\hat{B}_y$ and $\hat{E}_x$. This vector thus lies in the two-dimensional space orthogonal to the zero-frequency modes. Substituting it into eq. (16.28), and retaining only the bottom two rows, yields the 2×2 system

$$\omega \begin{bmatrix} \hat{B}_y \\ \hat{E}_x \end{bmatrix} = \begin{bmatrix} 0 & ck \\ \omega_p^2/ck + ck & 0 \end{bmatrix} \cdot \begin{bmatrix} \hat{B}_y \\ \hat{E}_x \end{bmatrix} \quad , \tag{16.32}$$

describing the other dynamics. We have introduced the *plasma frequency* for species s

$$\omega_{p,s} = \sqrt{\frac{4\pi e^2 n_0}{m_s}} \quad , \tag{16.33}$$

and for the entire plasma

$$\omega_p \;=\; \sqrt{\omega_{p,i}^2+\omega_{p,e}^2} \;=\; \omega_{p,e}\sqrt{1+\frac{m_e}{m_p}} \quad . \tag{16.34}$$

The eigenfrequencies for this 2×2 system are quickly found

$$\omega_{3,4} \;=\; \pm\sqrt{\omega_p^2+c^2k^2} \;=\; 0 \quad . \tag{16.35}$$

It is evident from the dispersion relation, eq. (16.35), that eigenfrequencies are bounded from below by the plasma frequency, $|\omega|\geq\omega_p$: that serves as a *cut-off frequency* for transverse waves in the plasma. If we attempt to drive waves of lower frequency, say by placing an antenna inside the plasma, they will decay with distance as $e^{-x\sqrt{\omega_p^2-\omega^2}/c}$. In the limit of very low frequency the waves decay over a scale $\delta=c/\omega_p$, known as the *collisionless skin depth* of the plasma. Waves with frequencies $|\omega|>\omega_p$ will propagate with phase speed

$$v_\phi \;=\; \frac{\omega}{k} \;=\; \frac{c}{\sqrt{1-\omega_p^2/\omega^2}} \quad . \tag{16.36}$$

For very high frequencies, $|\omega|\gg\omega_p$, this approaches the speed of light, which suggests that these transverse electromagnetic waves are related to traditional light waves in vacuum. It is notable that the phase speed is always *greater* than c, and approaches $v_\phi\to\infty$ at the cut-off frequency. This is very common for waves with lower cut-offs and serves to remind us that only the *group* velocity is required to remain sub-luminal.

The eigenmode itself has the property

$$\hat{B}_y \;=\; \frac{ck}{\omega}\hat{E}_x \;=\; \pm\sqrt{1-\frac{\omega_p^2}{\omega^2}}\,\hat{E}_x \quad . \tag{16.37}$$

At very high frequency, $|\omega|\gg\omega_p$, this goes over to the classic electromagnetic wave in vacuum: $\hat{\mathbf{B}}_1=\pm\hat{\mathbf{k}}\times\hat{\mathbf{E}}_1$. The plasma velocity of each species, found from (16.31), is

$$\hat{\mathbf{u}}_{1,s} \;=\; i\frac{q_s}{m_s}\,\frac{\hat{\mathbf{E}}_1}{\omega} \quad . \tag{16.38}$$

This is nothing more than Newton's second law for the particles of species s, experiencing force $q_s\hat{\mathbf{E}}_1$. These transverse waves are therefore electromagnetic waves in which the electrons and ions are accelerated by the electric field.

16B.2 Longitudinal modes

The longitudinal modes include the scalar quantities, $\hat{n}_{1,i}$, $\hat{n}_{1,e}$, $\hat{p}_{1,i}$ and $\hat{p}_{1,e}$ and the components of $\hat{\mathbf{u}}_{1,s}$ parallel to the wave vector. Denoting that component $\hat{u}_s$, and using eq. (16.27) to eliminate $\hat{E}_1$ yields the momentum equation

$$\omega\,\hat{u}_s \;=\; \frac{k}{m_s n_0}\,\hat{p}_{1,s} \;+\; \frac{q_s}{e}\,\frac{\omega_{p,s}^2}{k}\,\frac{\hat{n}_{1,i}-\hat{n}_{1,e}}{n_0} \quad . \tag{16.39}$$

Combining this with (16.21) and (16.23) for each species leads to the 6×6 system

$$\omega \begin{bmatrix} \hat{n}_i \\ \hat{u}_i \\ \hat{p}_i \\ \hat{n}_e \\ \hat{u}_e \\ \hat{p}_e \end{bmatrix} = \begin{bmatrix} 0 & n_0 k & 0 & 0 & 0 & 0 \\ \omega_{p,i}^2/n_0 k & 0 & k/n_0 m_p & -\omega_{p,i}^2/n_0 k & 0 & 0 \\ 0 & \gamma p_{0,i} k & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & n_0 k & 0 \\ -\omega_{p,e}^2/n_0 k & 0 & 0 & \omega_{p,e}^2/n_0 k & 0 & k/n_0 m_e \\ 0 & 0 & 0 & 0 & \gamma p_{0,e} k & 0 \end{bmatrix} \cdot \begin{bmatrix} \hat{n}_i \\ \hat{u}_i \\ \hat{p}_i \\ \hat{n}_e \\ \hat{u}_e \\ \hat{p}_e \end{bmatrix} \quad . \tag{16.40}$$

Entropy modes

This system admits two zero-frequency modes whose right eigenvectors are

$$\underline{\hat{\mathrm{U}}}^{(9)} \;=\; \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad , \quad \underline{\hat{\mathrm{U}}}^{(10)} \;=\; \begin{bmatrix} 1 \\ 0 \\ -2m_p\omega_{p,i}^2/k^2 \\ -1 \\ 0 \\ 2m_e\omega_{p,e}^2/k^2 \end{bmatrix} \quad . \tag{16.41}$$

The first is reminiscent of the entropy mode: a change to density but not pressure. Here the density of both ions and electrons changes together thereby preventing a net charge from being built up. The second mode is slightly more subtle with

$$\hat{p}_{1,i} \;=\; -\,\hat{p}_{1,e} \;=\; -\,\frac{8\pi e^2 n_0}{k^2}\,\hat{n}_{1,i} \;=\; \frac{8\pi e^2 n_0}{k^2}\,\hat{n}_{1,e} \tag{16.42}$$

after noting that $m_s\omega_{p,s}^2 = 4\pi e^2 n_0$ for each species. Averaging the final two expression shows that pressure perturbations are related to a net charge

$$\hat{p}_{1,i} \;=\; -\,\hat{p}_{1,e} \;=\; -\,\frac{4\pi e^2 n_0^2}{k^2}\,\frac{\hat{n}_{1,i} - \hat{n}_{1,e}}{n_0} \quad . \tag{16.43}$$

Forming the difference

$$\frac{\hat{p}_{1,i}}{p_{0,i}} - \frac{\hat{p}_{1,e}}{p_{0,e}} \;=\; -\left(\frac{4\pi e^2 n_0}{k_{\mathrm{B}} T_{0,i}\, k^2} + \frac{4\pi e^2 n_0}{k_{\mathrm{B}} T_{0,e}\, k^2}\right)\frac{\hat{n}_{1,i} - \hat{n}_{1,e}}{n_0} \;=\; -\,\frac{\hat{n}_{1,i} - \hat{n}_{1,e}}{\lambda_d^2 k^2\, n_0} \quad , \tag{16.44}$$

after defining the *Debeye length* for species s

$$\lambda_{d,s} \;=\; \sqrt{\frac{k_{\mathrm{B}} T_{0,s}}{4\pi e^2 n_0}} \;=\; \frac{v_{th,s}}{\omega_{p,s}} \quad . \tag{16.45}$$

and then the combined Debeye length

$$\frac{1}{\lambda_d^2} \;=\; \frac{1}{\lambda_{d,i}^2} + \frac{1}{\lambda_{d,e}^2} \quad , \tag{16.46}$$

which is less than either of the individual Debeye lengths.

The nature of this zero-frequency mode is revealed most clearly when the relative net charge is expressed in terms of perturbation to either pressure or temperatures of each species,

$$\frac{\hat{n}_{1,i} - \hat{n}_{1,e}}{n_0} \;=\; -\,k^2\lambda_d^2\left(\frac{\hat{p}_{1,i}}{p_{0,i}} - \frac{\hat{p}_{1,e}}{p_{0,e}}\right) \;=\; -\,\frac{k^2\lambda_d^2}{1+k^2\lambda_d^2}\left(\frac{\hat{T}_{1,i}}{T_{0,i}} - \frac{\hat{T}_{1,e}}{T_{0,e}}\right) \quad . \tag{16.47}$$

When the perturbations are on scales shorter than the Debeye length, $k\lambda_d \gg 1$, there is relatively little change in the pressure of either species. The densities and temperatures change in opposite senses to produce some perturbations to the charge density. The situation is very different for perturbations on scales much longer than the Debeye length, $k\lambda_d \ll 1$. In that case there is *very little* net charge. Instead the pressures and temperatures change in the same sense. The pressure gradients are opposite for the two species, and they are balanced by the electric field. This electric field, produced by the small amount of charge separation which does occur, essentially allows the ions to exert a force on the electrons and *vice versa*.

The left eigenvectors of these modes are the right eigenvectors of the matrix's transpose. It can be checked that the following two vectors

$$\underline{\hat{\mathrm{W}}}^{(9)} \;=\; \frac{n_0}{\gamma + 2/\lambda_{d,i}^2 k^2}\begin{bmatrix} \gamma/n_0 \\ 0 \\ -1/p_{0,i} \\ 0 \\ 0 \end{bmatrix} + \frac{n_0}{\gamma + 2/\lambda_{d,e}^2 k^2}\begin{bmatrix} 0 \\ 0 \\ 0 \\ \gamma/n_0 \\ -1/p_{0,e} \end{bmatrix} \tag{16.48}$$

$$\underline{\hat{\mathrm{W}}}^{(10)} \;=\; \begin{bmatrix} \gamma/n_0 \\ 0 \\ -1/p_{0,i} \\ -\gamma/n_0 \\ 0 \\ 1/p_{0,e} \end{bmatrix} \quad , \tag{16.49}$$

are null left eigenvectors; with more effort it can be verified that each is orthogonal to the opposing right eigenvector. Evidently, $\underline{\hat{\mathrm{W}}}^{(9)}$, produces entropy perturbations to the different species, while $\underline{\hat{\mathrm{W}}}^{(10)}$ is the difference between their entropies. Any change to the total entropy will excite mode 9 which is, therefore, the natural inheritor of the term *entropy mode*. Mode number 10 will be excited by any exchange of entropy between the species.

Adiabatic perturbations

Based on the foregoing we will remove both zero-frequency modes from further consideration by restricting the dynamics to be adiabatic for each species. We do this by setting

$$\hat{p}_{1,s} \;=\; \gamma\frac{p_{0,s}}{n_0}\,\hat{n}_{1,s} \;=\; \gamma\,k_{\mathrm{B}}T_{0,s}\,\hat{n}_{1,s} \;=\; \gamma\,m_s\,v_{th,s}^2\,\hat{n}_{1,s} \quad , \tag{16.50}$$

and dropping the energy equation. Doing so converts (16.40) to the 4×4 system

$$\omega \begin{bmatrix} \hat{n}_i \\ \hat{u}_i \\ \hat{n}_e \\ \hat{u}_e \end{bmatrix} = \begin{bmatrix} 0 & n_0 k & 0 & 0 \\ (\omega_{p,i}^2 + \gamma v_{th,i}^2 k^2)/n_0 k & 0 & -\omega_{p,i}^2/n_0 k & 0 \\ 0 & 0 & 0 & n_0 k \\ -\omega_{p,e}^2/n_0 k & 0 & (\omega_{p,e}^2 + \gamma v_{th,e}^2 k^2)/n_0 k & 0 \end{bmatrix} \cdot \begin{bmatrix} \hat{n}_i \\ \hat{u}_i \\ \hat{n}_e \\ \hat{u}_e \end{bmatrix} \quad . \tag{16.51}$$

This system has no zero-frequency modes. The determinental equation is found using reduction by minors along the third row

$$-\omega \begin{vmatrix} -\omega & n_0 k & 0 \\ (\omega_{p,i}^2 + \gamma v_{th,i}^2 k^2)/n_0 k & -\omega & 0 \\ -\omega_{p,e}^2/n_0 k & 0 & -\omega \end{vmatrix} - n_0 k \begin{vmatrix} -\omega & n_0 k & 0 \\ (\omega_{p,i}^2 + \gamma v_{th,i}^2 k^2)/n_0 k & -\omega & -\omega_{p,i}^2/n_0 k \\ -\omega_{p,e}^2/n_0 k & 0 & (\omega_{p,e}^2 + \gamma v_{th,e}^2 k^2)/n_0 k \end{vmatrix}$$

$$\begin{aligned} &= \omega^4 - \omega^2(\omega_{p,i}^2 + \gamma v_{th,i}^2 k^2) - \omega^2(\omega_{p,e}^2 + \gamma v_{th,e}^2 k^2) \\ &\quad - \omega_{p,i}^2 \omega_{p,e}^2 \; + \; (\omega_{p,i}^2 + \gamma v_{th,i}^2 k^2)(\omega_{p,e}^2 + \gamma v_{th,e}^2 k^2) \\ &= \omega^4 - \omega^2(\omega_{p,i}^2 + \omega_{p,e}^2 + \gamma v_{th,i}^2 k^2 + \gamma v_{th,e}^2 k^2)\omega^2 + \gamma \omega_{p,i}^2 \omega_{p,e}^2 (\lambda_{d,i}^2 + \lambda_{d,e}^2) k^2 + \gamma^2 v_{th,e}^2 v_{th,i}^2 k^4 \\ &= \omega^4 - \omega^2 \omega_{p,e}^2 \left(1 + \gamma \lambda_{d,e}^2 k^2\right) + \omega_{p,e}^4 \frac{m_e}{m_p} \gamma \left(\lambda_{d,i}^2 k^2 + \lambda_{d,e}^2 k^2 + \gamma \lambda_{d,i}^2 \lambda_{d,e}^2 k^4\right) \; = \; 0 \quad , \end{aligned} \tag{16.52}$$

where we have dropped comparable terms with ratio $m_e/m_i \ll 1$. The discriminant of the final expression is

$$\begin{aligned} \sqrt{b^2 - 4ac} &= \omega_{p,e}^2 \left[(1 + \gamma \lambda_{d,e}^2 k^2)^2 - 4\gamma \frac{m_e}{m_p} \left(\lambda_{d,i}^2 k^2 + \lambda_{d,e}^2 k^2 + \gamma \lambda_{d,i}^2 \lambda_{d,e}^2 k^4 \right) \right]^{1/2} \\ &= \omega_{p,e}^2 (1 + \gamma \lambda_{d,e}^2 k^2) \left[1 \; - \; 4\gamma \frac{m_e}{m_p} \frac{\lambda_{d,i}^2 k^2 + \lambda_{d,e}^2 k^2 + \gamma \lambda_{d,i}^2 \lambda_{d,e}^2 k^4}{(1 + \gamma \lambda_{d,e}^2 k^2)^2} \right]^{1/2} \\ &\simeq \omega_{p,e}^2 (1 + \gamma \lambda_{d,e}^2 k^2) - 2\gamma \omega_{p,i}^2 \frac{\lambda_{d,i}^2 k^2 + \lambda_{d,e}^2 k^2 + \gamma \lambda_{d,i}^2 \lambda_{d,e}^2 k^4}{1 + \gamma \lambda_{d,e}^2 k^2} \end{aligned}$$

after the square root is expanded in $m_e/m_i \ll 1$. Recalling that

$$\gamma \omega_{p,i}^2 (\lambda_{d,i}^2 + \lambda_{d,e}^2) \; = \; \gamma \frac{k_B T_{0,i} + k_B T_{0,e}}{m_p} \; = \; \gamma \frac{p_{0,i} + p_{0,e}}{m_p n_0} \; \simeq \; \gamma \frac{p_0}{\rho_0} \; = \; c_s^2 \tag{16.53}$$

since the mass density $\rho_0 = (m_e + m_p) n_0 \simeq m_p n_0$ and the total pressure is the sum of the partial pressures of the two species. With this in place we can write the discriminant

$$\sqrt{b^2 - 4ac} \; = \; \omega_{p,e}^2 (1 + \gamma \lambda_{d,e}^2 k^2) - 2 \frac{c_s^2 k^2 + \gamma^2 v_{th,i}^2 \lambda_{d,e}^2 k^4}{1 + \gamma \lambda_{d,e}^2 k^2} \quad . \tag{16.54}$$

The eigenfrequencies for the longitudinal modes may be found from the roots of the quadratic, eq. (16.52),

$$\omega_{11,12}^2 \; = \; \omega_{p,e}^2 (1 + \gamma \lambda_{d,e}^2 k^2) \tag{16.55}$$

$$\omega_{13,14}^2 \; = \; \frac{c_s^2 k^2 + \gamma^2 v_{th,i}^2 \lambda_{d,e}^2 k^4}{1 + \gamma \lambda_{d,e}^2 k^2} \quad , \tag{16.56}$$

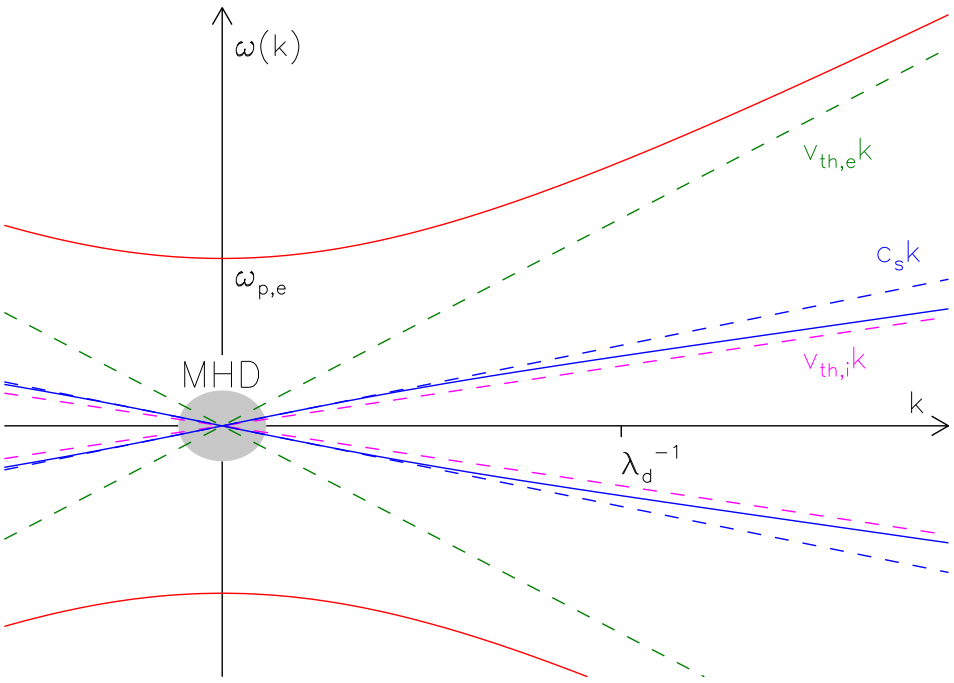

Figure 16.1: The dispersion relations, $\omega(k)$, given by eqs. (16.55) in red and (16.56) in blue. This is for the case $T_{0,i} = T_{0,e}$, and thus $\lambda_{d,i} = \lambda_{d,e} = \lambda_d$. The diagonal dashed lines are the asymptotic phase speeds $v_{th,e}$ (green), $v_{th,i}$ (megenta), and c_s (blue). A grey ellipse around the origin shows the region in which the MHD approximation is valid. (The ion-to-electron mass ratio is artificially reduced for clarity).

where the second term ($\sim m_e/m_p \ll 1$) in the discriminant (16.54), has been dropped from the first term, but retained in the second where the first is cancelled by the $-b$ of the quadratic formula. These are plotted in fig. 16.1.

As a consequence of the partial cancellation modes 11 and 12 (red in fig. 16.1) have much higher frequencies than modes 13 and 14 (blue). In the limit of short wavelengths, $\lambda_{d,e}k \gg 1$ the dispersion relations

$$\omega_{11,12} \;=\; \pm\sqrt{\gamma}\, v_{th,e}\, k \quad , \quad \omega_{13,14} \;=\; \pm\sqrt{\gamma}\, v_{th,i}\, k \quad , \tag{16.57}$$

are those for sound waves of the electron and ion fluids independently; these asymptotes are shown as green and magenta dashed lines in fig. 16.1 At these small scales the two species do not interact, and thus behave as independent fluids with independent normal modes.

At longer wavelengths, $\lambda_{d,e}k \ll 1$ the frequencies become

$$\omega_{11,12} \;=\; \pm\,\omega_{p,e} \quad , \quad \omega_{13,14} \;=\; \pm\, c_s\, k \quad . \tag{16.58}$$

Modes 13 and 14 are evidently classic acoustic waves with a restoring force derived from the combined pressures of the two species. Here again, the electric field is essential in communicating the electron pressure to the ions. This occurs on scales much larger than the electron Debeye length, $\lambda_{d,e}$, where quasi-neutrality obtains, swhile at shorter lengths the ions respond only to their own pressure. At these long wavelengths modes 11 and 12 are simple oscillations generically referred to as *plasma oscillations*. The full dispersion relation, eq. (16.55), shows that plasma waves are dispersive, at least for wavelengths approaching the Debeye length.

16B.3 Plasma as a Dielectric

The foregoing has treated the plasma as a fluid, following the method used throughout our study. It is also possible, though, to consider it as a dielectric medium in which a bound charge density ρ_b is induced by the electric field. The bound charge density of the plasma is

$$\rho_b \;=\; en_{1,i} - en_{1,e} \quad . \tag{16.59}$$

Because we have linearized the fluid equations the densities of each species, $n_{1,s}$, will be linear in the electric field. This means the bound charge will be proportional to electric field and the plasma will be a linear dielectric characterized by dielectric constant[2] ϵ according to

$$\nabla\cdot\mathbf{E} - \nabla\cdot\mathbf{D} = \nabla\cdot[(1-\epsilon)\mathbf{E}\,] \;=\; 4\pi\rho_b \;=\; 4\pi e(n_{1,i} - en_{1,e}) \quad , \tag{16.60}$$

where $\mathbf{D} = \epsilon\mathbf{E}$ is the electric displacement, whose divergence gives the *free charge* density. Using the Fourier transforms of the first order quantities, the dielectric constant of the plasma is

$$\epsilon(\omega,\mathbf{k}) \;=\; 1 \;-\; 4\pi e\,\frac{\hat{n}_{1,i} - \hat{n}_{1,e}}{i\mathbf{k}\cdot\hat{\mathbf{E}}_1} \quad . \tag{16.61}$$

(Since the dielectric constant is a function of wavenumber, and not really constant, the electric displacement is related to the electric field through convolution, rather the simple multiplication.)

Linearized fluid eqs. (16.21) and (16.22) can be combined to yield an equation for the density

$$\omega^2 m_s\,\hat{n}_{1,s} \;=\; k^2\hat{p}_{1,s} \;+\; q_s n_0\, i\mathbf{k}\cdot\hat{E}_1 \quad . \tag{16.62}$$

We henceforth assume that the plasma is polarized slowly enough, or on scales large enough, as to leave its zero-frequency longitudinal modes unexcited. This means we can assume adiabatic perturbation and use eq. (16.50) to eliminate pressure. The result shows explicitly how the density of each species is proportional to the electric field

$$\hat{n}_{1,s} \;=\; \frac{q_s n_0}{m_s}\,\frac{i\mathbf{k}\cdot\hat{E}_1}{\omega^2 - \gamma v_{th,s}^2 k^2} \quad . \tag{16.63}$$

[2]In MKS units, the variable ϵ denotes the permitivity, ϵ_0 is the permitivity of free space, and their ratio is the dielectric constant. In the cgs system $\epsilon_0 \to 1$ so ϵ plays both roles at once: it is the permitivity and the dielectric constant.

Placing the ion and electron version into eq. (16.61) yields the dielectric constant for a two-component, fully ionized plasma

$$\epsilon(\omega, \mathbf{k}) \;=\; 1 \;-\; \frac{\omega_{p,i}^2}{\omega^2 - \gamma v_{th,i}^2 k^2} \;-\; \frac{\omega_{p,e}^2}{\omega^2 - \gamma v_{th,e}^2 k^2} \quad . \tag{16.64}$$

The first use to which expression (16.64) may be put is an alternative derivation of the dispersion relationship for longitudinal waves. Recall that the electric displacement $\mathbf{D} = \epsilon \mathbf{E}$ is generated by *free charges*, rather than the bound charges composing the plasma. In the absence of free charge $\mathbf{D} = 0$, and if the situation is to be other than trivial (i.e. $\mathbf{E} = 0$) the electric field must take the form of a plane wave with frequency ω and wave vector $\mathbf{k}$ such that $\epsilon(\omega, \mathbf{k}) = 0$. Setting eq. (16.64) to zero leads to there relation

$$\begin{aligned} \omega^4 \;&-\; (\omega_{p,i}^2 + \omega_{p,e}^2 + \gamma v_{th,i}^2 k^2 + \gamma v_{th,e}^2 k^2)\omega^2 \\ &+\; \gamma(\omega_{p,i}^2 v_{th,e}^2 + \omega_{p,e}^2 v_{th,i}^2)k^2 \;+\; \gamma^2 v_{th,i}^2 v_{th,e}^2 k^4 \;=\; 0 \quad . \end{aligned} \tag{16.65}$$

Up to factors $\sim m_e/m_i \ll 1$ this matches eq. (16.52), showing that the dielectric treatment gives the same result as the fluid treatment.

Another application is to find the response of a plasma to the introduction of a point charge q_0. In that case the electric displacement must satisfy

$$\nabla \cdot \mathbf{D} \;=\; 4\pi\rho_f \;=\; 4\pi q_0\, \delta(\mathbf{x}) \quad . \tag{16.66}$$

The Fourier transform of this expression

$$i\mathbf{k} \cdot \hat{\mathbf{D}} \;=\; i\epsilon(\omega, \mathbf{k})\, \mathbf{k} \cdot \hat{\mathbf{E}} \;=\; 2q_0\, \delta(\omega) \quad , \tag{16.67}$$

since the Fourier transform of the spatial delta-function is a constant, and the Fourier transform of the temporally constant function yields a frequency delta-function. The electrostatic potential, defined by $\mathbf{E} = -\nabla\varphi$, therefore satisfies

$$\hat{\varphi}(\omega, \mathbf{k}) \;=\; \frac{2q_0}{k^2 \epsilon(0, \mathbf{k})}\, \delta(\omega) \;=\; \frac{2q_0}{k^2 + 1/\gamma\lambda_d^2}\, \delta(\omega) \quad , \tag{16.68}$$

after evaluating expression (16.64) at $\omega = 0$, and using eq. (16.46) to combine the individual Debeye lengths. The inverse Fourier transform yields the potential

$$\varphi(\mathbf{x}) \;=\; \frac{q_0}{2\pi^2} \int \frac{e^{i\mathbf{k}\cdot\mathbf{x}}}{k^2 + 1/\gamma\lambda_d^2}\, d^3k \;=\; \frac{q_0}{|\mathbf{x}|} \exp\left(-\frac{|\mathbf{x}|}{\sqrt{\gamma}\lambda_d}\right) \quad , \tag{16.69}$$

after performing the triple integral.

Expression (16.69) gives the electrostatic potential around a test charge in a plasma. The charge density of the combination

$$\rho_c(\mathbf{x}) \;=\; -\,\frac{\nabla^2\varphi}{4\pi} \;=\; q_0\, \delta(\mathbf{x}) \;-\; \frac{q_0}{4\pi\gamma\lambda_d^2 |\mathbf{x}|} \exp\left(-\frac{|\mathbf{x}|}{\sqrt{\gamma}\lambda_d}\right) \quad , \tag{16.70}$$

consists of the point charge, $q_0\delta(\mathbf{x})$, surrounded by a cloud of opposing bound charge — the *Debeye cloud*. It is evident from the exponential that the cloud extends only a few

Debeye lengths from the point charge. Beyond that there is no charge density, and the same exponential in eq. (16.69) means the potential, as well as the electric field vanish outside the Debeye cloud. The Debeye cloud has net charge equal and opposite to the point charge, which it therefore completely screens. It is expected that a conductor will completely screen any free charge, but it seems that in a plasma that screening only occurs on length scales significantly greater than a Debeye length.

In contrast to this, well *inside* a Debeye length the potential assumes the traditional Coulomb form, $\varphi = q_0/|\mathbf{x}|$, unaffected by the plasma. Only on such small scales do the particles composing the plasma become evident. The full potential given in eq. (16.69), switching between these limits, is sometimes called a *screened Coulomb potential*.

16C Linear waves in a magnetized two-fluid plasma

The introduction of a uniform magnetic field $\mathbf{B}_0$ does not compromise the homogeneity of the equilibrium, so plane waves will still be normal modes of the linearized system. It does, however, introduce a preferred direction and thus spoil isotropy. Waves in an anisotropic medium do not, in general, separate into longitudinal and transverse modes. The result is a very complicated situation, analogous to that of fast and slow magnetosonic waves when $\mathbf{k}$ and $\mathbf{B}_0$ are at an arbitrary angle. In the present case we could end up with a fully coupled 14×14 matrix for which we would seek 14 different eigenfrequencies.

The general theory of linear waves in a magnetized plasma is quite formidable. A common approach is to take the limit $p_{0,e}$, $p_{0,i} \to 0$, to obtain so-called *cold plasma* waves. This eliminates two equations leaving 12 whose eigenmodes form 6 pairs of cold plasma waves. This zoo of waves includes things called whistler waves, ion-cyclotron waves, O-modes and X-modes, among others; it is quite formidable by itself.

One way to avoid all this complication is to restrict consideration to the case where $\mathbf{k}$ and $\mathbf{B}_0$ are parallel. In this special case there is again a decoupling into longitudinal and transverse modes. In the case of MHD the fast mode and shear Alfvén mode become two degenerate polarizations of a transverse mode, while the slow mode becomes purely longitudinal.

In order to retain the simplicity of transverse/longitudinal decoupling we consider only the case of parallel wave propagation and take $\mathbf{B}_0 = B_0\hat{\mathbf{z}}$ and $\mathbf{k} = k\hat{\mathbf{z}}$. The introduction of an equilibrium magnetic field will affect only the linearized momentum equation, eq. (16.22)

$$\omega\, m_s n_0 \hat{\mathbf{u}}_{1,s} \;=\; k\hat{\mathbf{z}}\,\hat{p}_{1,s} \;+\; i\, q_s n_0 \hat{\mathbf{E}}_1 \;+\; i\,\frac{q_s n_0 B_0}{c}\,\hat{\mathbf{u}}_{1,s}\times\hat{\mathbf{z}} \quad . \tag{16.71}$$

The new term does not affect the $\hat{\mathbf{z}}$ component of the momentum equation so the longitudinal modes, discussed in § 16B.2, are completely unaffected by a magnetic field.

We henceforth consider only the transverse modes, consisting of vectors with $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ components. The cross product of the Lorentz force will, however, couple these two components. In order to decouple the system we write all vectors in *circularly polarized* components

$$\hat{\mathbf{E}}_1 \;=\; \hat{E}_+\,(\hat{\mathbf{x}} + i\hat{\mathbf{y}}) \;+\; \hat{E}_-\,(\hat{\mathbf{x}} - i\hat{\mathbf{y}}) \quad , \tag{16.72}$$

and likewise for $\hat{\mathbf{u}}_{1,i}$, $\hat{\mathbf{u}}_{1,e}$ and $\hat{\mathbf{B}}_1$. In order that complex plane waves can be combined into real solutions we have always imposed a relation on amplitudes with negative frequencies

and wave numbers: $\hat{\mathbf{E}}_1(-\omega,-k) = \hat{\mathbf{E}}_1^*(\omega,k)$. Applying this to eq. (16.72) shows that

$$\hat{E}_-(-\omega,-k) \;=\; \hat{E}_+^*(\omega,k) \quad . \tag{16.73}$$

The $\hat{E}_+$ component with positive frequency and wave number, when combined with its corresponding $\hat{E}_-(-\omega,-k)$, yields an electric field

$$\mathrm{Re}\Big[\,\hat{E}_+(\hat{\mathbf{x}}+i\hat{\mathbf{y}})e^{ikz-i\omega t}\,\Big] \;=\; \hat{E}_+\Big[\,\cos(kz-\omega t)\hat{\mathbf{x}} \;-\; \sin(kz-\omega t)\hat{\mathbf{y}}\,\Big] \quad , \tag{16.74}$$

after taking $\hat{E}_+$ to be real. At any given time the electric field traces out a left-handed helix, so this is termed the *left circularly* polarized component. We will henceforth refer to the coefficients of $(\hat{\mathbf{x}}+i\hat{\mathbf{y}})$ as left-hand polarization, even though they imply a contribution with opposing sign of frequency and wave number proportional to $(\hat{\mathbf{x}}-i\hat{\mathbf{y}})$.

The advantage of using circular polarization is that cross products do not mix the components:[3]

$$(\hat{\mathbf{x}}\pm i\hat{\mathbf{y}})\times\hat{\mathbf{z}} \;=\; -\,\hat{\mathbf{y}}\pm i\hat{\mathbf{x}} \;=\; \pm\, i(\hat{\mathbf{x}}\pm i\hat{\mathbf{y}}) \quad . \tag{16.75}$$

With this in hand eq. (16.71) can be expressed as separate equations for each circular component

$$\omega\,\hat{u}_{\pm,s} \;=\; i\,\frac{q_s}{m_s}\hat{E}_\pm \;\mp\; \frac{q_sB_0}{m_sc}\,\hat{u}_{\pm,s} \quad . \tag{16.76}$$

Faraday's law and Ampere's law, eqs. (16.24) and (16.25), similarly become

$$\omega\hat{B}_\pm \;=\; \mp i\,ck\hat{E}_\pm \tag{16.77}$$

$$\omega\hat{E}_\pm \;=\; \pm i\,ck\hat{B}_\pm \;-\; 4\pi i\,en_0(\hat{u}_{\pm,i}-\hat{u}_{\pm,e}) \quad . \tag{16.78}$$

The circular polarizations therefore decouple from one another leading to two separate 4×4 systems in place of eq. (16.28)

$$\omega\begin{bmatrix}\hat{u}_{i,\pm}\\ \hat{u}_{e,\pm}\\ \hat{B}_\pm\\ \hat{E}_\pm\end{bmatrix} \;=\; \begin{bmatrix}\mp\Omega_i & 0 & 0 & i\,e/m_p\\ 0 & \pm\Omega_e & 0 & -i\,e/m_e\\ 0 & 0 & 0 & \mp i\,ck\\ -4\pi ien_0 & 4\pi ien_0 & \pm i\,ck & 0\end{bmatrix}\cdot\begin{bmatrix}\hat{u}_{i,\pm}\\ \hat{u}_{e,\pm}\\ \hat{B}_\pm\\ \hat{E}_\pm\end{bmatrix} \quad , \tag{16.79}$$

where we have introduced the gyrofrequency for species s, $\Omega_s = eB_0/m_sc$, defined to be positive.[4] The upper and lower sign in this equation gives the equation for left and right circularly polarized components respectively.

This slightly more complicated matrix has no zero-frequency eigenvectors, so the shear modes, previously with $\omega = 0$, evidently have gained a non-vanishing frequency. Without the simplification offered by zero-frequency roots it is necessary to take the determinant by co-factors, as we have done several times before. Doing so leads to the determinental equation

$$(\omega^2-c^2k^2)(\omega\pm\Omega_i)(\omega\mp\Omega_e) \;-\; \omega_p^2\,\omega^2 \;=\; 0 \quad , \tag{16.80}$$

[3]The deeper issue is that rotation about the z-axis leaves the basis vector $\hat{\mathbf{x}}+i\hat{\mathbf{y}}$ unchanged, except for a complex factor. The same cannot be said of either $\hat{\mathbf{x}}$ or $\hat{\mathbf{y}}$.

[4]It is common to use, instead, a defintion including the sign of the charge, making the electron gyrofreqency a *negative* quantity. We opt against this convention.

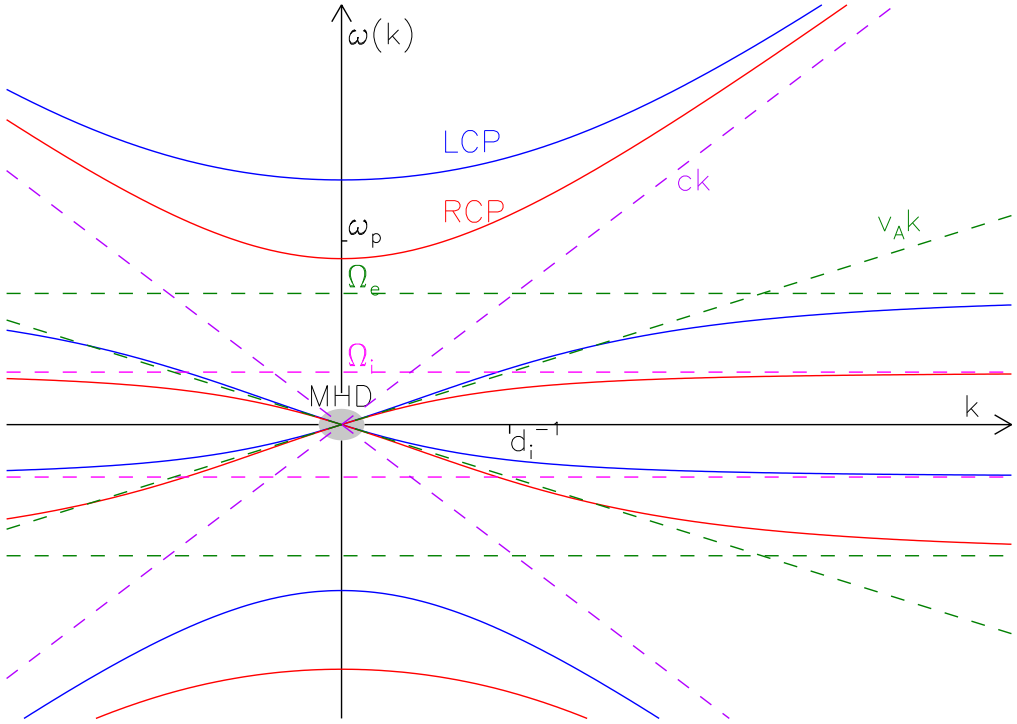

Figure 16.2: The dispersion relations, $\omega(k)$, found by solving eq. (16.80). The roots of the upper sign (LCP) are shown in blue, and of the lower sign (RCP) in red. Horizontal dashed lines show the resonance at $\pm\Omega_i$ (magenta) and $\pm\Omega_e$ (green). The diagonal dashed lines are the asymptotic phase speeds c (violet) and the modified Alfvén speed (green), given by eq. (16.84) but labelled v_A. A grey ellipse around the origin shows the region in which the MHD approximation is valid.

where ω_p is the plasma frequency defined by eq. (16.34). Equation (16.80) is a quartic equation for which we can expect two or four distinct real solutions. Figure 16.2 shows that all four real roots exist for both the upper sign (blue) and lower sign (red). In many previous cases we found polynomials in ω^2 so there were both positive and negative eigenfrequencies. These positive/negative pairs were necessary to permit the formation of *real* solutions in the end. The present case is more complicated, and neither the upper nor the lower version of eq. (16.80) has roots in positive/negative pairs. This is evident from the lack of vertical symmetry in a given color of fig. 16.2. Satisfying condition (16.73), however, requires that a positive root for the upper sign be matched with a negative root for the lower sign. It can bee seen that eq. (16.80) does satisfy this condition, and the vertical reflection of the red curves in fig. 16.2 gives the blue curves.

Since eq. (16.80) is a quartic equation, it is very difficult to factor analytically. It is,

however, linear in k^2, and thus much easier to solve that way:

$$c^2k^2 \;=\; \omega^2\,\frac{\omega^2 \mp \Omega_e\,\omega - (\Omega_e\Omega_i + \omega_p^2)}{(\omega \pm \Omega_i)(\omega \mp \Omega_e)} \quad , \tag{16.81}$$

after dropping one ω term in the numerator because $\Omega_e \gg \Omega_i$. It is evident from this that for very high frequencies, $|\omega| \gg \max(\Omega_e, \omega_p)$, the dispersion relation becomes that of a light wave in vacuum: $\omega^2 = c^2k^2$, for either polarization (the violet dashed lines in fig. 16.2). Using this in eq. (16.79), and taking $\omega/k = +c$ for concreteness, shows that the eigenmode will have $\hat{B}_\pm = i\hat{E}_\pm$, for either polarization; this is a circularly polarized light wave.

Both upper branches in fig.16.2 have a cut-off frequency where they cross $k = 0$ — i.e. the roots of the numerator of eq. (16.81). In the unmagnetized case this cut-off was at $\omega = \omega_p$ independent of polarization; in the magnetized case the right polarized branch has a lower cut-off than the left polarized branch.

Even as the two upper branches approach the $\omega = ck$ asymptote they remain separate. This means that left and right polarization travel at different speeds through a magnetized plasma. The phase speed for high frequencies

$$\frac{\omega}{k} \;\simeq\; c\left[1 + \frac{\omega_p^2}{2\omega^2}\left(1 \pm \frac{\Omega_e}{\omega}\right) + \cdots\right] \quad , \tag{16.82}$$

is slightly above c (as it is in eq. [16.36], for an unmagnetized plasma) and slightly higher for left than for right polarizations. This small difference in phase speeds, $\sim \omega_p^2\Omega_e/\omega^3$, will manifest as a gradual change in the plane of linear polarization known as *Faraday rotation*. The change in polarization direction over frequency will depend on the product of plasma density and magnetic field strength. This provides a means of measuring those properties in an astrophysical plasma.

A more interesting limit of eq. (16.81) is that for very low frequencies, $|\omega| \ll \min(\Omega_i, \omega_p)$, where the relation becomes

$$c^2k^2 \;\simeq\; \omega^2\left(1 + \frac{\omega_p^2}{\Omega_e\Omega_i}\right) \;\simeq\; \omega^2\left(1 + \frac{c^2}{v_{\rm A}^2}\right) \quad , \tag{16.83}$$

where $v_{\rm A} = B_0/\sqrt{4\pi n_0 m_i}$ is the standard Alfvén speed. This is a dispersionless wave propagating at

$$\left|\frac{\omega}{k}\right| \;=\; \frac{v_{\rm A}c}{\sqrt{v_{\rm A}^2 + c^2}} \;<\; c \;, \tag{16.84}$$

shown by the green dashed diagonal line in fig.16.2. In the case that $v_{\rm A} \ll c$ the phase speed becomes the the Alfvén speed, and we have a parallel-propagating shear Alfvén wave for both polarizations. These are, in fact, the degenerate fast-mode/shear Alfvén mode we found in MHD. This suggests that MHD is a reasonable approximation of a two-component plasma, at least of its linearized dynamics, for low frequencies and provided $v_{\rm A} \ll c$. In the case where $v_{\rm A} > c$ eq. (16.84) shows that the plasma wave still travels sub-luminally.

One more notable feature of eq. (16.81) is its behavior as the frequency approaches either of the gyrofrequencies. For the case of the upper sign (left-hand polarization) as $\omega \to \Omega_e$ the wave number diverges, $k \to \infty$. This appears as a green dashed horizontal

asymptote in fig.16.2. Because of the diverging wave number eq. (16.79) demands that the eigenmode have $\hat{E}_+ \simeq \hat{B}_+ \simeq 0$. It also must have $\hat{u}_{i,+} \simeq 0$ and therefore consists of simple motion of the electrons alone. The analog of eq. (16.74) shows that at any given point the electron velocity, $\hat{\mathbf{u}}_{1,e}$, rotates in the counter-clockwise sense. This the classic gyromotion of an electron in a static, uniform magnetic field ($B_0\hat{\mathbf{z}}$) subject to no electric field: counter-clockwise at frequency Ω_e. The $k \to \infty$ behavior is therefore termed a resonance. (The divergence at $\omega \to -\Omega_e$ of the lower sign is needed according to [16.73] to produce a real left-hand polarized wave.)

The same reasoning shows that the right-hand polarized wave (the lower sign) resonates (i.e. $k \to \infty$) at $\omega = \Omega_i$. The normal mode there corresponds to simple gyro motion of ions alone ($\hat{E}_- = \hat{B}_- = \hat{u}_{e,-} = 0$), in the clockwise sense. The separation of the polarizations at frequencies near or above Ω_i (see fig. 16.2) is at odds with the dispersionless, polarization-degenerate waves found in MHD. It is therefore necessary to restrict attention to $\omega \ll \Omega_i$ to obtain the MHD limit, at least its linear dynamics (the grey dashed region in fig. 16.2) . The means considering length scales much larger than $v_{\rm A}/\Omega_i = c/\omega_{p,i} = d_i$, called the ion collisions skin depth.

Chapter 17

Two Fluid Derivation of MHD

Magnetohydrodynamics, which occupies much of these notes, is a description of a plasma as a single fluid. This combines the electron fluid and the ion fluid into a single fluid, characterized by a single mass density, $\rho(\mathbf{x}, t)$, a single fluid velocity, $\mathbf{u}(\mathbf{x}, t)$, and single pressure, $p(\mathbf{x}, t)$. The linearized two-fluid treatment from the foregoing chapter provides some insight into this synthesis. A notable finding was that the only linear wave in which the local densities of electrons and ions differed significantly, thereby creating a charge density $e(n_i - n_e)$, was the plasma oscillation described by eq. (16.55). This wave oscillated at high frequency, $\omega_{11,12} > \omega_{p,e}$, so the charge density will not persist very long. Based on this insight we assume that when considering dynamics on longer time scales, $\tau \gg 2\pi/\omega_{p,e}$, there will be no charge *no charge* density. This restriction, equivalent to taking $n_i(\mathbf{x}) = n_e(\mathbf{x})$ at all points, is called *quasi-neutrality*. We had made this claim long ago, motivated by our belief that a good conductor will not admit any charge density inside it. We now see that to be a good conductor requires a tiny bit of patience: one must wait for the charges to neutralize themselves.

This gives us our first sense that the combined system of electrons and ion might behave as a single fluid, rather than two distinct fluids, at least on slow enough time scales. This is the basis of the single-fluid system we call MHD. The electron fluid responds to the motion of the ions and does not, therefore, behave as a separate degree of freedom.

The linearization of the previous chapter demonstrated more qualitatively how certain degrees of freedom are only manifest at higher frequencies and shorter length scales (higher wave numbers k). The electron fluid, ion fluid, electric, and magnetic fields account for a total of 14 degrees of freedom.[1] Linearization about a homogeneous equilibrium produces 6 waves, with positive and negative frequencies, and 2 zero-frequency modes. In the special case we consider, where $\mathbf{k}$ is parallel to $\mathbf{B}_0$, the 6 waves can be separated into 2 longitudinal waves, whose dispersion relation is plotted in fig. 16.1, and 4 transverse waves, plotted in fig. 16.2. It is notable that only one of the former and two of the latter, pass into the gray shaded region of low frequency and low wave number ($|\omega| \ll \Omega_i$ and $k \ll 1/d_i = \omega_{p,i}/c$). These correspond to the slow magnetosonic (longitudinal) and the fast magnetosonc and shear Alfvén waves (transverse) of traditional MHD. The other longitudinal wave, for which

[1] That is to say we ended up with a 14×14 linear system, namely eqs. (16.21) – (16.25), from which we obtained 14 eigenfrequencies for each wave vector $\mathbf{k}$. We appealed to a number of tricks to reduce this to more manageable sizes.

$|\omega| > \omega_{p,e}$, is the plasma oscillation described above. The other transverse waves, both with $|\omega| > \Omega_e$, are related to light waves. All of these represent degrees of freedom which manifest on very short times scales, and are therefore excluded from the single-fluid treatment.

Linearization thereby points the way to the single-fluid treatment of the plasma we know as MHD. The full set of MHD equations are, however, actually *non-linear*. We must return then to the full set of nonlinear equations for both electrons and ions, and somehow combine these into a set of non-linear equations for a single fluid: the MHD equations. In doing so we will appeal to the lessons from the linearization. We will assume time scales much longer that $2\pi/\omega_{p,e}$, in order invoke the approximation of *quasi-neutrality*: $n_e(\mathbf{x},t) = n_i(\mathbf{x},t)$.

17A MHD densities

The mass density of the combined MHD fluid is simply a sum of othe masses of its two components

$$\rho(\mathbf{x},t) \;=\; m_i n_i(\mathbf{x},t) + m_e n_e(\mathbf{x},t) \;\simeq\; (m_i + m_e)\, n_e(\mathbf{x},t) \;\simeq\; m_i n_e(\mathbf{x},t) \quad , \tag{17.1}$$

after using quasi-neutrality, $n_i = n_e$, and noting that $m_e \ll m_i$. The momentum density is

$$\rho\mathbf{u} \;=\; m_i n_i \mathbf{u}_i + m_e n_e \mathbf{u}_e \;=\; m_i n_e \left(\mathbf{u}_i + \frac{m_e}{m_i}\mathbf{u}_e\right) \quad , \tag{17.2}$$

so the fluid velocity

$$\mathbf{u}(\mathbf{x},t) \;=\; \mathbf{u}_i(\mathbf{x},t) + \frac{m_e}{m_i}\mathbf{u}_e(\mathbf{x},t) \quad , \tag{17.3}$$

is the velocity of the ion fluid up to corrections of order m_e/m_i.

The current density is determined by the *difference* between velocities of the two fluids

$$\mathbf{J} \;=\; e(n_i\mathbf{u}_i - n_e\mathbf{u}_e) \;=\; en_e(\mathbf{u}_i - \mathbf{u}_e) \;=\; en_e\mathbf{v}_d \quad , \tag{17.4}$$

where $\mathbf{v}_d = \mathbf{J}/en_e = \mathbf{u}_i - \mathbf{u}_e$ is the *drift velocity*. It is this velocity with which one associates the current density and *not* the velocity of the plasma itself. Equations (17.3) and (17.4) can be inverted to express electron and ion velocities in terms of MHD quantities

$$\mathbf{u}_i \;\simeq\; \mathbf{u} \;+\; \frac{m_e}{m_i}\mathbf{v}_d \quad , \tag{17.5}$$

$$\mathbf{u}_e \;\simeq\; \mathbf{u} \;-\; \mathbf{v}_d \tag{17.6}$$

after dropping comparable terms $\sim m_e/m_i$.

Since current in MHD satisfies Ampère's law without displacement current the drift velocity is

$$\mathbf{v}_d \;=\; \frac{1}{en_e}\mathbf{J} \;=\; \frac{c}{4\pi en_e}\nabla\times\mathbf{B} \quad . \tag{17.7}$$

Its magnitude is therefore

$$|\mathbf{v}_d| \;\sim\; \frac{c}{4\pi en_e}\frac{B}{L} \;=\; \frac{c}{\sqrt{4\pi e^2 n_e/m_i}\sqrt{4\pi m_i n_e}}\frac{B}{L} \;=\; \frac{c/\omega_{p,i}}{L}\, v_{\rm A} \quad , \tag{17.8}$$

when magnetic variations occur on length scale L. This is compared to a length $d_i = c/\omega_{p,i}$ called the *ion collisionless skin depth*. It is a scale intrinsic to the plasma and for densities characteristic of the solar corona, $n_e \sim 10^9\,\mathrm{cm}^{-3}$, it is only 7 meters! This means that magnetic field variations on typical coronal length scales, $L \sim 10^9$ cm, lead to drift velocities many orders of magnitude smaller than the Alfvén speed. When Lorentz forces are important they often accelerate fluids to speeds comparable to the Alfvén speed. In these cases the drift speed is very small compared to the typical flow speed. The electrons and ions mostly move together, with only a very small difference needed to sustain the electric current.

17B MHD momentum equation

Since quasi-neutrality demands the densities of the two species remain everywhere equal, the continuity equation of each species is therefore related. These can naturally be combined into a single mass continuity equation for the plasma,

$$\frac{\partial \rho}{\partial t} \;=\; -\nabla\cdot\left[\, n_e(m_i\mathbf{u}_i + m_e\mathbf{u}_e)\,\right] \;=\; -\nabla\cdot(\rho\mathbf{u}) \quad , \tag{17.9}$$

which is the traditional mass continuity equation — how could it be otherwise? The two momentum equations, however, are not so obviously redundant

$$\begin{aligned} m_i\frac{\partial(n_e\mathbf{u}_i)}{\partial t} \;&=\; -m_i\nabla\cdot(n_e\mathbf{u}_i\mathbf{u}_i) \;-\; \nabla\cdot\underline{\underline{\mathrm{P}}}_i \\ &\quad +\; en_e\mathbf{E} \;+\; en_e\frac{\mathbf{u}_i}{c}\times\mathbf{B} \;+\; \frac{m_e n_e}{\tau_{ei}}(\mathbf{u}_e-\mathbf{u}_e) \quad , \end{aligned} \tag{17.10}$$

$$\begin{aligned} m_e\frac{\partial(n_e\mathbf{u}_e)}{\partial t} \;&=\; -m_e\nabla\cdot(n_e\mathbf{u}_e\mathbf{u}_e) \;-\; \nabla\cdot\underline{\underline{\mathrm{P}}}_e \\ &\quad -\; en_e\mathbf{E} \;-\; en_e\frac{\mathbf{u}_e}{c}\times\mathbf{B} \;+\; \frac{m_e n_e}{\tau_{ei}}(\mathbf{u}_i-\mathbf{u}_e) \quad , \end{aligned} \tag{17.11}$$

where the drag terms have been written so as to obviously satisfy Newton's third law. In a simple sum of the equations the electric force vanishes due to quasi-neutrality and the drag terms cancel each other. The leaves

$$\frac{\partial(\rho\mathbf{u})}{\partial t} \;=\; -\nabla\cdot(n_e m_i\mathbf{u}_i\mathbf{u}_i + n_e m_e\mathbf{u}_e\mathbf{u}_e) - \nabla\cdot(\underline{\underline{\mathrm{P}}}_i + \underline{\underline{\mathrm{P}}}_e) \;+\; \frac{1}{c}\mathbf{J}\times\mathbf{B} \quad , \tag{17.12}$$

after using eq. (17.2). The term inside the first divergence on the right may be expressed

$$n_e m_i\mathbf{u}_i\mathbf{u}_i + n_e m_e\mathbf{u}_e\mathbf{u}_e \;\simeq\; \rho\left(\mathbf{u}_i\mathbf{u}_i + \frac{m_e}{m_i}\mathbf{u}_e\mathbf{u}_e\right) \;=\; \rho\left(\mathbf{u}\mathbf{u} + \frac{m_e}{m_i}\mathbf{v}_d\mathbf{v}_d\right) \quad . \tag{17.13}$$

Since $m_e/m_i \ll 1$ and $|\mathbf{v}_d|$ is typically smaller, often much smaller, that $|\mathbf{u}|$, we can neglect the second term. This leaves the standard MHD momentum equation

$$\frac{\partial(\rho\mathbf{u})}{\partial t} + \nabla\cdot(\rho\mathbf{u}\mathbf{u}) \;=\; -\nabla\cdot(\underline{\underline{\mathrm{P}}}_i + \underline{\underline{\mathrm{P}}}_e) \;+\; \frac{1}{c}\mathbf{J}\times\mathbf{B} \quad , \tag{17.14}$$

with a pressure tensor formed from a sum of the the two species.

While it seems natural that the electron pressure tensor, $\underline{\underline{\mathrm{P}}}_e$, would appear in the single-fluid moment equation, eq. (17.14), it is worth pausing to understand why it does. It was remarked upon above, and shown explicitly in eq. (17.5), that the fluid velocity is approximately the same as the ion velocity, $\mathbf{u} \simeq \mathbf{u}_i$. This follows from the fact that the ions account for nearly all the momentum of the fluid, and momentum is the basis of the fluid velocity. The random thermal motions of the ions give rise to the ion pressure, $\underline{\underline{\mathrm{P}}}_i$, and therefore naturally affect the mean ion momentum. But how do the random thermal motions of the electrons affect that momentum? The electric force on the electrons was present in the electron momentum equation, eq. (17.11), but was cancelled by the equal and opposite force on the ions when the two equations were combined. If this term had been large by itself it would then it would need to be balanced by the electron pressure in eq. (17.11), since all other terms are multiplied by the very small electron mass. It therefore seems that the electron pressure gradient is equal to the electric field, and the ions *do feel* that force directly. It is in this seemingly round-about way that the electron pressure enters the single-fluid momentum equation.

17C MHD energy equation

The pressure tensor entering eq. (17.14) is a combination of the electron pressure tensor and the ion pressure tensor. These can each be decomposed into isotropic and traceless contributions

$$\underline{\underline{\mathrm{P}}}_i + \underline{\underline{\mathrm{P}}}_e \;=\; (p_i + p_e)\,\underline{\underline{\mathrm{I}}} \;+\; \underline{\underline{\pi}}_i + \underline{\underline{\pi}}_e \;=\; p\,\underline{\underline{\mathrm{I}}} \;+\; \underline{\underline{\pi}} \quad . \tag{17.15}$$

The isotropic contribution to the sum is the scalar total pressure $p = p_i + p_e$ combining the contributions of the electron pressure and the ion pressure. Each of these will be 2/3 of the thermal energy density of the species. The total pressure is, as we expected, 2/3 of the combined thermal energy. The total pressure, p, will therefore be determined by the evolution of the thermal energies of the two species summed together. This is a fairly straightforward task, and we will not pursue it further here. We should note, however, that energy transferred between the species will not appear in the MHD energy equation.

We have noted above that the pressure of a particular species is computed from the second moment of the distribution function of that species. There is a pressure regardless of what that distribution function happens to be. In the special case that the distribution is a *Maxwellian*, however, we can define a temperature T_s for the species — and *only* in this special case. When both electrons and ions have Maxwellian distributions we can write the total pressure

$$p \;=\; p_i + p_e \;=\; k_{\mathrm{B}} n_i T_i \;+\; k_{\mathrm{B}} n_e T_e \;=\; k_{\mathrm{B}}\, n_e\, (T_i + T_e) \quad , \tag{17.16}$$

after making use once again of quasi-neutrality. The MHD energy equation can then be constructed by combing equations for the evolutions of T_i and T_e independently. It is not necessary to assume these temperatures are actually equal, although that assumption is sometimes made. Indeed, collisions between the species will exchange energy with the aim of achieving thermodynamic equilibrium: $T_i = T_e$. Even if this state is not achieved, however, there will be an MHD energy equation describing how the combined thermal energy density evolves.

When there is a gradient in the temperature of either species, then its distribution cannot remain Maxwellian by itself. We have seen, however, that if collisions are frequent enough, the distribution will be nearly Maxwellian, and the small departure will result in conductive heat flux down the temperature gradient

$$\mathbf{q}_s \;=\; -\,\kappa_s\,\nabla T_s \quad , \tag{17.17}$$

where κ_s is the conductivity of species s, as given in eq. (15.75). The net energy flux will be a sum of the fluxes in each of the two species,

$$\mathbf{q} \;=\; -\,\kappa_i\,\nabla T_i \;-\; \kappa_e\,\nabla T_e \quad . \tag{17.18}$$

Only in the case that the two temperatures are equal, $T_e = T_i = T$, can we write the heat flux using the gradient of a single temperature, $\mathbf{q} = -\kappa\nabla T$. In this case the conductivity is a sum of the two contributions

$$\kappa \;=\; \kappa_i \;+\; \kappa_e \;\simeq\; \kappa_e \;=\; \frac{5}{8\pi}\,\frac{k_{\rm B}^{7/2}}{e^4\sqrt{m_e}\ln\Lambda}\,T^{5/2} \quad , \tag{17.19}$$

but is dominated by the electron contribution since $m_e \ll m_i$, appear in the denominators of eq. (15.75).

17D Generalized Ohm's law

Since the drift velocity, $\mathbf{v}_d$, is the difference between species velocities it is natural to look for an equation for its evolution in the difference between the two momentum equations. Multiplying eq. (17.10) by m_e/m_i, and subtracting from it (17.11) yields

$$\begin{aligned} m_e\frac{\partial(n_e\mathbf{v}_d)}{\partial t} \;&=\; -m_e\nabla\cdot(n_e\mathbf{u}_i\mathbf{u}_i - n_e\mathbf{u}_e\mathbf{u}_e) \;-\; \nabla\cdot\left(\frac{m_e}{m_i}\underline{\underline{\mathrm{P}}}_i - \underline{\underline{\mathrm{P}}}_e\right) + \; en_e\mathbf{E} \\ &\quad +\; \frac{en_e}{c}\left(\frac{m_e}{m_i}\mathbf{u}_i + \mathbf{u}_e\right)\times\mathbf{B} \;+\; \frac{m_e n_e}{\tau_{ei}}(\mathbf{u}_e - \mathbf{u}_i) \quad . \end{aligned}$$

The first term on the left may be expressed

$$n_e\mathbf{u}_i\mathbf{u}_i - n_e\mathbf{u}_e\mathbf{u}_e \;\simeq\; n_e(\mathbf{u}\mathbf{v}_d + \mathbf{v}_d\mathbf{u} - \mathbf{v}_d\mathbf{v}_d) \;=\; \frac{1}{e}(\mathbf{u}\mathbf{J} + \mathbf{J}\mathbf{u}) \quad . \tag{17.20}$$

to leading order in m_e/m_i. The Lorentz force term becomes

$$\left(\frac{m_e}{m_i}\mathbf{u}_i + \mathbf{u}_e\right)\times\mathbf{B} \;\simeq\; \mathbf{u}\times\mathbf{B} \;-\; \mathbf{v}_d\times\mathbf{B} \;=\; \mathbf{u}\times\mathbf{B} \;-\; \frac{1}{en_e}\mathbf{J}\times\mathbf{B} \quad , \tag{17.21}$$

to leading order in m_e/m_i. Introducing these expressions, and dropping $(m_e/m_i)\underline{\underline{\mathrm{P}}}_i$ in the face of $\underline{\underline{\mathrm{P}}}_e$, yields an equation governing the time evolution of the current density

$$\begin{aligned} \frac{m_e}{e}\left[\frac{\partial\mathbf{J}}{\partial t} + \nabla\cdot(\mathbf{u}\mathbf{J} + \mathbf{J}\mathbf{u})\right] \;&=\; \nabla\cdot\underline{\underline{\mathrm{P}}}_e \\ +\, en_e\underbrace{\left(\mathbf{E} + \frac{1}{c}\mathbf{u}\times\mathbf{B}\right)}_{\mathbf{E}'} \;&-\; \frac{1}{c}\mathbf{J}\times\mathbf{B} \;-\; \frac{m_e}{e\tau_{ei}}\mathbf{J} \quad . \end{aligned} \tag{17.22}$$

The evolution of $\mathbf{v}_d$, and thus $\mathbf{J}$, appears to be governed by a number of forces including that of the electric field in the reference frame of the fluid, $\mathbf{E}'$.

The relation between electric field and current, or current change, is a kind of Ohm's law; it is typically called the *generalized Ohm's law*. It is more common to express Ohm's law with electric field equal to other terms. For further clarity we follow standard practice and decompose the electron pressure tensor into an isotropic scalar piece and a viscous stress contribution: $\underline{\underline{\mathrm{P}}}_e = p_e \underline{\underline{\mathrm{I}}} + \underline{\underline{\pi}}_e$. Rearranging the resulting equation and dividing by en_e the entire relation yields a full form of generalized Ohm's law where the electric field

$$\mathbf{E}' = \mathbf{E} + \frac{1}{c}\mathbf{u}\times\mathbf{B} = \overbrace{\frac{4\pi}{\omega_{p,e}^2\tau_{ei}}\mathbf{J}}^{\text{Ohmic}} + \overbrace{\frac{4\pi}{\omega_{p,e}^2}\left[\frac{\partial \mathbf{J}}{\partial t} + \nabla\cdot(\mathbf{u}\mathbf{J}+\mathbf{J}\mathbf{u})\right]}^{e\text{ inertia}} - \overbrace{\frac{1}{en_e}\nabla p_e}^{\text{polarization}} - \underbrace{\frac{1}{en_e}\nabla\cdot\underline{\underline{\pi}}_e}_{e\text{ viscosity}} + \underbrace{\frac{1}{en_ec}\mathbf{J}\times\mathbf{B}}_{\text{Hall}} \quad , \tag{17.23}$$

is balanced by a combination of five terms designated with a commonly used names. The first term, labeled "Ohmic", arises from the drag due to electron-ion collisions. This is the one we associate with traditional Ohm's law for a medium with conductivity σ: $\mathbf{E}' = \mathbf{J}/\sigma$. It seems that electron-ion collisions gives rise, in a fully-ionized plasma, to electrical conductivity

$$\sigma = \frac{\omega_{p,e}^2\tau_{ei}}{4\pi} = \frac{e^2 n_e}{m_e}\tau_{ei} \quad . \tag{17.24}$$

The final expression is sometimes derived using the so-called *Drude* model for a conductor.

It would appear that when we make use of the traditional Ohm's law we are dropping all the other terms on the right of eq. (17.23). Is this justified? We show below that those terms are all typically small compared to those on the left hand side, namely $\mathbf{u}\times\mathbf{B}/c$, so it is usually permissible to discard them. The Ohmic term is *also* generally very small, as we have remarked before, so this should be discarded as well, leaving $\mathbf{E}' = 0$ — ideal Ohm's law. The Ohmic term is, in fact, often smaller than some of the other terms on the right, so the traditional form of Ohm's law, including only $\mathbf{J}/\sigma$, is typically *not* a good approximation in an astrophysical plasma.

Let us first consider the magnitude of the Ohmic term. Scaling it agains the motional EMF, on the left,

$$\frac{\text{Ohm}}{|\mathbf{u}\times\mathbf{B}|/c} \sim \frac{cB/\omega_{p,e}^2\tau_{ei}L}{uB/c} = \frac{c^2/\omega_{p,e}^2\tau_{ei}}{uL} = \frac{4\pi c^2/\sigma}{uL} = \frac{\eta}{uL} = \frac{1}{R_m} \quad , \tag{17.25}$$

where $\eta = 4\pi c^2/\sigma$ is the magnetic diffusivity and R_m is the magnetic Reynold's number. We have seen before that $R_m \gg 1$ in virtually all astrophysical contexts. A second view of this results follows from

$$R_m = \frac{|\mathbf{u}\times\mathbf{B}|/c}{\text{Ohm}} \sim \frac{uL}{c^2/\omega_{p,e}^2\tau_{ei}} = \frac{L^2}{d_e^2}\frac{u\tau_{ei}}{L} = \frac{L^2}{d_e^2}\frac{\tau_{ei}}{\tau} \quad , \tag{17.26}$$

where $d_e = c/\omega_{p,e}$ is the electron collisonless skin depth (43 times smaller than the ion skin depth, which is already very small) and $\tau = L/u$ is the hydrodynamic time scale. This

shows clearly why magnetic Reynold's number are so enormous in astrophysical plasmas. While collision times can often be short, d_e is so small compared to L that even short collision times are overwhelmed by the ratio $L/d_e \gg 1$ squared.

It is relatively easy to compare the electron inertia term, at least the $\partial \mathbf{J}/\partial t$ part of it, to the Ohmic term because of their very similar forms,

$$\frac{e \text{ inertia}}{\text{Ohm}} \sim \frac{\tau_{ei}}{\tau} \quad . \tag{17.27}$$

We can therefore neglect electron inertia in the face of electron-ion drag when collision times are short compared to fluid time scales. The same limit is invoked to close the fluid equations — short collision times are needed to keep the distribution function close to Maxwellian. This is what is done in traditional Ohm's law: the collision time is assumed to be so short that the electrons reach "terminal velocity" almost immediately. It is for that reason that we find a relation whereby a "force" (namely electric field) is proportional to a "velocity" (namely current), rather than to an acceleration as Newton's law would have. In discarding the electron inertia term, however, we eliminate the very term which would permit us to update the current in response to a driving force. We are left instead with an instantaneous relation which the current density must satisfy.

When collision times are not short we can compare the electron inertia term directly to the motional EMF. The ratio of eqs. (17.27) to (17.26) gives

$$\frac{e \text{ inertia}}{|\mathbf{u} \times \mathbf{B}|/c} \sim \frac{d_e^2}{L^2} \quad , \tag{17.28}$$

which will be *very small* unless the magnetic field is structured on scales comparable to $d_e = c/\omega_{p,e}$ — *very* small scales. Thus it is possible in virtually all astrophysical circumstances to discard the electron inertia term. This leaves the state of affairs, now familiar in MHD, that current is determined by an instantaneous relation rather than evolved like velocity.

Comparing the Hall term to the motion EMF gives

$$\frac{\text{Hall}}{|\mathbf{u} \times \mathbf{B}|/c} \sim \frac{|\mathbf{v}_d|}{|\mathbf{u}|} \sim \frac{d_i}{L}\frac{v_\text{A}}{u} \quad , \tag{17.29}$$

after using (17.8). Here again we see a comparison of the global length scale to a skin depth — in this case the ion skin depth. This ratio is so small that even when the flow is very sub-Alfvénic, the Hall term can usually be ignored. Since the ratio involves the small ratio to the first power, the Hall term is, in fact, typically larger than either the Ohmic term or the electron inertia term. It would therefore be difficult to justify keeping the Ohmic term and dropping the Hall term, although this is precisely what is done when using traditional Ohm's law.

The viscous stress tensor, $-\underline{\underline{\pi}}_e$, is proportional to the electron rate of strain tensor, in the traditional fluid closure. The electron velocity matches the ion velocity to leading order so the rate of strain will scale as $\partial u_i/\partial x_j \sim u/L$. In a magnetized plasma the viscosity coefficient is a fourth-rank tensor with five distinct elements. The largest of these acts parallel to the magnetic field and is the same as the unmagnetized viscosity $\mu \sim \tau_{ee} p_e$, where τ_{ee} is the electron-electron collision time. Using these values gives a magnitude estimate

$$\frac{e \text{ viscosity}}{|\mathbf{u} \times \mathbf{B}|/c} \sim \frac{\tau_{ee} p_e c/e n_e}{L^2 B} \sim \frac{\tau_{ee} v_{th,e}^2}{L^2 \Omega_e} \sim \frac{r_{ge}}{L}\frac{\ell_{mfp,e}}{L} \quad , \tag{17.30}$$

where $r_{ge} = v_{th,i}/\Omega_e$ is the electron gyroradius, and $\ell_{mfp,e} = \tau_{ee} v_{th,e}$ is the electron mean free path. Both of these length scales are very small, and the electron viscosity coefficient is therefore negligible. The viscous stress across the magnetic field is smaller than the parallel component by at least a factor of $r_{ge}/\ell_{mfp,e}$. Were we to use this contribution instead the ratio $\sim (r_{ge}/L)^2$ would be on par with the electron inertia contribution in eq. (17.28).

The final contribution, the polarization electric field, does actually make a contribution to traditional MHD, even if it is often omitted when writing Ohm's law. The MHD momentum equation principally expresses the evolution of the ions, since they carry most of the momentum. In spite of this, there is a contribution from the electron pressure, which adds to the ion pressure to form the total plasma pressure. While this seems natural, it should give pause since pressure is not a body force, so there is no way for the ions to "feel" the pressure of the electrons directly. What has actually happened is that the electron pressure is balanced in part by the electric field. When the electron momentum equation was added to the ion equation the electric field contributions from the two species cancelled, but the balanjng electron pressure remained. This electron pressure term mostly represents the electric field which the ions do feel.

Polarization electric field can, for example, arise in response to the need for quasi-neutrality. Say there is some force on the ion fluid, perhaps simple inertia, which works to increase its density. The electron fluid must be compressed as well, otherwise there would be a significant charge imbalance. The polarization electric field is a small manifestation of this effect, and it is responsible for compressing the electron fluid in order to maintain charge neutrality. This same field exerts an equal and opposite force on the ions, thereby serving as an intermediary through which the ions may "feel" the pressure of the electrons as they resist compression.

The polarization electric field can have a substantial electrostatic component. If, for example, the electron fluid is isothermal then the polarization electric field is

$$\mathbf{E}_{\rm pol} \;=\; -\frac{\nabla p_e}{e n_e} \;=\; -\nabla\left[\frac{k_{\rm B} T_e}{e}\,\ln(n_e)\right] \quad . \tag{17.31}$$

In a million Kelvin isothermal atmosphere the electrostatic potential drops 86 Volts over every scale height risen. The downward electric field is responsible for confining the electrons to the same stratification as the ions and thereby maintaining quasi-neutrality.

Appendix A

Useful numbers for Solar Plasmas

$c =$	2.99E+10	cm/sec	$k_{\rm B} =$	1.38E−16	erg/K	$G =$	6.67E−8	erg cm g^{-2}
$m_e =$	0.911E−27	gm	$m_p =$	1.67E−24	gm	$e =$	4.80E−10	esu

SOLAR PROPERTIES

Mass	$M_\odot$ = 1.99E+33	gm	Luminosity	$L_\odot$ = 3.83E+33	erg/sec
Radius	$R_\odot$ = 6.96E+10	cm	surface gravity	$g_\odot$ = 2.74E+4	cm/sec^2
1 AU	$D_\odot$ = 1.50E+13	cm	escape speed	v_∞ = 6.18E+7	cm/sec
Synodic period†	2.38E+6	sec	Siderial frequency†	$\Omega_\odot$ = 2.84E−6	rad/sec

† Equatorial; Synodic/siderial period: 27.56/25.62 days (doppler shift — Snodgrass [1984])

Definitions: $T = T_6 \times 10^6\,{\rm K}$, $n_e = n_i = n_9 \times 10^9\,{\rm cm}^{-3}$, $B = B_2 \times$ 100G

PLASMA PROPERTIES			electron	proton	×	units
plasma frequency	$\omega_{ps} =$	$\sqrt{4\pi n_s e^2/m_s}$	1.8E+9	4.2E+7	$n_9^{1/2}$	rad/sec
gyro-frequency	$\Omega_s =$	$eB/m_s c$	1.8E+9	.96E+6	B_2	rad/sec
Coulomb logarithm	$\ln\Lambda$		18 + ln($T_6^{3/2} n_9^{-1/2}$)			—
thermal speed	$v_s =$	$\sqrt{k_b T_s/m_s}$	3.9E+8	.91E+7	$T_6^{1/2}$	cm/sec
gyro-radius	$\rho_s =$	v_s/Ω_s	0.22	9.5	$T_6^{1/2} B_2^{-1}$	cm
plasma skin depth	$d_s =$	c/ω_{ps}	17	718	$n_9^{-1/2}$	cm
Debeye length	$\lambda_s =$	v_s/ω_{ps}	0.22	0.22	$T_6^{1/2}\, n_9^{-1/2}$	cm
collision time	$\tau_s =$	$0.30\,(m_s^2/e^4) v_s^3/n \ln\Lambda$	1.5E−2	.93†	$T_6^{3/2}\, n_9^{-1}$	sec
collision frequency	$\nu_s =$	$1/\tau_s$	65	1.08	$T_6^{-3/2} n_9$	Hz
mean-free path	$\ell_s =$	$v_s \tau_s$	5.9E+6	8.4E+6	$T_6^2\, n_9^{-1}$	cm
Stopping column	$N =$	$E^2/6\pi e^4 \ln\Lambda$	1.4E+17		$E_{\rm keV}^2$	cm^{-2}

† $\tau_i = \sqrt{2}\sqrt{m_i/m_e}\tau_e$

MHD PROPERTIES				×	units
Alfvén speed	$v_{\rm A} =$	$B/\sqrt{4\pi\rho}$	6.9E+8	$B_2\, n_9^{-1/2}$	cm/sec
sound speed	$c_{\rm s} =$	$\sqrt{2\gamma k_b T/m_p}$	1.7E+7	$T_6^{1/2}$	cm/sec
plasma β	$\beta =$	$8\pi p/B^2$	6.9E−4	$T_6\, n_9/B_2^2$	—
scale height	$H_p =$	$2k_bT/m_pg_\odot$	6.0E+9	T_6	cm
electric conductivity† ..	$\sigma =$	$0.16\omega_e^2\tau_e$	7.8E+15	$T_6^{3/2}$	$\sec^{-1}$
thermal conductivity† .	$\kappa =$	$3.2\, k_b n_e\, v_e^2 \tau_e$	1.0E+9	$T_6^{5/2}$	erg (cm s K)$^{-1}$
Spitzer current (cgs) ...	$I_{\rm sp}/c =$	$\eta\sqrt{\rho/4\pi}$	1.4E−4	$T_6^{-3/2}\, n_9^{1/2}$	G cm
(MKS)	$I_{\rm sp} =$	$\eta\sqrt{\rho/\mu_0}$	1.4E−3	$T_6^{-3/2}\, n_9^{1/2}$	Amps
Dreicer field (cgs)	$E_{\rm D} =$	$e\ln\Lambda/\lambda_D^2$	1.8E−7	n_9/T_6	G
(MKS)	$E_{\rm D} =$	$e\ln\Lambda/4\pi\epsilon_0\lambda_D^2$ ...	5.9E−3	n_9/T_6	Volts/m
conductive time	$\tau_{\rm cond} =$	$2n_ek_bL^2/\kappa$	270	$n_9\, T_6^{-5/2} L_9^2$	sec
radiative time°	$\tau_{\rm rad} =$	$2k_bT/n_e\Lambda(T)$	2.3E+3	$T_6^{3/2}/n_9$	sec
diffusion coefficients					
viscosity†	$\nu =$	$0.96\, v_i^2\tau_i$	7.3E+13	$T_6^{5/2}\, n_9^{-1}$	cm^2/sec
magnetized viscosity‡ ..	$\nu_\perp =$	$0.3\, \rho_i^2/\tau_i$	29	$T_6^{-1/2}\, n_9\, B_2^{-2}$	cm^2/sec
thermal conductivity† .	$\tilde\kappa =$	$(\gamma-1)\kappa/k_bn_e$ ...	4.9E+15	$T_6^{5/2}\, n_9^{-1}$	cm^2/sec
resistivity†	$\eta =$	$c^2/4\pi\sigma$	.92E+4	$T_6^{-3/2}$	cm^2/sec

† Component $\parallel$ to $\mathbf{B}$ $\quad \sigma_\perp = 0.51\sigma$ and $\eta_\perp = 1.96\eta$

‡ Coupling rate-of-strain & stress $\perp$ to $\mathbf{B}$

∘ $\Lambda(T) = 1.2\times10^{-22}\, T_6^{-1/2}$ erg cm^3/s good for $300,000\ {\rm K} < T < 30$ MK.

Appendix B: 2d Magnetic Equilibria

Complex potentials: $A_z(x,y) = \mathrm{Re}\{\hat{F}(x+iy)\}$

$$\begin{aligned} B_y + iB_x &= -\hat{F}'(w) \\ \text{wire: } \hat{F}(w) &= -\frac{4\pi I}{c}\frac{\ln w}{2\pi} \end{aligned}$$

Green-Syrovatskii CS of length $\Delta = 2\delta$

$$\begin{aligned} \hat{F}(w) &= \tfrac{1}{2}B'w\sqrt{w^2-\delta^2} \\ &\quad -\tfrac{1}{2}B'\delta^2\ln(w+\sqrt{w^2-\delta^2}) \\ B_y + iB_x &= -B'\sqrt{w^2-\delta^2} \\ \frac{4\pi I}{c} &= \pi B'\delta^2 = \frac{\pi}{4}B'\Delta^2 \end{aligned}$$

Axisymmetric fields: $\mathbf{B} = \nabla f \times \nabla\phi$

$$\begin{aligned} \Phi(\varpi<\rho) &= 2\pi f(\rho,z) \\ \nabla\times\mathbf{B} &= \varpi^2\nabla\cdot\left(\frac{\nabla f}{\varpi^2}\right)\nabla\phi \equiv \Delta^* f\,\nabla\phi \\ &= \left[\frac{\partial^2 f}{\partial r^2} + \frac{1-\mu^2}{r^2}\frac{\partial^2 f}{\partial\mu^2}\right]\nabla\phi \\ &= \left[\varpi\frac{\partial}{\partial\varpi}\left(\frac{1}{\varpi}\frac{\partial f}{\partial\varpi}\right) + \frac{\partial^2 f}{\partial z^2}\right]\nabla\phi \end{aligned}$$

2^p-pole potential fields: $f_p(r,\mu)$

$$\begin{aligned} f_0 &= 1-\mu = 1 - \frac{z}{\sqrt{\varpi^2+z^2}} \\ f_1 &= \frac{1-\mu^2}{r} = \frac{\varpi^2}{(\varpi^2+z^2)^{3/2}} \\ f_2 &= \frac{\mu(1-\mu^2)}{3r^2} = \frac{z\varpi^2}{3(\varpi^2+z^2)^{5/2}} \end{aligned}$$

Current ring @ $(\varpi,z) = (\rho,0)$

$$\begin{aligned} f(\varpi,z) &= \frac{4\pi I}{c}\frac{\sqrt{\rho\varpi}}{4\pi k}\left[(2-k^2)K(k) - 2E(k)\right] \\ k^2 &= \frac{4\rho\varpi}{(\varpi+\rho)^2+z^2} \\ K(k) &= \int_0^{\pi/2}\frac{d\theta}{\sqrt{1-k^2\cos^2\theta}} \\ &\simeq \tfrac{1}{2}\ln\left(\frac{16}{1-k^2}\right)\ ,\ k\to 1 \end{aligned}$$

Force-free equilibria:
Lundquist field (const.-α)

$$\mathbf{B} = B_0[\,J_0(\alpha\varpi)\hat{\mathbf{z}} + J_1(\alpha\varpi)\hat{\boldsymbol{\phi}}\,]\ ,\ \nabla\times\mathbf{B} = \alpha\mathbf{B}$$

$$\begin{aligned} \Phi(\varpi<\rho) &= \frac{2\pi\rho B_0}{\alpha^2}J_1(\alpha\rho) \\ \Phi(\varpi<j_{0,1}/\alpha) &= 0.431\,\pi(j_{0,1}/\alpha)^2 B_0 \end{aligned}$$

Gold-Hoyle field (const. pitch $q = B_\phi/\varpi B_z$)

$$\mathbf{B} = \frac{B_0(\hat{\mathbf{z}} + q\varpi\hat{\boldsymbol{\phi}})}{1+q^2\varpi^2}\ ,\ \nabla\times\mathbf{B} = \frac{2q\mathbf{B}}{1+q^2\varpi^2}$$

$$\begin{aligned} \Phi(\varpi<\rho) &= \frac{\pi B_0}{q^2}\ln(1+q^2\rho^2) \\ \frac{4\pi I}{c}(\varpi<\rho) &= \frac{2\pi B_0}{q}\frac{q^2\rho^2}{1+q^2\rho^2} \end{aligned}$$

Axisymmetric constant-α field:

$$\mathbf{B} = \nabla f\times\nabla\phi + \alpha f\nabla\phi\ \ ,\ \ B_\phi = \alpha f/\varpi$$

$$\Delta^* f = -\alpha^2 f$$

2^p-pole solutions:

$$\begin{aligned} f_0 &= -(1-\mu)\alpha r\, n_0(\alpha r) = (1-\mu)\cos(\alpha r) \\ f_1 &= -(1-\mu^2)\,\alpha^2 r\, n_1(\alpha r) \\ &= (1-\mu^2)\left[\frac{\cos(\alpha r)}{r} + \alpha\sin(\alpha r)\right] \\ f_2 &= -\mu(1-\mu^2)\,\alpha^3 r\, n_2(\alpha r) \end{aligned}$$

Point-source: (Chiu & Hilton, 1977)

$$f = \cos(\alpha\mu r) - \mu\cos(\alpha r)$$

Spheromak: $\nabla\times\mathbf{B} = \alpha = 0\ \ ,\ \ r > a$

$$f = \begin{cases} f_0\frac{\xi r}{a}j_1\left(\frac{\xi r}{a}\right)\sin^2\theta, & r\le a \\ f_0 C\xi\left(\frac{a}{r} - \frac{r^2}{a^2}\right)\sin^2\theta & r > a \end{cases}$$

$$\begin{aligned} j_1(x) &= \frac{\sin x}{x^2} - \frac{\cos x}{x}\ ,\ j_1(\xi) = 0 = \tan\xi - \xi \\ \xi &= 4.49341\ ,\ C = -\tfrac{1}{3}\sin\xi = 1.46204 \\ \alpha &= \xi/a\ ,\ \text{axis: } r_0 = 0.61067a \end{aligned}$$

Appendix C

Moments of a Maxwellian Distribution

The Maxwellian distribution function for a local density, n, mean velocity $\mathbf{u}$ and temperature T, is defined

$$f_0(\mathbf{u}) \;=\; \frac{n}{(2\pi k_{\rm B}T/m)^{3/2}}\exp\left[-\frac{|\mathbf{u}-\mathbf{u}|^2}{2k_{\rm B}T/m}\right] \;=\; \frac{n}{(2\pi)^{3/2}v_{\rm th}^3}\exp\left[-\frac{|\mathbf{u}-\mathbf{u}|^2}{2v_{\rm th}^2}\right] \quad , \qquad \text{(C.1)}$$

where $v_{\rm th} = \sqrt{k_{\rm B}T/m}$ is the thermal speed. Using this in a normalized velocity $\boldsymbol{\xi} = (\mathbf{u}-\mathbf{u})/v_{\rm th}$, converts the distribution function to

$$f_0(\mathbf{u})d^3v \;=\; f_0(\xi)\,d^3\xi \;=\; \frac{n}{(2\pi)^{3/2}}\,e^{-\xi^2/2}\,d^3\xi \quad . \qquad \text{(C.2)}$$

The general moment of this distribution is

$$\begin{aligned}\int \xi^{2\mu} f_0(\xi)\,d^3\xi \;&=\; \frac{4\pi n}{(2\pi)^{3/2}}\int_0^\infty \xi^{2+2\mu}e^{-\xi^2/2}\,d\xi \\ &=\; \frac{2n}{\sqrt{2\pi}}\,2^{\mu+1/2}\int_0^\infty s^{\mu+1/2}e^{-s}\,ds \;=\; \frac{2^{\mu+1}n}{\sqrt{\pi}}\Gamma\left(\mu+\tfrac{3}{2}\right)\end{aligned}$$

Noting that $\Gamma(1/2) = \sqrt{\pi}$ and $\Gamma(3/2) = \sqrt{\pi}/2$ and $\Gamma(5/2) = 3\sqrt{\pi}/4$ we can compile the results

$$\begin{aligned}\int f_0(\xi)\,d^3v \;&=\; n \\ \int \xi^2 f_0(\xi)\,d^3v \;&=\; 3n \\ \int \xi^4 f_0(\xi)\,d^3v \;&=\; 3\cdot 5n \\ \int \xi^6 f_0(\xi)\,d^3v \;&=\; 3\cdot 5\cdot 7n\end{aligned}$$

The second rank tensor moments are

$$\int \boldsymbol{\xi}\boldsymbol{\xi} f_0(\xi)\,d^3v \;=\; n\,\underline{\underline{I}}$$

$$\begin{aligned}\int \xi^2 \boldsymbol{\xi}\boldsymbol{\xi} f_0(\xi)\, d^3v &= 5n\, \underline{\underline{I}} \\ \int \xi^4 \boldsymbol{\xi}\boldsymbol{\xi} f_0(\xi)\, d^3v &= 35n\, \underline{\underline{I}}\end{aligned}$$

where $\underline{\underline{I}}$ is the identity tensor.

The fourth moment will be non-vanishing only when the four indices occur in pairs. This means it has the general form

$$\int \xi_i\xi_j\xi_k\xi_l\, f_0(\xi)\, d^3v = n[\, b_1(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) + b_2\, \delta_{ij}\delta_{jk}\delta_{kl}\,] \tag{C.3}$$

We can find the two coefficients from considering $i = j = 1$ and $k = l = 2$:

$$n^{-1}\int \xi_x^2\xi_y^2 f_0(\xi)\, d^3v = b_1 = \left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} s^2 e^{-s^2/2} ds\right]^2 = 1$$

and then $i = j = k = l = 1$

$$n^{-1}\int \xi_x^4 f_0(\xi)\, d^3v = 3 + b_2 = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} s^4 e^{-s^2/2} ds = \frac{2^{5/2}}{\sqrt{2\pi}}\Gamma\left(\tfrac{5}{2}\right) = 3$$

which means $b_2 = 0$. In other words the fourth moment is

$$\int \xi_i\xi_j\xi_k\xi_l\, f_0(\xi)\, d^3v = n(\delta_{ij}\delta_{kl} + \delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) \tag{C.4}$$

References

Batchelor, G. K. 1967, Introduction to Fluid Dynamics (Cambridge: Cambridge Univ. Press)

Born, M., & Wolf, E. 1980, Principles of Optics: Electromagnetic theory of propagation and diffraction of light (sixth ed.) (New York, NY: Pergamon Press)

Choudhuri, A. R. 1998, Physics of the Fluids and Plasmas: An introduction for astrophysicists (Cambridge, UK: Cambridge Univ. Press)

Christensen-Dalsgaard, J., et al. 1996, Science, 272, 1286

Clayton, D. D. 1983, Principles of stellar evolution and nucleosynthesis (Chicago: University of Chicago Press)

Cook, J. W., Cheng, C.-C., Jacobs, V. L., & Antiochos, S. K. 1989, ApJ, 338, 1176

Craig, I. J. D., McClymont, A. N., & Underwood, J. H. 1978, A&A, 70, 1

Feldman, U. 1992, Physica Scripta, 46, 202

Fetter, A. L., & Walecka, J. D. 1980, Theoretical Mechanics of Particles and Continua (New York, NY: McGraw Hill)

Foukal, P. V. 2004, Solar Astrophysics (Weinheim: Wiley)

Guckenheimer, J., & Holmes, P. 1983, Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields, Vol. 42 (New York: Springer-Verlag)

Kulsrud, R. M. 2005, Plasma Physics for Astrophysics (Pinceton, NJ: Princeton University Press)

Lamb, H. 1932, Hydrodynamics (6th ed.) (New York: Dover)

Landau, L. D., & Lifshitz, E. M. 1959, Fluid Mechanics, Vol. 6 (New York: Pergamon)

Landau, L. D., & Lifshitz, E. M. 1986, Theory of Elasticity (3rd ed.) (Butterworth-Heinemann)

Lifshitz, E. M., & Pitaevskii, L. P. 1981, Physical Kinetics (Oxford, UK: Butterworth-Heinemann)

Marion, J. B. 1970, Classical Dynamics of Particles and Systems (second ed.) (Academic Press, New York)

Marion, J. B., & Thornton, S. T. 1988, Classical Dynamics of Particles and Systems (third ed.) (Harcourt Brace Jovanovich, San Diego)

Martens, P. C. H., Kankelborg, C. C., & Berger, T. E. 2000, ApJ, 537, 471

Meyer, J.-P. 1985, ApJ Supp, 57, 173

Meyer, R. E. 1971, Introduction to Mathematical Fluid Dynamics (New York: Dover)

Moffatt, H. K. 1978, Magnetic Field Generation in Electrically Conducting Fluids (Cambridge: Cambridge University Press)

Priest, E. R. 1982, Solar Magnetohydrodynamics, Vol. 21 (Boston: D. Reidel)

Riley, K. F., Hobson, M. P., & Bence, S. J. 2006, Mathematical Methods for Physics and Engineering (third edition ed.) (Cambridge, UK: Cambridge University Press)

Rosner, R., Tucker, W. H., & Vaiana, G. S. 1978, ApJ, 220, 643

Rybicki, G. B., & Lightman, A. P. 1979, Radiative Processes in Astrophysics (New York, NY: Wiley and Sons)

Schwartzschild, M. 1958, Structure and Evolution of the Stars (New York, N.Y.: Dover)

Thompson, M. J. 2006, An Introduction to Astrophysical Fluid Dynamics (London, UK: Imperial College Press)

Tritton, D. J. 1977, Physical Fluid Dynamics (Wokingham, UK: Van Nostran Reinhold)

Weinberg, S. 1962, Phys. Rev., 126, 1899

Index