Cambridge University Press
978-1-107-03019-0 - Accretion Processes in Astrophysics: XXI Canary Islands Winter School of Astrophysics
Edited by Ignacio González Martínez-País, Tariq Shahbaz and Jorge Casares Velázquez
Excerpt
More information

1. Accretion disks

HENK SPRUIT

Abstract

In this lecture the basic theory of accretion disks is reviewed, with emphasis on aspects relevant for X-ray binaries and cataclysmic variables. The text gives a general introduction as well as a selective discussion of a number of more recent topics.

1.1 Introduction

Accretion disks are inferred to exist as objects of very different scales: millions of kilometers in low mass X-ray binaries (LMXB) and cataclysmic variables (CV), solar-radius-to-AU–scale disks in protostellar objects, and AU-to-parsec-scale disks in active galactic nuclei (AGN).

An interesting observational connection exists between accretion disks and jets (such as the spectacular jets from AGN and protostars) and outflows (the "CO-outflows" from protostars and the "broad-line regions" in AGN). Lacking direct (i.e., spatially resolved) observations of disks, theory has tried to provide models, with varying degrees of success. Uncertainty still exists with respect to some basic questions. In this situation, progress made by observations or modeling of a particular class of objects has direct impact on the understanding of other objects, including the enigmatic connection with jets.

In this lecture I concentrate on the more basic aspects of accretion disks, but an attempt is made to mention topics of current interest as well. Some emphasis is on those aspects of accretion disk theory that connect to the observations of LMXB and CVs. For other reviews on the basics of accretion disks, see Pringle (1981) and Papaloizou and Lin (1995). For a more extensive introduction, see the textbook by Frank *et al.* (2002). For a comprehensive text on CVs, see Warner (1995).

1.2 Accretion: general

Gas falling into a point mass potential

$$\Phi = -\frac{GM}{r} \tag{1.1}$$

from a distance r_0 to a distance r converts gravitational into kinetic energy by an amount $\Delta\Phi = GM(1/r - 1/r_0)$. For simplicity, assuming that the starting distance is large, $\Delta\Phi = GM/r$. The speed of arrival, the *free-fall speed* $v_{\rm ff}$, is given by

$$\frac{1}{2}v_{\rm ff}^2 = \frac{GM}{r}. \tag{1.2}$$

If the gas is then brought to rest, for example at the surface of a star, the amount of energy e dissipated per unit mass is

$$e = \frac{1}{2}v_{\rm ff}^2 = \frac{GM}{r} \qquad \text{(rest)}. \tag{1.3}$$

If, instead, it goes into a circular Kepler orbit at distance r:

$$e = \frac{1}{2}\frac{GM}{r} \qquad \text{(orbit)}. \tag{1.4}$$

The dissipated energy may go into internal energy of the gas, and into radiation, which escapes to infinity (usually in the form of photons, but neutrino losses can also play a role in some cases).

1.2.1 *Adiabatic accretion*

Consider first the case when radiation losses are neglected. Any mechanical energy dissipated stays locally in the flow. This is called an *adiabatic* flow (not to be confused with *isentropic* flow). For an ideal gas with constant ratio of specific heats γ, the internal energy per unit mass is

$$e = \frac{P}{(\gamma - 1)\rho}. \tag{1.5}$$

With the equation of state

$$P = \frac{\mathcal{R}\rho T}{\mu}, \tag{1.6}$$

where $\mathcal{R}$ is the gas constant and μ the mean atomic weight per particle, we find the temperature of the gas after the dissipation has taken place (assuming that the gas goes into a circular orbit):

$$T = \frac{1}{2}(\gamma - 1)T_{\mathrm{vir}}, \tag{1.7}$$

where T_{vir}, the *virial temperature*, is defined as

$$T_{\mathrm{vir}} = \frac{GM\mu}{\mathcal{R}r} = \frac{g\,r\mu}{\mathcal{R}}, \tag{1.8}$$

where g is the acceleration of gravity at distance r. In an atmosphere with temperature near T_{vir}, the sound speed $c_{\mathrm{s}} = (\gamma\mathcal{R}T/\mu)^{1/2}$ is close to the escape speed from the system, and the hydrostatic pressure scale height, $H \equiv \mathcal{R}T/(\mu g)$, is of the order of r. Such an atmosphere may evaporate on a relatively short time scale in the form of a stellar wind. This is as expected from energy conservation: if no energy is lost through radiation, the energy gained by the fluid while falling into a gravitational potential is also sufficient to move it back out again.

A simple example is spherically symmetric adiabatic accretion (Bondi, 1952). An important result is that such accretion is possible only if $\gamma \leq 5/3$. The lower γ, the lower the temperature in the accreted gas (eq. 1.7), and the easier it is for the gas to stay bound in the potential. A classical situation where adiabatic and roughly spherical accretion takes place is a supernova implosion: when the central temperature becomes high enough for the radiation field to start disintegrating nuclei, γ drops and the envelope collapses onto the forming neutron star via an accretion shock. Another case is Thorne-Zytkow objects (e.g., Cannon *et al.*, 1992), where γ can drop to low values due to pair creation, initiating an adiabatic accretion onto the black hole.

Adiabatic spherical accretion is fast, taking place on the dynamical time scale: something on the order of the free fall time scale, or Kepler orbital time scale,

$$\tau_{\mathrm{d}} = \frac{r}{v_{\mathrm{K}}} = \Omega_{\mathrm{K}}^{-1} = \left(\frac{r^3}{GM}\right)^{1/2}, \tag{1.9}$$

where v_{K} and Ω_{K} are the Kepler orbital velocity and angular frequency, respectively.

When radiative loss becomes important, the accreting gas can stay cool irrespective of the value of γ, and Bondi's critical value $\gamma = 5/3$ plays no role. With such losses, the temperatures of accretion disks are usually much lower than the virial temperature.

1.2.2 *Temperature near compact objects*

For accretion onto a neutron star surface, $R = 10$ km, $M = 1.4$ M$_\odot$, we have a free fall speed $v_{\mathrm{ff}}/c \approx 0.4\,c$ (this in Newtonian approximation; the correct value in general

Cambridge University Press
978-1-107-03019-0 - Accretion Processes in Astrophysics: XXI Canary Islands Winter School of Astrophysics
Edited by Ignacio González Martínez-País, Tariq Shahbaz and Jorge Casares Velázquez
Excerpt
More information

relativity is quantitatively somewhat different). The corresponding virial temperature would be $T_{\rm v} \sim 2\ 10^{12}$ K, equivalent to an average energy of 150 MeV per particle.

This is not the actual temperature we should expect, since other things happen before such temperatures are reached. If the accretion is adiabatic, one of these is the creation of a very dense radiation field. The energy liberated per infalling particle is still the same, but it is shared among a large number of photons. At temperatures above the electron rest mass (≈ 0.5 MeV), electron-positron pairs $e^{\pm}$ are produced in addition to photons. These can take up most of the accretion energy, limiting the temperature typically to a few MeV.

In most observed disks, however, temperatures do not get even close to 1 MeV, because accretion is rarely adiabatic. Energy loss takes place by escaping photons (or, under more extreme conditions, neutrinos). Exceptions are the radiatively inefficient accretion flows discussed in Section 1.13.

Radiative loss

Next to the adiabatic temperature estimates, a useful characteristic number is the *blackbody effective temperature*. Here, the approximation made is that the accretion energy is radiated away from an optically thick surface of some geometry, under the assumption of a balance between the heating rate by release of accretion energy and cooling by radiation. For a specific example, consider the surface of a star of radius R and mass M accreting via a disk. Most of the gravitational energy is released close to the star, in a region with surface area of the order (let us call it) $4\pi R^2$. If the surface radiates approximately as a blackbody of temperature T (make a note of the fact that this is a bad approximation if the opacity is dominated by electron scattering), the balance would be

$$\dot{M}\frac{GM}{R} \approx 4\pi R^2 \sigma_{\rm r} T^4, \tag{1.10}$$

where $\sigma_{\rm r}$ is the Stefan-Boltzmann constant, $\sigma_{\rm r} = a_{\rm r} c/4$, if $a_{\rm r}$ is Planck's radiation constant. For a neutron star with $M = 1.4\ {\rm M}_\odot = 3 \times 10^{33}$ g, $R = 10$ km, accreting at a typical observed rate (near Eddington rate; see later), $\dot{M} = 10^{18}\ {\rm g\ s^{-1}} \approx 10^{-8}\ {\rm M}_\odot\ {\rm yr}^{-1}$, this temperature would be $T \approx 10^7$ K, or $\approx$1 keV per particle. Radiation with this characteristic temperature is observed in accreting neutron stars and black holes in their so-called soft X-ray states (as opposed to their hard states, in which the spectrum is very far from a blackbody).

The kinds of processes involved in radiation from accretion flows form a large subject in itself and are not covered here. For introductions, see Rybicki and Lightman (1979) and Frank *et al.* (2002). At the moderate temperatures encountered in protostars and white dwarf accretors, the dominant processes are the ones known from stellar physics: molecular and atomic transitions and Thomson scattering. Up to photon energies around 10 keV (the Lyman edge of an iron nucleus with one electron left), these processes also dominate the spectra of neutron star and black hole accretors. Above this energy, observed spectra become dominated by Compton scattering. Cyclotron/synchrotron radiation plays a role when a strong magnetic field is present, which can happen in most classes of accreting objects.

1.2.3 *Critical luminosity*

Objects of high luminosity have a tendency to blow their atmospheres away because of the radiative force exerted when the outward-traveling photons are scattered or absorbed. Consider a volume of gas on which a flux of photons is incident from one side. Per gram of matter, the gas presents a scattering (or absorbing) surface area of κ cm^2 to the escaping radiation. The force exerted by the radiative flux F on 1 g is $F\kappa/c$. The force of gravity pulling back on this 1 g of mass is GM/r^2. The critical flux at which the two forces balance (energy per unit area and time) is

$$F_{\rm E} = \frac{c}{\kappa}\frac{GM}{r^2}. \tag{1.11}$$

Assuming that this flux is *spherically symmetric*, it can be converted into a luminosity,

$$L_{\rm E} = \frac{4\pi GMc}{\kappa}, \tag{1.12}$$

the Eddington critical luminosity, popularly called the Eddington limit (e.g., Rybicki and Lightman, 1979). If the gas is fully ionized, its opacity is dominated by electron scattering, and for solar composition, κ is then of the order 0.3 cm^2 g^{-1} (about a factor of 2 lower than for fully ionized helium). With these assumptions,

$$L_{\rm E} \approx 1.7\,10^{38}\frac{M}{M_\odot}\ \ {\rm erg\ s^{-1}} \approx 4\,10^4\frac{M}{M_\odot}\ {\rm L_\odot}. \tag{1.13}$$

This number is different if the opacity is not dominated by electron scattering. In partially ionized gases of solar composition, bound-bound and bound-free transitions can increase the opacity by a factor up to 10^3; the Eddington flux is then correspondingly lower.

If the luminosity results from accretion, one can define a corresponding Eddington characteristic accretion rate $\dot M_{\rm E}$:

$$\frac{GM}{r}\dot M_{\rm E} = L_{\rm E} \quad\rightarrow\quad \dot M_{\rm E} = \frac{4\pi rc}{\kappa}. \tag{1.14}$$

With $\kappa = 0.3$:

$$\dot M_{\rm E} \approx 1.3\,10^{18} r_6\ \ {\rm g\ s^{-1}} \approx 2\,10^{-8} r_6\ \ {\rm M_\odot\ yr^{-1}}, \tag{1.15}$$

where r_6 is the radius of the accreting object in units of 10^6 cm. The characteristic accretion rate thus scales with the *size* of the accreting object, while the critical luminosity scales with *mass*.

Whereas $L_{\rm E}$ is a critical value that in several circumstances plays the role of a limit, the Eddington characteristic accretion rate is less of a limit. For more on exceptions to $L_{\rm E}$ and $\dot M_{\rm E}$, see Section 1.2.3.

Eddington luminosity at high optical depth

The Eddington characteristic luminosity was derived earlier under the assumption of a radiation flux passing through an optically thin medium surrounding the radiation source. What changes if the radiation passes through an optically thick medium, such as a stellar interior? At high optical depth, the radiation field can be assumed to be nearly isotropic, and the *diffusion approximation* applies (cf. Rybicki and Lightman, 1979). The radiative heat flux can then be written in terms of the radiation pressure $P_{\rm r}$ as

$$F_{\rm r} = c\frac{{\rm d}P_{\rm r}}{{\rm d}\tau}, \tag{1.16}$$

where τ is the optical depth

$${\rm d}\tau = \kappa\rho\,{\rm d}s \tag{1.17}$$

along a path s, and κ an appropriate frequency-averaged opacity (such as the Rosseland mean). Balancing the gradient of the radiation pressure against the force of gravity gives the maximum radiation pressure that can be supported:

$$\nabla P_{\rm r,max} = \mathbf{g}\rho, \tag{1.18}$$

where $\mathbf{g}$ is the acceleration of gravity. With eq. 1.16, this yields the maximum radiation flux at a given point in a static gravitating object:

$$\mathbf{F}_{\rm r,max} = \frac{c\,\mathbf{g}}{\kappa}, \tag{1.19}$$

that is, the same as the critical flux in the optically thin case.

Limitations of the Eddington limit

The derivation of F_{E} assumed that the force relevant in the argument is gravity. Other forces can be larger. An example would be a neutron star with a strong magnetic field. The curvature force $B^2/(4\pi r_{\mathrm{c}})$ in a loop of magnetic field (where r_{c} is the radius of curvature of the field lines) can balance a pressure gradient $\sim P/r_{\mathrm{c}}$, so the maximum pressure that can be contained in a magnetic field[1] is of order $P \sim B^2/8\pi$. If the pressure is due to radiation, assuming an optical depth $\tau \geq 1$ so the diffusion approximation can be used, the maximum radiative energy flux is then of the order

$$F_{\mathrm{r,max}} \approx \frac{cP_{\mathrm{r}}}{\tau} \approx c\frac{B^2}{8\pi\tau}. \tag{1.20}$$

In the range of validity of the assumptions made this has its maximum for an optical depth of order unity: $F_{\mathrm{r,max}} \approx cB^2/8\pi$. For a neutron star of radius $R = 10^6$ cm and a field strength of 10^{12} G, this gives $L_{\mathrm{r,max}} \approx 10^{46}$ erg s^{-1}, many orders of magnitude higher than the Eddington value L_{E}. (This explains the enormous luminosities that can be reached in so-called magnetar outbursts; Hurley *et al.*, 2005.)

The Eddington argument considers only the radiative flux. Larger energy fluxes are possible if energy is transported by other means, for example by convection.

Since L_{E} depends on opacity, it can happen that L_{E} is lower in the atmosphere of a star than in its interior. A luminous star radiating near its (internal) Eddington rate will then blow off its atmosphere in a *radiatively driven stellar wind*; this happens, for example, in Wolf-Rayet stars. In the context of protostellar accretion, the opacity in the star-forming cloud from which the protostar accretes is high due to atomic and molecular transitions. As a result, the radiation pressure from a massive (proto-)star, with a luminosity approaching equation 1.13, is able to clear away the molecular cloud from which it formed. This is believed to set a limit on the mass that can be reached by a star formed in a molecular cloud.

Neutron stars versus black hole accretors

In deriving the critical accretion rate, it was assumed that the gravitational energy liberated is emitted in the form of radiation. In the case of a black hole accretor, this is not necessary, since mass can flow through the hole's horizon, taking with it all energy contained in it. Instead of being emitted as radiation, the energy adds to the mass of the hole. This becomes especially important at high accretion rates, $\dot{M} > \dot{M}_{\mathrm{E}}$ (and in the ion-supported accretion flows discussed in Section 1.13). The parts of the flow close to the hole then become optically thick; the radiation stays trapped in the flow and, instead of producing luminosity, is swallowed by the hole. The accretion rate on a black hole can thus be arbitrarily large in principle (see also Section 1.11, and chapter 10 in Frank *et al.*, 2002).

A neutron star cannot absorb this much energy (only a negligible amount is taken up by conduction of heat into its interior), so $\dot{M}_{\mathrm{E}}$ is more relevant for neutron stars than for black holes. It is not clear to what extent it actually limits the possible accretion rate, however, since the limit was derived under the assumption of spherical symmetry. It is possible that accretion takes place in the form of an optically thick disk, while the energy released at the surface produces an outflow along the axis, increasing the maximum possible accretion rate (cf. discussion in Section 1.11). This has been proposed (e.g., King, 2004) as a possible conservative interpretation of the so-called ultraluminous X-ray sources (ULX), rare objects with luminosities of 10^{39} to 10^{41} erg s^{-1}. These are alternatively and more excitingly suggested to harbor intermediate mass black holes (above ≈ 30 M$_\odot$).

[1] Depending on circumstances the actual maximum is less than this because a magnetically contained plasma tends to "leak across" field lines through MHD instabilities.

1.3 Accretion with angular momentum

When the accreting gas has a nonzero angular momentum with respect to the accreting object, it cannot accrete directly. A new time scale then plays a role: the time scale for outward transport of angular momentum. Because this is in general much longer than the dynamical time scale, much of what was said about spherical accretion needs modification for accretion with angular momentum.

Consider the accretion in a close binary consisting of a compact (white dwarf, neutron star, or black hole) primary of mass M_1 and a main sequence companion of mass M_2. The mass ratio is defined as $q = M_2/M_1$ (note: in the literature, q is just as often defined the other way around).

If M_1 and M_2 orbit each other in a circular orbit and their separation is a, the orbital frequency Ω is

$$\Omega^2 = \frac{G(M_1 + M_2)}{a^3}. \tag{1.21}$$

The accretion process is most easily described in a coordinate frame that corotates with this orbit, and with its origin in the center of mass. Matter that is stationary in this frame experiences an effective potential, the *Roche potential* (chapter 4 in Frank *et al.*, 2002), given by

$$\phi_{\rm R}(\mathbf{r}) = -\frac{GM}{r_1} - \frac{GM}{r_2} - \frac{1}{2}\Omega^2\varpi^2, \tag{1.22}$$

where $r_{1,2}$ are the distances of point $\mathbf{r}$ to stars 1, 2, and ϖ is the distance from the rotation axis (the axis through the center of mass, perpendicular to the orbit). Matter that does *not* corotate experiences a very different force (due to the Coriolis force). The Roche potential is therefore useful only in a rather limited sense. For non-corotating gas, intuition based on the Roche geometry can be misleading. Keeping in mind this limitation, consider the equipotential surfaces of equation 1.22. The surfaces of stars $M_{1,2}$, assumed to corotate with the orbit, are equipotential surfaces of equation 1.22. Near the centers of mass (at low values of $\phi_{\rm R}$), they are unaffected by the other star; at higher Φ, they are distorted; and at a critical value Φ_1, the two parts of the surface touch. This is the critical Roche surface S_1 whose two parts are called the *Roche lobes.*

Binaries lose angular momentum through gravitational radiation and a magnetic wind from the secondary (if it has a convective envelope). Through this loss, the separation between the components decreases and both Roche lobes decrease in size. Mass transfer starts when M_2 fills its Roche lobe and continues as long as the angular momentum loss from the system lasts. Mass transfer can also be due to expansion of the secondary in the course of its evolution, and mass transfer can be a runaway process, depending on mass and internal structure of the secondary. This is a classical subject in the theory of binary stars; for an introduction, see Warner (1995).

A stream of gas then flows through the point of contact of the two parts of S_1, the inner Lagrange point L_1. If the force acting on it were derivable entirely from equation 1.22, the gas would just fall in radially onto M_1. As soon as it moves, however, it no longer corotates, and its orbit under the influence of the Coriolis force is different (Fig. 1.1).

Since the gas at L_1 is usually very cold compared with the virial temperature, the velocity it acquires already exceeds the sound speed after moving a small distance from L_1. The flow into the Roche lobe of M_1 is therefore highly *supersonic*. Such hypersonic flow is essentially ballistic, that is, the stream flows approximately along the path taken by freely falling particles.

Though the gas stream on the whole follows a path close to that of a free particle, a strong shock develops at the point where the path intersects itself.[2] After this, the gas

[2]In practice, shocks develop shortly after passing the pericenter at M_1, when the gas is decelerated again. Supersonic flows that are decelerated, by whatever means, in general develop

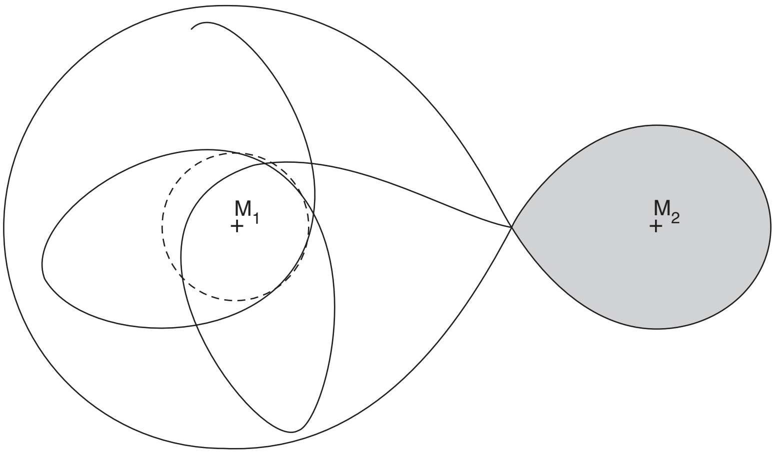

FIGURE 1.1. Roche geometry for $q = 0.2$, with free particle orbit from L_1 (as seen in a frame corotating with the orbit). Dashed: circularization radius.

settles into a ring, into which the stream continues to feed mass. If the mass ratio q is not too small, this ring forms fairly close to M_1. An approximate value for its radius is found by noting that near M_1, the tidal force due to the secondary is small, so that the angular momentum of the gas with respect to M_1 is approximately conserved. If the gas continues to conserve angular momentum while dissipating energy, it settles into the minimum-energy orbit with the specific angular momentum j of the incoming stream. The value of j is found by a simple integration of the orbit starting at L_1 and measuring j at some point near the pericenter. The radius of the orbit, the nominal *circularization radius*, r_c, is then defined through $(GM_1 r_c)^{1/2} = j$. In units of the orbital separation a, r_c and the distance r_{L1} from M_1 to L_1 are functions of the mass ratio only. As an example, for $q = 0.2$, $r_{L1} \approx 0.66a$ and the circularization radius $r_c \approx 0.16a$. In practice, the ring forms somewhat outside r_c, because there is some angular momentum redistribution in the shocks that form at the impact of the stream on the ring. The evolution of the ring depends critically on nature and strength of the angular momentum transport processes. If sufficient "viscosity" is present, it spreads inward and outward to form a disk.

At the point of impact of the stream on the disk, the energy dissipated is a significant fraction of the orbital kinetic energy; hence, the gas heats up to a significant fraction of the virial temperature. For a typical system with $M_1 = 1\ \mathrm{M}_\odot$, $M_2 = 0.2\ \mathrm{M}_\odot$, and an orbital period of 2 hours, the observed size of the disk (e.g., Wood *et al.*, 1989b; Rutten *et al.*, 1992) $r_d/a \approx 0.3$, the orbital velocity at r_d is about 900 km s^{-1}, and the virial temperature at r_d is $\approx 10^8$ K. The actual temperatures at the impact point are much lower, due to rapid cooling of the shocked gas. Nevertheless, the impact gives rise to a prominent "hot spot" in many systems, and an overall heating of the outermost part of the disk.

1.4 Thin disks: properties

1.4.1 *Flow in a cool disk is supersonic*

Ignoring viscosity, the equation of motion in the potential of a point mass is

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \cdot \nabla \mathbf{v} = -\frac{1}{\rho} \nabla P - \frac{GM}{r^2} \hat{\mathbf{r}}, \tag{1.23}$$

shocks (e.g., Courant and Friedrichs, 1948; Massey, 1968). The effect can be seen in action in the video published in Różyczka and Spruit (1993).

where $\hat{\mathbf{r}}$ is a unit vector in the spherical radial direction r. To compare the order of magnitude of the terms, choose a position r_0 in the disk and choose as typical time and velocity scales the orbital time scale $\Omega_0^{-1} = (r_0^3/GM)^{1/2}$ and velocity $\Omega_0 r_0$. For simplicity of the argument, assume a fixed temperature T. The pressure gradient term is then

$$\frac{1}{\rho}\nabla P = \frac{\mathcal{R}}{\mu} T \nabla \ln P. \tag{1.24}$$

In terms of the dimensionless quantities

$$\tilde{r} = \frac{r}{r_0}, \qquad \tilde{v} = \frac{v}{(\Omega_0 r_0)}, \tag{1.25}$$

$$\tilde{t} = \Omega_0 t, \qquad \tilde{\nabla} = r_0 \nabla, \tag{1.26}$$

the equation of motion becomes

$$\frac{\partial \tilde{\mathbf{v}}}{\partial \tilde{t}} + \tilde{\mathbf{v}} \cdot \tilde{\nabla} \tilde{\mathbf{v}} = -\frac{T}{T_{\rm vir}} \tilde{\nabla} \ln P - \frac{1}{\tilde{r}^2} \hat{\mathbf{r}}. \tag{1.27}$$

All terms and quantities in this equation are of order unity by the assumptions made, except the pressure gradient term, which has the coefficient $T/T_{\rm vir}$ in front. If cooling is important, so that $T/T_{\rm vir} \ll 1$, the pressure term is negligible to first approximation. Equivalent statements are also that the gas moves hypersonically on nearly Keplerian orbits, and that the disk is thin, as is shown next.

1.4.2 *Disk thickness*

The thickness of the disk is found by considering its equilibrium in the direction perpendicular to the disk plane. In an axisymmetric disk, using cylindrical coordinates (ϖ, ϕ, z), consider the forces at a point $\mathbf{r}_0$ $(\varpi, \phi, 0)$ in the midplane, in a frame rotating with the Kepler rate Ω_0 at that point. The gravitational acceleration $-GM/r^2\hat{\mathbf{r}}$ balances the centrifugal acceleration $\Omega_0^2 \hat{\boldsymbol{\varpi}}$ at this point, but not at some distance z above it because gravity and centrifugal acceleration work in different directions. Expanding both accelerations near $\mathbf{r}_0$, one finds a residual acceleration toward the midplane of magnitude

$$g_z = -\Omega_0^2 z. \tag{1.28}$$

Assuming an isothermal gas at temperature T, the condition for equilibrium in the z-direction under this acceleration yields a hydrostatic density distribution

$$\rho = \rho_0(\varpi) \exp\left(-\frac{z^2}{2H^2}\right). \tag{1.29}$$

$H(\varpi)$, called the *scale height* of the disk or simply the disk thickness, is given in terms of the isothermal sound speed $c_{\rm i} = (\mathcal{R}T/\mu)^{1/2}$ by

$$H = \frac{c_{\rm i}}{\Omega_0}. \tag{1.30}$$

We define $\delta \equiv H/r$, the *aspect ratio* of the disk; it can be expressed in several equivalent ways:

$$\delta = \frac{H}{r} = \frac{c_{\rm i}}{\Omega r} = \frac{1}{\rm Ma} = \left(\frac{T}{T_{\rm vir}}\right)^{1/2}, \tag{1.31}$$

where Ma is the Mach number of the orbital motion.

1.4.3 *Viscous spreading*

The shear flow between neighboring Kepler orbits in the disk causes friction if some form of viscosity is present. The frictional torque is equivalent to exchange of angular momentum between these orbits. But because the orbits are close to Keplerian, a change in angular momentum of a ring of gas also means it must change its distance from the central mass. If the angular momentum is increased, the ring moves to a larger radius. In a thin disk, angular momentum transport (more precisely, a nonzero divergence of the angular momentum flux) therefore automatically implies redistribution of mass in the disk.

A simple example (Lüst, 1952; Lynden-Bell and Pringle, 1974) is a narrow ring of gas at some distance r_0. If at $t = 0$ this ring is released to evolve under the viscous torques, one finds that it first spreads into an asymmetric hump. The hump quickly spreads inward onto the central object while a long tail spreads slowly outward to large distances. As $t \to \infty$, almost all the *mass* of the ring accretes onto the center, while a vanishingly small fraction of the gas carries almost all the *angular momentum* to infinity.

As a result of this asymmetric behavior, essentially all the mass of a disk can accrete even if the total angular momentum of the disk is conserved. In practice, however, there is often an external torque removing angular momentum from the disk: when the disk results from mass transfer in a binary system. The tidal forces exerted by the companion star take up angular momentum from the outer parts of the disk, limiting its outward spread.

1.4.4 *Observational evidence of disk viscosity*

Evidence for the strength of the angular momentum transport processes in disks comes from observations of variability time scales. This evidence is not good enough to determine whether the processes really behave in the same way as viscosity, but if this is assumed, estimates can be made of its magnitude.

Observations of cataclysmic variables (CVs) give the most detailed information. These are binaries with white dwarf (WD) primaries and (usually) main sequence companions (for reviews, see Meyer-Hofmeister and Ritter, 1993; Warner, 1995). A subclass of these systems, the dwarf novae, show semiregular outbursts. In the currently most developed theory, these outbursts are due to an instability in the disk (Smak, 1971; Meyer and Meyer-Hofmeister, 1981; King, 1995; Hameury *et al.*, 1998). The outbursts are episodes of enhanced accretion of mass from the disk onto the primary, involving a significant part of the whole disk. The decay time of the burst is thus a measure of the viscous time scale of the disk (the quantitative details depend on the model; see Cannizzo *et al.*, 1988; Hameury *et al.*, 1998):

$$t_{\rm visc} = \frac{r_{\rm d}^2}{\nu}, \tag{1.32}$$

where $r_{\rm d}$ is the size of the disk, $\sim 10^{10}$ cm for a CV. With decay times on the order of days, this yields viscosities of the order 10^{15} cm^2 s^{-1}, some 14 orders of magnitude above the microscopic "molecular" viscosity of the gas.

Other evidence comes from the inferred time scale on which disks around protostars disappear, which is of the order of 10^7 years (e.g., Strom *et al.*, 1993).

1.4.5 α-*Parameterization*

Several processes have been proposed to account for these short time scales and the large apparent viscosities inferred from them. One of these is that accretion does not in fact take place through a viscous-like process as described earlier at all, but results from angular momentum loss through a magnetic wind driven from the disk surface (Bisnovatyi-Kogan and Ruzmaikin, 1976; Blandford, 1976), much in the way sunlike stars spin down by angular momentum loss through their stellar winds. The extent to which this plays a role in accretion disks is still uncertain. It would be a "quiet" form of accretion, since it can do without energy-dissipating processes such as viscosity. It has been proposed as the explanation for the low ratio of X-ray luminosity to jet power in many radio sources (see Migliari and Fender, 2006, and references therein). This low ratio, however, is also plausibly attributed to the low efficiency with which X-rays are produced in the "ion-supported accretion flows" (see Section 1.13) that are expected to be the source of the jet outflows observed from X-ray binaries and AGN.

The most quantitatively developed mechanism for angular momentum transport is a form of magnetic viscosity (anticipated already in Shakura and Sunyaev, 1973). This is discussed below in Section 1.8.1. It requires the accreting plasma to be sufficiently electrically conducting. This is often the case, but not always: it is questionable for example in the cool outer parts of protostellar disks. Other mechanisms thus still play a role in the discussion, for example spiral shocks (Spruit *et al.*, 1987) and self-gravitating instabilities (Paczyński, 1978; Gammie, 1997).

In order to compare the viscosities of disks in objects with (widely) different sizes and physical conditions, introduce a dimensionless viscosity α:

$$\nu = \alpha \frac{c_{\rm i}^2}{\Omega}, \tag{1.33}$$

where $c_{\rm i}$ is the isothermal sound speed as before. The quantity α was introduced by Shakura and Sunyaev (1973) as a way of parametrizing our ignorance of the angular momentum transport process in a way that allows comparison between systems of very different size and physical origin. (Their definition of α differs a bit, by a constant factor of order unity.)

1.4.6 *Causality limit on turbulent viscosity*

How large can the value of α be, on theoretical grounds? As a simple model, let's assume that the shear flow between Kepler orbits is unstable to the same kind of shear instabilities found for flows in tubes, in channels, near walls, and in jets. These instabilities occur so ubiquitously that the fluid mechanics community considers them an automatic consequence of a high Reynolds number:

$$\mathrm{Re} = \frac{LV}{\nu} \tag{1.34}$$

where L and V are characteristic length and velocity scales of the flow. If this number exceeds about 1,000 (for some forms of instability much less), instability and turbulence are generally observed. It has been argued (e.g., Zel'dovich, 1981) that for this reason hydrodynamic turbulence is the cause of disk viscosity. Let's look at the consequences of this assumption. If an eddy of radial length scale l develops because of shear instability, it will rotate at a rate given by the rate of shear in the flow, σ, in a Keplerian disk:

$$\sigma \approx r\frac{\partial \Omega}{\partial r} = -\frac{3}{2}\Omega. \tag{1.35}$$

The velocity amplitude of the eddy is $V = \sigma l$, and a field of such eddies would produce a turbulent viscosity of the order (leaving out numerical factors of order unity)

$$\nu_{\rm turb} \approx l^2 \Omega. \tag{1.36}$$

In compressible flows, there is a maximum to the size of the eddy set by causality considerations. The force that allows an instability to form a coherently overturning eddy is the pressure, which transports information about the flow at the speed of sound. The eddies formed by a shear instability can therefore not rotate faster than $c_{\rm i}$; hence, their size does not exceed $c_{\rm i}/\sigma \approx H$ (eq. 1.31). At the same time, the largest eddies formed also have the largest contribution to the exchange of angular momentum. Thus we should expect that the turbulent viscosity is given by eddies with size of the order H:

$$\nu < H^2 \Omega, \tag{1.37}$$