The dynamics of perturbation modes in protoplanetary discs - new effects of self-gravity and velocity shear

---

George Mamatsashvili

Institute for Astronomy

School of Physics and Astronomy

Doctor of Philosophy

The University of Edinburgh

January 2011

Abstract

Protoplanetary discs, composed of gas and dust, usually surround young stellar objects and serve two main purposes: they determine the accretion of matter onto the central object and also represent sites of planet formation. The accretion proceeds through the transport of angular momentum outwards allowing the disc matter to fall towards the centre. A mechanism responsible for the transport can be turbulence, waves or other coherent structures originating from various instabilities in discs that could, in addition, play a role in the planet formation process. For an understanding of these instabilities, it is necessary to study perturbation dynamics in differentially rotating, or sheared media. Thus, this thesis focuses on new aspects in the perturbation dynamics in non-magnetised protoplanetary discs that arise due to their self-gravity and velocity shear associated with the disc's differential rotation. The analysis is carried out in the framework of the widely employed local shearing box approximation. We start with 2D discs and then move on to 3D ones.

In 2D discs, there are two basic perturbation types/modes – spiral density waves and vortices – that are responsible for angular momentum transport and that can also contribute to accelerating planet formation. First, in the linear regime, we demonstrate that the vortical mode undergoes large growth due to self-gravity and in this process generates density waves via shear-induced linear mode coupling phenomenon. This is noteworthy, because commonly only density waves are considered in self-gravitating discs. Then we investigate vortex dynamics in the non-linear regime under the influence of self-gravity by means of numerical simulations. It is shown that vortices are no longer well-organised and long-lived structures, unlike those occurring in non-self-gravitating discs. They undergo recurring phases (lasting for a few disc rotation periods) of formation, growth and eventual destruction. We also discuss the dust trapping capability of such transient vortices.

Perturbation dynamics in 3D vertically stratified discs is richer, as there are more mode types. We first consider non-axisymmetric modes in non-self-gravitating discs and then only axisymmetric modes in the more complicated case when self-gravity is present. Specifically, in non-self-gravitating discs with superadiabatic vertical stratification, motivated by the recent results on the transport properties of incompressible convection, we show that when compressibility is taken into account, the non-axisymmetric convective mode excites density waves via the same shear-induced linear mode coupling mechanism mentioned above. These generated density waves transport angular momentum outwards in the trailing phase, and we suggest that they may aid and enhance the transport due solely to convection in the non-linear regime, where the latter becomes outward.

In the final part of the thesis, we carry out a linear analysis of axisymmetric vertical normal modes in stratified self-gravitating discs. Although axisymmetric modes do not display shear-induced couplings, their analysis provides insight into how gravitational instabilities develop in the 3D case and their onset criterion. We examine how the structure of dispersion curves and eigenfunctions of 3D modes are influenced by self-gravity, which mode first becomes gravitationally unstable and thus determines the onset criterion and nature of the gravitational instability in stratified discs. We also contrast the more exact instability criterion obtained with our 3D model with that of density waves in 2D discs. Based on these findings, we discuss the origin of 3D behaviour of perturbations involving noticeable disc surface distortions, as seen in some numerical simulations of self-gravitating discs.

Declaration

I declare that this thesis is not substantially the same as any that I have submitted for a degree or diploma or other qualification at any other University. I further state that no part of my thesis has already been or is being concurrently submitted for any such degree, diploma or other qualification.

This thesis is the outcome of my own work except where specifically indicated in the text.

George Mamatsashvili

Edinburgh,
6 January 2011

Acknowledgements

I would like to thank my principal supervisor Ken Rice, who has acted as my teacher, my guide in the theory of protoplanetary discs and planet formation, my main advisor and a person to rely on in my postgraduate research activity and simply a good friend. He made my stay at the Institute for Astronomy (IfA) highly enjoyable and productive. During these three years, I have learnt many valuable scientific lessons from Ken, which will no doubt be very helpful throughout my future career as a researcher, and I am extremely grateful for that.

I would also like to thank my second supervisor Avery Meiksin for advice and help and other staff members for maintaining a friendly and scientifically stimulating atmosphere at the IfA.

I greatly appreciate many useful and stimulating discussions with my colleagues outside the IfA. Particularly, I wish to express my thanks to Charles Gammie for kindly providing a version of the ZEUS code used in this work and for elucidating many aspects of numerical simulations of vortices and waves in astrophysical discs. I thank Gordon Ogilvie for illuminating discussions regarding the mode dynamics in discs (on the material in chapter 5) and constructive comments on the subsequent paper. Also big thanks to my georgian collaborators, George Chagelishvili – my MSc. supervisor, who first introduced me into the broad and fascinating field of astrophysical discs, and Alexander Tevzadze, who greatly helped me to understand velocity shear-induced phenomena in discs.

Special thanks go to my examiners, Gordon Ogilvie and James Dunlop, who spent time patiently and critically reading the thesis and gave very useful comments and suggestions.

I am grateful to Graduate School Secretary, Jane Patterson, and IfA Teaching and General Support Secretary, Paula Wilkie, for helping many times with making various formal documents as part of my postgraduate studies.

I would also like to thank my fellow students, Neil Phillips and David Sobral, who have always been willing to provide assistance with programming and computer matters and much more, as well as postdocs at the IfA, Eric Tittley, Wing-Fai Thi and Peder Norberg, for many useful discussions scientific or non-scientific, interesting conversations and good times.

I thank my family and friends, who have given me continual support and encouragement during the whole period of my postgraduate studies.

Finally, I would like to gratefully acknowledge the generous financial support granted to me by the Scottish Universities Physics Alliance (SUPA) as a honourable prize studentship, without which my postgraduate studies at the University of Edinburgh, all my research activity and hence the present thesis would not be possible.

Contents

List of Figures

Chapter 1

Overview of protoplanetary discs

1.1 Disc formation

The formation of stars and planetary systems begins with the gravitational collapse of dense cores in a turbulent molecular cloud. This very first phase is called the collapse phase when most of the cloud material is initially concentrated in dense regions, or cores, and is starting to contract under its own gravity.[1] Such cores are mainly composed of H_2 gas with number densities $\sim 10^5$ cm^{-3}, sizes ~ 0.1 pc, masses of a few solar mass $M_\odot$ and temperatures ~ 10 K (Shu et al., 1987; Bodenheimer, 1995). Usually they rotate uniformly with angular velocities $\sim 10^{-13}$ s^{-1} (Goodman et al., 1993), so the corresponding angular momentum of the core turns out to be about three orders of magnitude larger than that of a single star, even if rotating at breakup speed. This implies that some loss or redistribution of angular momentum must take place if a star is to form. In other words, the core should somehow get rid of excess angular momentum and, for this reason, its collapse proceeds through the formation of a rotating disc that takes up the angular momentum, instead of matter directly falling into the centre (e.g., Ulrich, 1976; Cassen and Moosman, 1981; Terebey et al., 1984; Shu et al., 1993; Yorke et al., 1993; Yorke and Bodenheimer, 1999).[2] As a result, a system of a protostar (or stellar core) surrounded by a so-called protostellar disc forms, which is embedded

---

[1] A core at this early stage of collapse is classified as a Class 0 object (see classification system below).

[2] More complex situations involve the formation of binary or multiple stellar systems, which are not considered here; we describe only the formation of a single low-mass ($\lesssim 2.5 M_\odot$) star.

in an optically thick envelope of gas and dust. Such a system is commonly classified as a Class I object. The material falling from the envelope is channeled through the disc onto the protostar, which at the beginning has a very low mass ($\sim 0.01 M_\odot$) and thus builds up by accretion of matter from the disc. This accretion process is driven by various instabilities operating in the disc, which by exerting torques, transport angular momentum outwards, thereby allowing the disc matter to fall towards the centre (see chapter 2). Specifically, at this early collapse stage, the disc is massive with a mass comparable to the mass of the protostar. As a consequence, self-gravity (gravitational instability) of the disc plays an important dynamical role during this first phase of the accretion process and is likely to be the chief cause of angular momentum and mass transport in young embedded discs (e.g., Laughlin and Bodenheimer, 1994; Osorio et al., 2003; Piétu et al., 2005; Vorobyov and Basu, 2006; Lodato, 2007; Kratter et al., 2008, see also section 2.2.3 in chapter 2). The resulting mass accretion rate within the disc is highest at this infall stage reaching $10^{-6} - 10^{-5}\ M_\odot/\mathrm{yr}$, but quickly drops to $10^{-7} - 10^{-9}\ M_\odot/\mathrm{yr}$ at later times once the infall phase has ended (Fig. 1.1). However, this average disc accretion rate during the embedded phase is more than an order of magnitude lower than the envelope infall rate. Equivalently, Class I protostars are, on average, about an order of magnitude less luminous than expected from the steady release of gravitational energy as the envelope falls onto the protostar over the lifetime of the embedded phase (Kenyon et al., 1990; Eisner et al., 2005). This implies that mass transport across the disc due to its self-gravity proceeds in an episodic manner, periodically reaching accretion rates as high as $10^{-4}\ M_\odot/\mathrm{yr}$ during relatively short intervals of $1 - 10$ yr and then decaying on the time-scale of 100 yr in the course of evolution (FU Orionis-type outbursts, Vorobyov and Basu, 2006; Zhu et al., 2009; Lodato, 2007; Dunham et al., 2010).

After a few 10^5 yr, the infall phase is essentially over, the surrounding envelope is almost completely depleted, revealing the protostar-disc system. At this point, the protostar starts the second, pre-main-sequence quasi-static gravitational contraction stage of evolution, at which it is often called a T Tauri star and is classified as a Class II object. Now it can be considered as an almost fully formed star undergoing accretion only from the remaining disc of gas and dust as well as perhaps from some tenuous amounts of envelope material too. Mass accretion rates in this later phase, as mentioned above, are smaller, of the order of $10^{-7}\ M_\odot/\mathrm{yr}$ (e.g., Gullbring et al., 1998), than those during the earlier infall phase and fall off with time approximately as $t^{-1.5}$ (Hartmann et al., 1998). However, the above-mentioned episodic rises and drops in accretion rate are sometimes also observed at the early T Tauri stage (Hartmann and Kenyon, 1996), possibly because of mass piling up and following gravitational instability

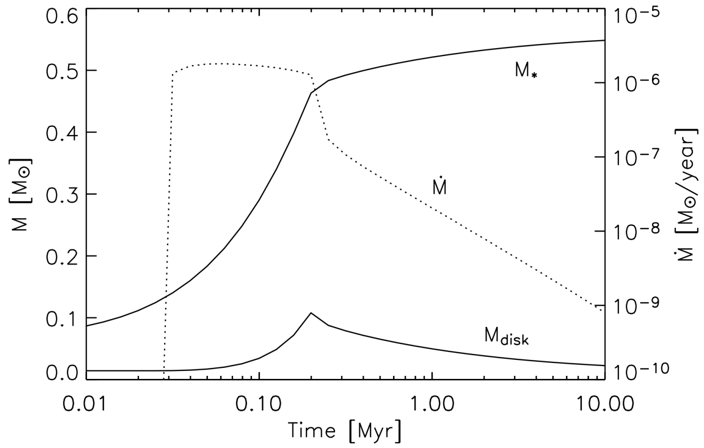

Figure 1.1: Evolution of disc mass (lower solid line), mass accretion rate onto the protostar (dotted line) and protostar mass itself (upper solid line) with time after the onset of cloud core collapse (adapted from Hueso and Guillot, 2005). The jump in the accretion rate curve marks the disc formation. The collapse phase is finished by 2×10^5 yr. From this point, the accretion rate and disc mass are gradually decaying, because the infalling envelope has already been depleted and there is no other supply of mass for the disc. After a few million years the disc is almost completely dispersed and the protostar becomes a normal star beginning its main-sequence phase.

in a magnetically inactive ('dead') zone within the disc (see chapter 2). By that time, the disc mass has been reduced as well and constitutes a few percent of the central stellar mass (see the next section). Due to this, gravitational instability in the disc, though possibly still playing a certain role, is no longer the main driver of mass accretion in the T Tauri phase, so that these low accretion rates are primarily attributable to other mechanisms (instabilities) responsible for transport in this regime. Because of the absence of optically thick envelopes, discs around Class II young stellar objects are more easily accessible to observations mostly in optical, infrared and (sub-)millimeter bands that largely determined their discovery in the mid-1980s (e.g., Beckwith et al., 1986; Hartmann et al., 1986; Adams et al., 1987; Sargent and Beckwith, 1987; Strom et al., 1989; Beckwith et al., 1990; Dutrey et al., 1994; Padgett et al., 1999). Such discs – a natural evolutionary progression of earlier protostellar discs – are referred to as 'protoplanetary' (Fig. 1.2), since not only they provide mass for accretion and subsequent evolution of the central star, but also represent primary arenas for dynamical processes leading to planet formation. In the next section, we describe the basic physical properties of these discs, as they are the central subject of the present thesis.

We should also mention that protoplanetary (i.e., T Tauri) discs are often described within the framework of a prototypical disc model – the Minimum Mass Solar Nebula (MMSN,

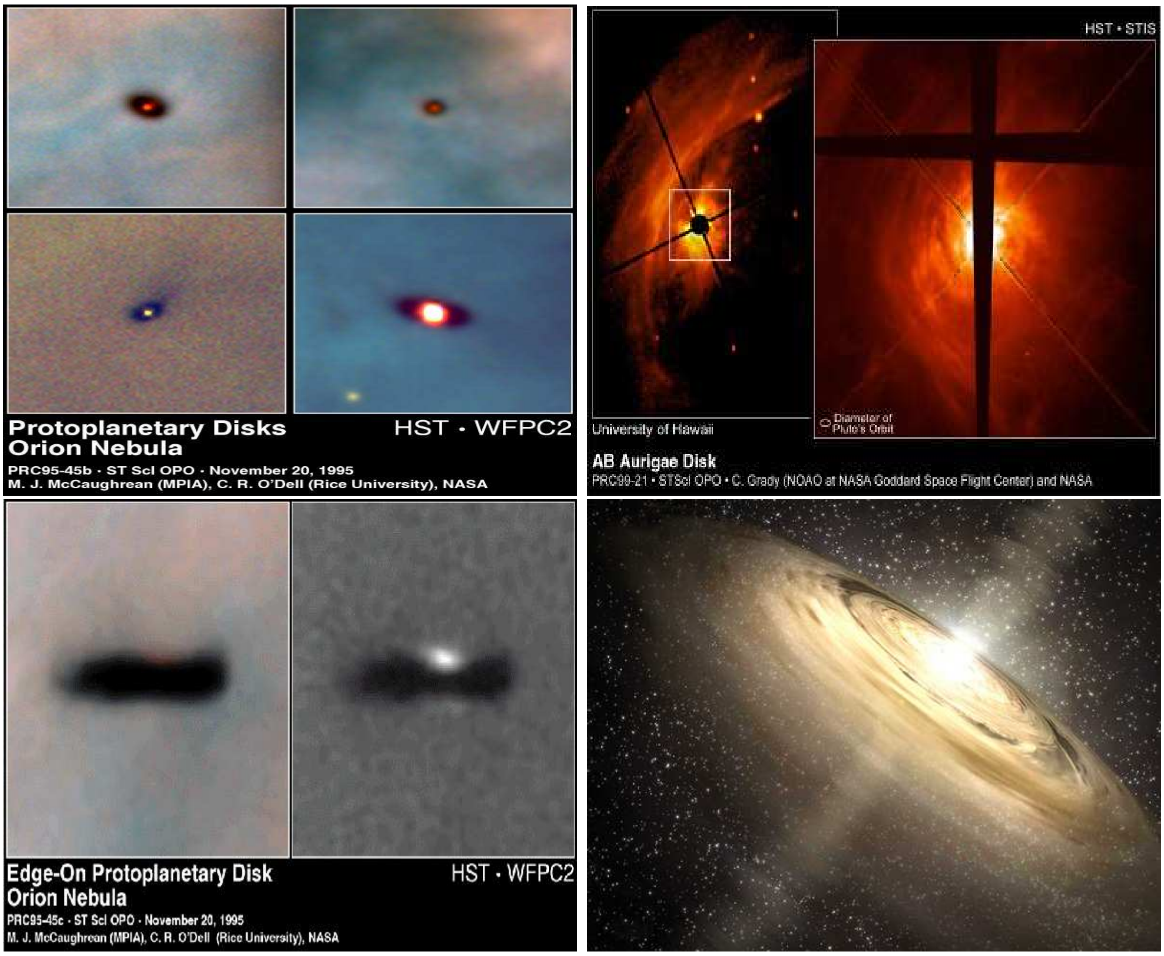

Figure 1.2: Left upper and lower panels show the face-on and edge-on images (silhouettes) of protoplanetary discs around young stars in the Orion nebula taken with the Hubble Space Telescope (HST). Upper right panel shows the face-on images of the AB Aurigae disc taken with HST. Lower right plot is an artist's impression of a protoplanetary disc. (All these plots are courtesy of NASA.)

Weidenschilling, 1977; Hayashi, 1981). The MMSN is constructed by augmenting the masses of planets in the Solar System with enough hydrogen and helium to get solar composition. Then, the Solar System is divided into radial annuli, with one planet per annulus and the augmented mass for each planet is distributed uniformly across its corresponding annulus to obtain a surface density profile, which turns out to be $\Sigma \propto r^{-3/2}$. The resulting gaseous disc has the total mass of $0.01M_{\odot}$ (from Venus to Neptune), which is actually the minimum mass required to build the Solar System's planets and appears comparable to the typical masses of protoplanetary discs around solar mass stars as inferred from observations (Andrews and Williams, 2005, see also the next section). However, we should keep in mind that, though useful in theoretical analyses, the MMSN is a hypothetical disc model and more realistic theoretical disc models as well as observations tend to indicate shallower density profiles similar to $\Sigma \propto r^{-1}$ (e.g., Bell et al., 1997; Wilner et al., 2000; Kitamura et al., 2002, see also the next section).

Finally, after about 10^7 yr the disc gas gradually disperses during $10^4 - 10^5$ yr (Simon and Prato, 1995; Wolk and Walter, 1996) leaving a remnant 'debris disc' mostly composed of rocks and pebbles, leftover from the planet formation process, which has nearly finished by this time. Thus, the typical lifetimes of protoplanetary discs are a few million years, with a maximum of 10 Myr (Haisch et al., 2001; Sicilia-Aguilar et al., 2006). At this stage, the star becomes a Class III object, in which the characteristic features of accretion – excess infrared and (sub-)millimeter continuum emission from cold dust in the outer disc – are no longer distinctly noticeable. The main mechanism causing disc dispersal is photoevaporation – escaping of gas from the disc surface into space due to heating by the central star's ultraviolet radiation (see e.g., Dullemond et al., 2007). Figure 1.1 illustrates the evolution of disc mass, accretion rate, and star mass through all the evolutionary stages described above.

At the end of this section, we outline the classification system of low-mass stars mentioned at the beginning of this chapter. The radiation produced during the main infall phase is reprocessed by the dust in the surrounding envelope and subsequently re-emitted mainly in the infrared and (sub-)millimeter wavelengths (e.g., Adams et al., 1987). Similarly, at the T Tauri stage, dust grains in the disc absorb stellar radiation and then re-emit it in the same wavelength bands (e.g., Adams et al., 1988; Kenyon and Hartmann, 1987, 1995; Chiang and Goldreich, 1997). This re-radiated light is characterised by a spectral energy distribution (SED) and young low-mass stars, or young stellar objects, are usually classified into different types according to the shape (slope) of their SED from near- to mid-infrared ($2 - 50$ μm) bands (Lada and Wilking, 1984; Adams et al., 1987; Lada, 1987; Greene et al., 1994). Let us define the parameter $\alpha_{IR} = d\log(\lambda F_\lambda)/d\log\lambda$, where F_λ is the observed flux at wavelengths λ. This parameter characterises the infrared excess of the SED with respect to stellar radiation spectrum that is associated with the presence of a disc. Depending on the value of α_{IR}, the classification scheme is as follows:

Class 0: Sources at the initial stage of collapse with a central protostar extremely faint (hardly detectable) at optical and near-infrared (< 10 μm) and having a significant sub-millimeter luminosity – the spectrum peaks in the far-infrared (~ 100 μm) or sub-millimeter. Most of the circumstellar material still resides in envelopes although a nascent stellar core is noticeable. Class 0 objects are thought to be precursors of Class I objects.

Class I: Sources of this class have $\alpha_{IR} > 0$ in the wavelengths range from 2.2 μm to $10 - 25$ μm. Such sources, as mentioned above, are evolved protostars with distinct accretion discs and envelopes. The material from the infalling envelope is being accumulated in a disc and then channeled onto the centre.

Class II: Sources of this class have $-1.5 < \alpha_{IR} < 0$ and represent classical T Tauri stars with large protoplanetary discs and with a negligible amount of infalling envelopes. In this case, infrared continuum thermal emission comes primarily from the warm dust in the disc, heated by stellar light perhaps with some contribution through viscously dissipated gravitational energy that is liberated during accretion. However, flat-spectrum (small $|\alpha_{IR}| < 1$, $\alpha_{IR} < 0$) may occur if a disc around a T Tauri star is massive enough to be self-gravitating (Adams et al., 1988; Lodato, 2007). The spectrum in this case is mostly generated by internal viscous dissipation in the disc rather than by dust-reprocessed stellar light.

Class III: Sources of this class have $\alpha_{IR} < -1.5$ and represent pre-main-sequence stars that have already accreted all their disc material and are therefore characterised by little or no excess infrared emission. The emission in this case comes from the remaining hot dust in the debris disc devoid of gas. At this stage, the star already has a planetary system.

These four classes can be viewed as an evolutionary sequence of the formation process of a single low-mass star from beginning to end (Lada and Wilking, 1984; Adams et al., 1987).

1.2 Characteristics of protoplanetary discs

1.2.1 Radius

Estimates of disc radii are obtained via detection at (sub-)millimeter wavelengths of rotational lines of cold (~ 10 K) gas (mostly CO, the primary gaseous species H_2 is difficult to detect) and continuum emission from cold dust (e.g., Beckwith et al., 1990; Dutrey et al., 1996; Andrews and Williams, 2007b). However, it has been found recently that the gas component usually extends much farther out than the dust component, so only the observations of the former can give correct values for a disc's total radial extent (Piétu et al., 2005; Isella et al., 2007; Hughes et al., 2008). The gas surface density profile follows a power-law throughout most of the disc (see Eq. 1.1 below), but decays approximately exponentially near the outer boundary (i.e., has a tapered edge, Hughes et al., 2008), which itself moves slowly outwards because of the disc's viscous (secular) evolution (see e.g., Lynden-Bell and Pringle, 1974; Hartmann et al., 1998; Armitage, 2010, and also chapter 2). So, the determination of the disc radial size is reduced to locating a characteristic radius, r_c, that delineates where the gas surface density profile begins to steepen from a power-law. Recent high-resolution millimeter observations of a sample of T Tauri discs indicate that typically $r_c \simeq 30 - 230$ AU (e.g., Hughes et al., 2008; Andrews and Williams, 2007b; Andrews et al., 2009; Isella et al., 2009; Andrews et al., 2010), with older discs having larger radii (sometimes extending up to several hundreds of AU, see also e.g., Guilloteau and Dutrey, 1998; Kitamura et al., 2002; Isella et al., 2007) due to viscous spreading, while younger discs, including the youngest ones in Class I objects, tend to have smaller radii, $r_c \lesssim 70$ AU, as they have not yet undergone substantial slow viscous evolution (see also Kenyon et al., 1993; Whitney et al., 1997).

1.2.2 Mass

Disc masses are determined from (sub-)millimeter observations, because at these wavelengths the continuum emission from dust is usually optically thin except in the innermost regions where column densities are high (e.g., Beckwith et al., 1990). Therefore, knowing the dust opacity, observed fluxes can be easily converted into dust mass. Then, assuming a similar canonical dust-to-gas mass ratio of 1:100 for discs as it is in the interstellar medium, one obtains a total disc mass. Surveys of discs around solar mass stars in the Taurus-Auriga and Ophiuchus complexes indicate that disc masses, measured in the above manner, are in the range $(10^{-4} - 10^{-1})M_\odot$ with a median disc mass of $0.005M_\odot$ and with higher mass discs being rare (e.g., Beckwith et al., 1990; Natta et al., 2000; Kitamura et al., 2002; Andrews and Williams, 2005; Andrews et al., 2009). With these disc masses, self-gravity is expected to play a minor role for the majority of T Tauri (i.e., for Class II) discs (Isella et al., 2009; Andrews et al., 2010), as opposed to younger discs around Class I objects, and therefore they can be fairly accurately described within the MMSN model. However, in a few cases, discs more massive than $0.1M_\odot$ in the Class II phase have also been reported (Eisner and Carpenter, 2006; Andrews and Williams, 2007a), which are likely to be self-gravitating. Because disc masses at the Class II stage are generally a small fraction of the central stellar mass, their motion (rotation) is expected to be Keplerian, dominated by gravity of the central star. Observations of CO spectral lines indicate that indeed in most cases discs are in Keplerian rotation on radial scales of a few hundred AU (e.g., Guilloteau and Dutrey, 1998; Simon et al., 2000; Dutrey, 2000).

It should be noted, however, that disc mass measurements at millimeter wavelengths are still subject to large uncertainties due to the amount of optically thick emission from the inner disc, to the gas-to-dust mass ratio and most notably to the growth of dust grains to relatively large sizes ($\gtrsim 1$ cm) compared to those in the interstellar medium, causing reduction in opacity. These large grains contribute little to the (sub-)millimeter flux due to their small surface area-to-mass ratio, so that most of their total mass appears effectively hidden from observations. In addition, these measurements are sensitive to the dust content of the disc, that is, to the dust-to-gas ratio and there is no guarantee that in discs it is similar to the interstellar medium value 1:100, which is usually adopted when estimating disc masses. Consequently, (sub-)millimeter observations may have a tendency to underestimate real disc masses (Andrews and Williams, 2005; Hartmann et al., 2006; Andrews and Williams, 2007a). This implies that even in many standard T Tauri discs, gravitational instability might still contribute to a certain degree to total angular momentum transport.

1.2.3 Structure

Surface density

Imaging protoplanetary discs at (sub-)millimeter wavelengths provides not only a measure of their total masses and radii, but also allows us to determine the mass surface density distribution with radius. Until recently, the disc surface density profile was typically characterised by a simple power-law $\Sigma \propto r^{-p}$, with $p = 0 - 1$, deduced from observations of a sample of T Tauri stars (e.g., Wilner et al., 2000; Kitamura et al., 2002; Andrews and Williams, 2007b). From recent observations it has been realised that in fact the surface density distribution, as already mentioned above, is better described by a power-law within most of the disc, which is truncated exponentially at large radii (Hughes et al., 2008; Andrews et al., 2009; Isella et al., 2009; Andrews et al., 2010)

$$\Sigma(r) \propto r^{-n} \mathrm{exp}\left[-\left(\frac{r}{r_c}\right)^{2-n}\right], \tag{1.1}$$

where r_c is the truncation radius being identified as the disc's outer boundary, as defined in section 1.2.1, and n is the new power-law exponent. This form also agrees well with physical models of viscous accretion discs (e.g., Lynden-Bell and Pringle, 1974; Hartmann et al., 1998). By analysing a large sample of T Tauri discs with different modelling techniques, these observations find that the mean $\langle n \rangle = 0.1 - 0.9$. Note that this is shallower than the MMSN density profile $\Sigma \propto r^{-3/2}$. To get an idea of the typical values of surface densities, Andrews et al. (2009, 2010) find surface densities in the range $\Sigma \approx 10 - 100$ g cm^{-2} at about 20 AU in their sample of Ophiuchus discs. (For the sake of comparison, in the MMSN model, $\Sigma \approx 19$ g cm^{-2} at 20 AU, Hayashi 1981.) However, these inferred surface densities are increasingly uncertain closer to the star (at radii $\lesssim 10$ AU) due to the limited resolution of the observations and also because of the emission becoming optically thick. These issues are to be addressed by future high sensitivity and high resolution observations.

Temperature

Disc temperature depends on the heating by irradiation from the central star and to a lesser extent also on the heat generation due to the accretion process (i.e., viscous dissipation). Accretion heating might be important at the early stages of disc evolution, but turns out to be small compared with stellar irradiation at the T Tauri stage except probably in the inner regions (e.g., Calvet et al., 2000). So, in theoretical modelling (see e.g., Dullemond et al., 2007, for a review), viscous dissipation is neglected to a first approximation[3] and temperature distribution is obtained by locally equating the stellar flux impinging onto the disc (that also depends on the disc shape, namely on the degree of flaring) to that re-emitted by dust grains (i.e., radiative cooling). The latter is calculated by solving a radiative transfer problem for dust (infrared) continuum along the disc vertical structure.[4] Then, knowing the temperature distribution, the corresponding SED can be computed and directly compared with observationally inferred SEDs. This, in turn, allows us to more accurately constrain the disc temperature structure. As a result, combining theoretical modelling based on radiative transfer calculations with observed SEDs of T Tauri stars at infrared and (sub-)millimeter wavelengths, it was demonstrated that a typical temperature profile has a power-law form $T \propto r^{-q}$, with q in the range $0.4-0.7$ (e.g., Kenyon and Hartmann, 1987; Beckwith et al., 1990; Calvet et al., 1992; Chiang and Goldreich, 1997; D'Alessio et al., 1998; Wilner et al., 2000; Chiang et al., 2001; Kitamura et al., 2002; Andrews and Williams, 2005). Interferometric CO images also yield a similar radial distribution of temperature (Guilloteau and Dutrey, 1998; Piétu et al., 2007). As a reference value, observationally obtained disc temperatures at 1 AU are usually in the range $100 - 300$ K (e.g., Beckwith et al., 1990; Kitamura et al., 2002).

1.3 Summary

The existence of protoplanetary, or generally circumstellar, discs around young stars was established in the mid-1980s thanks to advances in infrared and millimeter astronomy. It is also this wavelength band through which most of the disc characteristics, such as sizes, masses, surface density and temperature distributions, etc. as well as dust properties are

---

[3] Models including heating due to viscous dissipation (e.g., that of Bell et al., 1997; D'Alessio et al., 1998) usually adopt turbulent viscosity parameterised by a constant Shakura-Sunyaev α coefficient (introduced in chapter 2).

[4] In most theoretical models, gas and dust temperatures are assumed to be equal within most of the disc interior except in the uppermost surface layers, where gas number density is small, below $10^4 - 10^5$ cm^{-3} (see e.g., Dullemond et al., 2007).

inferred and analysed. The interpretation of data is still complicated by various factors, such as uncertainties in our knowledge of grain growth and settling, opacities, gas dispersal, etc. Nevertheless, to date we have a fairly sound understanding of the general picture of disc formation and evolution as well as of the physical processes responsible for these. In this section, we summarise the basic facts about discs, as reviewed in this chapter.

- Discs form as a result of angular momentum conservation after a molecular cloud core starts to collapse. At the early stage, they are massive and therefore evolve/accrete due primarily to self-gravity, exhibiting recurring large bursts of mass accretion rate.
- After about 10^5 yr, when a central young star becomes visible, the mass of the new protoplanetary disc is reduced (less than 10% that of the star). The disc surface density declines with radius as a power-law with an exponentially tapered edge (see Eq. 1.1), which, depending on the disc age, extends from a few tens to hundreds of AU. The disc temperature also decreases with radius as a power-law. The disc evolution in this phase is steadier and quieter than that during the early collapse phase, because self-gravity is no longer a main agent driving evolution and other mechanisms of accretion, such as turbulent viscosities (see the next chapter), are at work. Observations also unanimously demonstrate that discs rotate with a Keplerian velocity profile.
- After $10^6 - 10^7$ yr discs disperse quickly due to a photoevaporation process caused by stellar irradiation. Once the disc has disappeared, it leaves a debris disc of rocks and pebbles – a leftover from the planet formation process within the disc – which has nearly been completed by this time.

With this basic information, we finish the description of observational characteristics of protoplanetary discs. A much fuller account of disc observations can be found in a recent review article by Williams and Cieza (2011).

Chapter 2

Dynamics of protoplanetary discs

The evolution of protoplanetary discs proceeds, as mentioned in chapter 1, through the redistribution of angular momentum allowing accretion of disc matter onto the central star. During this process, the matter in the outer parts of the disc takes up angular momentum from the matter in the inner parts, so the latter losing angular momentum, gradually falls onto the central star (e.g., Lynden-Bell and Pringle, 1974; Pringle, 1981). This transport of mass inwards and angular momentum outwards can be provided by a variety of physical processes, such as turbulence originating from various instabilities in the disc or, in the case of relatively massive discs, by disc self-gravity. These processes play a dual role here: firstly, they drive outward angular momentum transport, which is necessary for accretion to occur, by exerting torques on the disc and, secondly, they provide a channel of conversion of gravitational energy, liberated as mass falls towards the centre, into thermal energy. The dissipated energy, in turn, can influence the observed SED. The angular momentum transport by turbulent torques (stresses) can be conveniently characterised in terms of so-called *turbulent, or 'anomalous' viscosity*[1], the concept of which in the theory of astrophysical discs was first introduced by von Weizsäcker (1948) and further developed by Shakura and Sunyaev (1973) (see also comprehensive reviews by Balbus and Hawley 1998 and Balbus 2003). In this chapter, we outline mechanisms that can provide angular momentum transport in discs and a possible nature of this anomalous/enhanced viscosity.

---

[1] We will show below that ordinary molecular viscosity of gas is far too small to yield typical time-scales of secular evolution of accretion discs.

To first gain insight into how viscosity, irrespective of its origin, causes mass and angular momentum redistribution in discs, following Lynden-Bell and Pringle (1974); Pringle (1981); Frank et al. (2002), let us consider the simplest model of a protoplanetary disc – a razor-thin (2D) gaseous disc rotating around a central star with considerably larger mass $M_\star$, so that disc self-gravity can be neglected (we will return to self-gravitating discs later). We adopt cylindrical coordinates (r, ϕ, z) with the central star at the origin and the disc lying in the $z = 0$ plane. The razor-thin disc approximation implies that the typical length-scale in the vertical direction – the disc scale height $H \simeq c_s/\Omega_0$, where c_s is the gas sound speed and Ω_0 is the angular velocity of disc rotation, is much smaller than the radial distance from the central star, that is, the aspect ratio $H/r \ll 1$.[2] In this case, basic hydrodynamical quantities are vertically averaged, that is, integrated in the vertical direction all over the disc height, and thus are made 2D (see section 2.3). The disc is then characterised by its surface density Σ, which is a mass per unit surface area of the disc, obtained by vertically averaging the gas density ρ. The thin disc approximation is also equivalent to the requirement that the sound speed c_s be much less than the rotation velocity $r\Omega_0(r)$, that is, the disc flow is strongly supersonic. From the latter condition, in turn, it follows that the radial gradients of pressure can be ignored, so that the disc angular velocity is given by a balance between the gravity force of the central star and centrifugal force due to rotation, resulting in the Keplerian rotation profile (Pringle, 1981; Frank et al., 2002; Lodato, 2007)

$$\Omega_0(r) = \left(\frac{GM_\star}{r^3}\right)^{1/2}.$$

In this case, a disc is said to be rotationally supported. The disc accretes, meaning that in addition to azimuthal Keplerian velocity it also possesses radial ‘drift’ velocity, u_r, directed towards the central star. As will be clear below, this radial velocity is much smaller than the Keplerian velocity (the ordering $u_r \ll c_s \ll r\Omega_0$ holds) and is directly related to the viscosity ν. The disc is assumed to be axisymmetric, so that all quantities are functions of only radius r and time t (z-dependence is absorbed in the vertical averaging). The height-integrated continuity equation for surface density is

$$\frac{\partial \Sigma}{\partial t} + \frac{1}{r}\frac{\partial}{\partial r}(r\Sigma u_r) = 0. \tag{2.1}$$

We also need another height-integrated equation for angular momentum conservation, which

---

[2] In protoplanetary discs, typically $H/r \simeq 0.05 - 0.1$ (e.g., Durisen et al., 2007), so the condition for thin disc limit is marginally satisfied. Nevertheless, many studies are limited just to this approximation, because it helps to understand basic physical processes and instabilities, which can then be generalised to thick 3D discs.

can be similarly derived. The only difference is that now we should include the viscous stress tensor, which in this special case is dominated by $r\phi$-component proportional to ν,

$$W_{r\phi} = \nu \frac{d\Omega_0}{d\ \ln\ r}.$$

As a result, we obtain

$$\frac{\partial}{\partial t}(r^2\Sigma\Omega_0) + \frac{1}{r}\frac{\partial}{\partial r}(r^3\Sigma\Omega_0 u_r) = \frac{1}{r}\frac{\partial}{\partial r}(r^2\Sigma W_{r\phi}). \tag{2.2}$$

Equations (2.1-2.2) can be combined into a single equation for surface density evolution

$$\frac{\partial\Sigma}{\partial t} = \frac{3}{r}\frac{\partial}{\partial r}\left[r^{1/2}\frac{\partial}{\partial r}(\nu r^{1/2}\Sigma)\right]. \tag{2.3}$$

In deriving this equation we have made use of the fact that the rotation is stationary and Keplerian. Equation (2.3), commonly used in accretion disc modelling, is the basic equation governing the secular, or viscous evolution of the disc. This equation is a non-linear diffusion-type equation for Σ, where ν can be a complex function of local variables (surface density, radius, temperature, ionisation fraction, etc.) so it should generally be solved numerically. However, if viscosity can be expressed as some power of radius, then analytic solutions are feasible (Lynden-Bell and Pringle, 1974; Armitage, 2010) that allow us to get the general feeling of a disc's viscous evolution.

A simple inspection of the structure of Eq. (2.3) shows that the radial velocity is proportional to viscosity,

$$u_r = -\frac{3}{r^{1/2}\Sigma}\frac{\partial}{\partial r}(\nu r^{1/2}\Sigma) \sim \frac{\nu}{r}$$

and the characteristic time-scale of the surface density evolution due to viscosity is

$$t_{visc} = \frac{r^2}{\nu} \sim \frac{r}{u_r}.$$

In real discs, this secular evolution time-scale is usually much larger than the dynamical/orbital time $t_{dyn} \equiv \Omega_0^{-1}$ (see below), implying, as noted above, that the radial accretion velocity u_r is much smaller than the disc's rotational velocity.

To illustrate viscous evolution more quantitatively, we consider a simple, analytically tractable case of constant viscosity coefficient, $\nu = const.$ Suppose that initially at $t = 0$ matter is concentrated in a very thin ring with radius r_0 and mass m. The surface density of such a ring has the form

$$\Sigma(r, t = 0) = \frac{m}{2\pi r_0}\delta(r - r_0).$$

An analytic solution can be constructed in terms of Bessel functions describing the time-development of this initial ring-shaped profile of surface density (e.g., Pringle, 1981; Frank

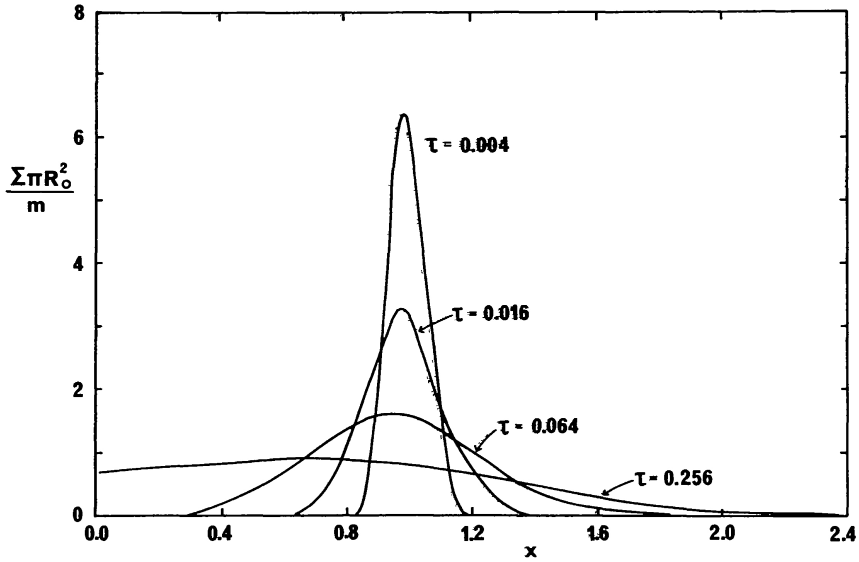

Figure 2.1: Evolution of the surface density of a Keplerian spreading ring under the action of constant viscosity ν, shown as a function of the normalised radius $x = r/r_0$, where r_0 is the initial radius of the ring, and of non-dimensional time $\tau = 12\nu t/r_0^2$. Adapted from Pringle (1981).

et al., 2002; Lodato, 2007), which we do not give here. Instead, we show it graphically in Fig. 2.1 at different times. It is clear from this figure that viscosity acts to spread the ring both inwards and outwards on the time-scale of t_{visc}. During this process, mass in the inner parts of the disc, losing angular momentum, moves inwards and eventually falls onto the central star. At the same time, matter in the outer parts moves to larger radii to take up outwardly flowing angular momentum. The net angular momentum is conserved; it is merely redistributed between different annuli. A detailed analysis of the corresponding analytical solution shows that the boundary between inwardly and outwardly propagating regions of the spreading ring expands outwards as $r_b \simeq t\nu/r_0$ (Pringle, 1981; Frank et al., 2002; Lodato, 2007). As a result, eventually (at $t \to \infty$) a singular equilibrium state is reached – almost all the matter stripped of its angular momentum is accreted at the centre, while a negligibly small amount of mass at infinitely large radii carries all the angular momentum. In another case, where viscosity has a power-law dependence on radius, an analytic self-similar solution can also be obtained that gives a qualitatively similar evolutionary picture (Lynden-Bell and Pringle, 1974; Lodato, 2007; Armitage, 2010).

The above analysis shows that viscosity is crucial for disc evolution – it is a primary means of outward transport of angular momentum. However, a major unresolved problem here concerns the physical nature of viscosity ν. It had soon been realised that standard molecular viscosity is too small to account for observed mass accretion rates and time-scales for disc secular evolution. Indeed, molecular viscosity is given by the product of random velocity of molecules, which is of the order of the sound speed c_s, and their collisional mean free path λ, $\nu = \lambda c_s$. In a typical protoplanetary disc with the number density $n \sim 10^{13}$ cm^{-3} of molecules having the cross-section $\sigma_{coll} \sim 10^{-16}$ cm^2, the mean free path is $\lambda = 1/(n\sigma_{coll}) \sim 10^3$ cm. So, the ratio of viscous to dynamical times is $t_{visc}/t_{dyn} = r^2/(H\lambda) = Re$, where $Re \equiv r^2\Omega_0/\nu$ is the Reynolds number of disc flow. If we now take $r \sim 100$ AU for the disc radial size and $H/r \sim 0.1$, we obtain $t_{visc}/t_{dyn} = Re \sim 10^{13}$. Since the dynamical timescale for protoplanetary discs is of the order of a few years, the viscous time $t_{vis} \sim 10^{13}$ yr, during which the whole disc accretes onto the central star, turns out to be much larger than observationally inferred disc lifetimes of $10^6 - 10^7$ yr. This means that there must exist some other mechanisms capable of producing orders of magnitude larger, or anomalous, effective viscosity resulting in mass accretion rates that will be consistent with observed disc lifetimes.

Very large Reynolds numbers, $Re \sim 10^{13}$, associated with the disc flow seem to offer a clue to resolve this problem. In laboratory experiments, hydrodynamic flows usually tend to become turbulent at high enough Reynolds numbers (e.g., Taylor, 1936; Bayly et al., 1988; Richard and Zahn, 1999; Drazin and Reid, 1981; Longaretti, 2002). Based on this, it was originally thought that gas flow in discs with such huge Reynolds numbers should naturally be strongly turbulent.[3] In this case, angular momentum exchange occurs not through collisions of molecules, but due to the mixing of fluid elements in turbulent motion. The typical length-scales of such turbulent motions can be several orders of magnitude larger than the collisional mean free path of molecules and therefore transport becomes much more efficient. The question is then how such an enhanced turbulent transport can be accommodated in the framework of a viscous disc model described above, or, in other words, what effective viscosity coefficient should be used to represent the turbulence.

To understand this, we note that the total gas velocity in a turbulent Keplerian disc can be written as a sum of the background Keplerian rotation and turbulent fluctuation $\mathbf{u}_t$,

$$\mathbf{u} = r\Omega_0(r)\mathbf{e}_\phi + \mathbf{u}_t, \tag{2.4}$$

where $\mathbf{e}_\phi$ is the unit vector in the azimuthal ϕ-direction. In this case, the mean accretion velocity is $u_r = \langle u_{tr} \rangle$, where the angle brackets denote vertical and azimuthal averages. We assume that this accretion velocity is much smaller than the fluctuation velocity amplitude and hence rotational velocity, so that the ordering $\langle u_{tr} \rangle \ll u_t \ll r\Omega_0$ holds (see e.g., Balbus,

---

[3] We will see below that care is needed in generalising this statement to accretion disc flows, especially to Keplerian rotation profiles (see section 2.2.1).

2003). Taking decomposition (2.4) into account, the angular momentum equation (without molecular viscosity) takes the form (Balbus and Hawley, 1998; Balbus and Papaloizou, 1999; Balbus, 2003)

$$\frac{\partial}{\partial t}(r^2\Sigma\Omega_0) + \frac{1}{r}\frac{\partial}{\partial r}(r^3\Sigma\Omega_0 u_r) = -\frac{1}{r}\frac{\partial}{\partial r}(r^2\Sigma\langle u_{tr}u_{t\phi}\rangle).$$

Comparing this equation to Eq. (2.2), we see that the averaged turbulent stress $\langle u_{tr}u_{t\phi}\rangle$ plays a similar role as the viscous stress $W_{r\phi}$. Consequently, we can define the effective turbulent viscosity as

$$\nu_t = \langle u_{tr}u_{t\phi}\rangle \left|\frac{d\Omega_0}{d\ \ln\ r}\right|^{-1} \sim \lambda_t u_t,$$

where $\lambda_t \sim u_t/\Omega_0$ is the characteristic length-scale of the turbulent motions. This expression resembles that for molecular viscosity. However, since turbulence is generally a very complex phenomenon, we do not know how to determine exactly the characteristic length-scales, λ_t, and velocities, u_t, associated with a turbulent flow without a proper knowledge of the underlying physical mechanisms causing the transition to turbulence in discs. Nevertheless, we can use physical arguments to put some constraints on these two quantities. Firstly, the typical size of the largest turbulent eddies cannot exceed the disc thickness, $\lambda_t \lesssim H$, otherwise turbulence would not be isotropic and transport would not be local, i.e., the characterisation of turbulent transport in terms of viscosity coefficient would not be possible. Secondly, it is unlikely that turbulence in discs is very supersonic, because in this case strong shocks would appear resulting in a high dissipation rate of turbulent motions. As a result, turbulent velocity would be damped to subsonic values, $u_t \lesssim c_s$. Based on these two basic arguments, in their seminal paper, Shakura and Sunyaev (1973) came up with the following parametrisation of turbulent viscosity:

$$\nu_t = \alpha c_s H,$$

where, because of the above limits on the size and velocity of turbulent motions, the non-dimensional parameter α is less than unity. This formulation, also mentioned above, is known as the 'α-prescription' for turbulent viscosity. In fact, α contains in itself all the uncertainties regarding the onset mechanism and properties of turbulence in discs, that is, the source of anomalous viscosity. Doing such a scaling, we just replace one unknown parameter ν_t with the other unknown one, α. But the usefulness of this approach is that now we measure the turbulent stress, which determines the angular momentum transport rate, in units of the local pressure and we know that in these units, if turbulence is local, it should not exceed unity. In other words, the characterisation of angular momentum transport in terms of the

α parameter is an inherently local description valid for thin ($H \ll r$) discs.[4] Although, as numerical simulations show, α may display both temporal and spatial variations in the disc (e.g., Papaloizou and Nelson, 2003), in analytical modelling it is often assumed to be constant. This actually gives reasonably good results as far as disc secular evolution is concerned. Observations indicate that $\alpha \sim 0.01$ for protoplanetary discs (e.g., Hartmann et al., 1998; Hueso and Guillot, 2005; Andrews et al., 2009), which corresponds to disc evolutionary timescales of the order of observed disc lifetimes of a few Myr. Thus, the α parameter is a key quantity as it can be accessed observationally and can provide a check on the theory of turbulence and anomalous transport in discs.

It is crucial to have a proper understanding of disc turbulence and its onset mechanisms, as apart from driving the accretion process, it can also have a profound effect on other processes occurring in discs. This, in turn, allows us to determine α in different parts of the disc, critically examine the validity of the constant-α assumption and also make comparisons with observations. Great progress has been made in recent years in identifying the mechanisms of transition to turbulence in astrophysical discs (see e.g., Balbus, 2003). As in the case of hydrodynamic turbulence in shear flows, turbulence in discs (that in fact represent a special case of shear flow as discs almost always rotate differentially) should continuously extract energy from the disc flow at the largest scales ($\sim H$) and transfer it to motions at smaller and smaller scales until eventual dissipation on the smallest viscous scales. The extraction of energy at the largest scales, in turn, is possible owing to various types of instabilities in the disc. Turbulence should also provide ‘right’ outward transport of angular momentum for accretion. Thus, the starting point in tackling the disc turbulence problem is to explore what types of instabilities can develop in accretion discs. The Keplerian differential rotation plays a special role here because, as observations indicate (e.g., Dutrey, 2000; Simon et al., 2000), in most cases protoplanetary discs do rotate with Keplerian or nearly Keplerian velocities unless they are strongly self-gravitating. Therefore, research in this direction is mainly focused on investigating the stability of the Keplerian rotation profile, its subsequent (possible) transition to turbulence and the properties (e.g., angular momentum transport rate) of the resulting turbulent state itself when other physical factors specific to discs (magnetic fields, self-gravity, radial and vertical stratification, etc.) are involved. However, we should note in this respect that not all types of disc instabilities in the non-linear regime necessarily lead to turbulence in the classical sense, but can result in some kind of more or-

---

[4] Note that there are also other non-local, long-range angular momentum transport mechanisms, for example, gravitational instability described in section 2.2.3 below.

ganised, quasi-steady, saturated non-linear states (e.g., coherent vortices, quasi-steady spiral structure, convective rolls) that can also act to maintain angular momentum transport (in such cases, λ_t and u_t simply correspond to the characteristic sizes and velocities of these organised non-linear motions). Mechanisms/instabilities causing enhanced angular momentum transport in discs, broadly speaking, can be divided into two basic categories: magnetohydrodynamic (MHD) mechanisms, primarily giving rise to magnetorotational instability (MRI), and non-magnetic mechanisms, such as a purely hydrodynamical (i.e., driven by shear of disc flow velocity) turbulence, baroclinic/Rossby wave instability and coherent vortices, spiral density waves, convection and gravitational instability if the disc is self-gravitating. Below we describe/review each of these mechanisms separately and then put them into the context of the main thesis work.[5] However, we will not expand much on the MRI and put more emphasis on non-magnetic means of transport, since the latter is the central subject of study in the present thesis. Besides, MHD instabilities (mostly the MRI) in accretion discs have been the focus of numerous studies since the 1990s (see reviews by Balbus and Hawley, 1998; Balbus, 2003) and today are accepted as the most likely cause of turbulence and enhanced angular momentum transport in many accretion discs, but not necessarily in protoplanetary discs. There can be certain conditions in the protoplanetary disc under which the MRI cannot operate and that is where one has to appeal to non-magnetic means of transport.

In order for the MRI to develop, the disc should necessarily have regions where gas is sufficiently ionised to couple effectively with magnetic field, so that the latter can affect the gas dynamics (Blaes and Balbus, 1994; Gammie, 1996; Sano et al., 2000; Fromang et al., 2002; Sano and Stone, 2002; Salmeron and Wardle, 2003; Desch, 2004). The minimum ionisation fraction (i.e., the ratio of electron concentration, which are the main carriers of current, to that of neutrals) required for the onset of the MRI in protoplanetary discs is very low, typically 10^{-13} (e.g., Gammie, 1996). Above this minimum ionisation, magnetic diffusion is small and cannot suppress the MRI, while at lower ionisation fractions the MRI is quenched due to dissipative effects. Disc ionisation can be provided by thermal collisions of molecules, X-rays and cosmic rays. Thermal ionisation of alkali metals (potassium) can occur at $T \gtrsim 10^3$ K in the inner parts ($\lesssim 0.1$ AU) of the disc (Umebayashi, 1983) and keep the ionisation level there above the threshold value, so that these inner regions are maintained magnetically active. Ionisation due to cosmic rays or due to illumination from the central star's X-ray

---

[5] We do not consider here external factors, such as the effects of a companion star, which can also produce anomalous angular momentum transport by tidally inducing shocks of spiral density waves in the disc (e.g., Rozyczka and Spruit, 1993; Lin and Papaloizou, 1993).

and UV radiation dominate at larger radii, where the disc is cooler and thermal ionisation is ineffective. These non-thermal sources, entering the disc from the top and bottom, can ionise gas only in surface layers with column densities up to a maximum value of 100 g cm^{-2} above which they cannot penetrate deeper due to attenuation (Umebayashi and Nakano, 1981). As a result, a large magnetically inactive region – so-called 'dead zone' – forms near the disc midplane outside ~ 0.1 AU, where the ionisation level is much smaller than that required for magnetic coupling. This dead zone with almost no MHD turbulent activity thus appears sandwiched between two magnetically active, MRI-turbulent surface layers and can extend to radii as large as 30 AU (layered accretion disc model, see e.g., Gammie, 1996; Fromang et al., 2002; Fleming and Stone, 2003). Beyond this radius, the disc surface density drops below the maximum value, so that cosmic rays can propagate all along the height and fully ionise the gas, thereby making it turbulent again and the layered structure disappears. Analogous almost neutral regions can also exist in cataclysmic variable (CV) discs (Gammie and Menou, 1998) and also in the outer parts of cool active galactic nuclei (AGN) discs (Menou and Quataert, 2001). The absence of sufficient magnetic coupling, and therefore of MHD turbulence, within such large portions of the disc necessitates a search for other alternative non-magnetic mechanisms of angular momentum transport. However, it should be noted in this regard that it is also possible that the turbulence in the magnetically active surface layers induces small angular momentum transport in the dead zone by driving velocity fluctuations (hydrodynamical waves) there and/or through diffusive penetration of magnetic fields to the midplane (Fleming and Stone, 2003; Turner et al., 2007; Oishi et al., 2007; Turner and Sano, 2008; Oishi and Mac Low, 2009).

2.1 Magnetorotational instability

The MRI is a linear instability occurring in differentially rotating, ionised gaseous discs threaded by magnetic field. The remarkable feature of this instability is that it grows on a characteristic time-scale of the order of the orbital period and even a very weak imposed magnetic field is sufficient for its activation. In other words, the MRI persists and retains its key properties in the limit of vanishing magnetic field, $B \to 0$, or equivalently, for weak ionisation, but, as mentioned above, the ionisation fraction of the disc must still be larger than a certain minimum value, which can be very small. We note in this respect that the stability properties of weakly ionised discs are qualitatively different from those of non-magnetised discs.

The MRI was first discovered by Velikhov (1959) in the stability analysis of magnetised Taylor-Couette flows – inhomogeneous flows between two rotating cylinders threaded by magnetic field along the vertical axes – and then examined in more detail by Chandrasekhar (1960). The relevance and immediate significance of the MRI to accretion disc turbulence problem was recognised after the seminal paper by Balbus and Hawley (1991), who demonstrated for the first time that the MRI can in fact operate in discs threaded by a (stellar) magnetic field in the presence of weak ionisation and differential rotation. This is clearly seen from the local dispersion relation for small axisymmetric perturbations that can be derived using linearised basic MHD equations (e.g., Balbus and Hawley, 1998; Balbus, 2003)

$$\omega^4 - \omega^2 \left[\kappa^2 + 2\left(\mathbf{k}\cdot\mathbf{v_a}\right)^2\right] + \left(\mathbf{k}\cdot\mathbf{v_a}\right)^2 \left[\left(\mathbf{k}\cdot\mathbf{v_a}\right)^2 + \frac{d\Omega_0^2}{d\ \ln\ r}\right] = 0$$

where ω and $\mathbf{k}$ are the frequency and wavenumber of perturbations, $\mathbf{v_a} = \mathbf{B}/\sqrt{4\pi\rho}$ is the Alfvén speed and κ is the epicyclic frequency ($\kappa = \Omega_0$ for Keplerian rotation). This dispersion relation has unstable roots ($\omega^2 < 0$), that is, the MRI is at work, if the angular velocity, $\Omega_0(r)$, of disc rotation is a decreasing function of radius

$$\frac{d\Omega_0^2}{dr} < 0.$$

Note that the magnetic field is not present in this inequality. The Keplerian rotation falls into this category and is therefore subject to the MRI. The maximum growth rate can be easily found

$$|\omega_{max}| = \frac{1}{2}\left|\frac{d\Omega_0}{d\ \ln\ r}\right|,$$

which is indeed of the order of the angular velocity of disc rotation and is also independent of the magnetic field. As mentioned above, this means that the MRI can, in principle, operate for arbitrarily small imposed magnetic fields. This is the consequence of the field-freezing (zero resistivity) condition ensuring the conservation of magnetic field flux. What depends on magnetic field is the wavenumber $\mathbf{k}_{max}$ corresponding to this maximum growth rate,

$$\left(\mathbf{k}_{max}\cdot\mathbf{v_a}\right)^2 = \left(\frac{1}{4} + \frac{\kappa^2}{16\Omega_0^2}\right)\left|\frac{d\Omega_0^2}{d\ \ln\ r}\right|.$$

From this expression it follows that the critical wavenumber increases with decreasing magnetic field. So, weak subthermal ($v_a \lesssim c_s$) magnetic fields tend to give rise to local instability.

A simple mechanical analogy can help to grasp the basic physics underlying the MRI (here we follow Balbus and Hawley, 1998; Balbus, 2003). Imagine two fluid elements in a Keplerian disc initially at the same orbital location and then displaced to slightly different radii, one element inwards and the other outwards. The elements are connected by a massless weak spring, which can be shown to mimic the action of a magnetic force on fluid elements that depends linearly on their displacement. Because of Keplerian rotation, the element at a smaller radius is rotating faster than that at a larger radius. This increases the separation between these two elements. As a result, the connecting spring is stretched and develops tension that pulls the faster element backwards decreasing its angular momentum and, at the same time, the slower element forwards increasing its angular momentum. Thus, the faster element is continuously losing angular momentum to the slower one via the spring. This means that these two elements cannot stay at their original orbits – the faster element drops further inwards to even smaller radii to accommodate its reduced angular momentum; conversely, the slower element moves to even larger radii corresponding to increasing angular momentum. In this process, the spring tension grows with increasing element separation, which, in turn, leads to larger angular momentum changes of the elements and thus to the runaway growth of their separation. This is the essence of the MRI in differentially rotating systems.

The development of the MRI into turbulence was first followed by means of non-linear numerical simulations in the local shearing sheet/box approximation[6] by Brandenburg et al. (1995); Hawley et al. (1995); Stone et al. (1996) and by many others. These simulations explored the influence of (numerical) viscosity, resistivity, numerical resolution, vertical stratification, etc. on the saturated values of α in a developed turbulent state (e.g., Sano and Stone, 2002; Fromang and Papaloizou, 2007; Fromang et al., 2007; Bodo et al., 2008; Davis et al., 2010; Longaretti and Lesur, 2010, see also review by Balbus 2003 and references therein). In the context of global discs, it was first investigated by Armitage (1998); Hawley (2001); Papaloizou and Nelson (2003) for different magnetic field geometries. All these studies unanimously demonstrate that in the non-linear regime, the MRI quite robustly leads to the breakdown of the Keplerian shear flow into self-sustained MHD turbulence that transports angular momentum outwards predominantly via Maxwell stresses. The resulting values of α obtained from simulations depend on the above-mentioned factors and are typically in the range $10^{-3} - 10^{-1}$ that more or less agree with observed values (see e.g., Balbus, 2003; King et al., 2007). As mentioned above, today MRI-driven turbulence is considered to be the most likely candidate providing angular momentum transport in accretion discs and is still the subject of on-going active research.

---

[6] The shearing sheet/box approximation is described in section 2.3.

2.2 Non-magnetic mechanisms of angular momentum transport

2.2.1 Hydrodynamical instabilities

In laboratory experiments, hydrodynamic shear flows become turbulent when the Reynolds number exceeds a certain critical value (see e.g., Drazin and Reid, 1981; Bayly et al., 1988). Turbulence generally sets in as a result of the non-linear development of various (linear or non-linear) instabilities associated with the mean velocity shear of flow. Owing to these instabilities, energy is extracted from the background flow and goes into feeding turbulence, thereby making it self-sustained. A key feature of shear-driven turbulence is that it is characterised by a high degree of correlation between streamwise and shearwise velocity fluctuations ensuring that momentum (or angular momentum in the case of accretion discs) is transported in the 'right' direction (i.e., towards reducing shear), which in discs with decreasing angular velocity profile would correspond to outward transport.

As mentioned above, in the majority of cases protoplanetary discs are in Keplerian differential rotation, which is linearly stable to *axisymmetric* disturbances for arbitrarily large Reynolds numbers according to Rayleigh's criterion (specific angular momentum increases with radius and has no inflection point, Rayleigh, 1880). However, a number of laboratory experiments (e.g., Taylor, 1936; Bidokhti and Tritton, 1992; Dauchot and Daviaud, 1995; Tillmark and Alfredsson, 1996; Bottin et al., 1997; Richard, 2001; Richard et al., 2001; Longaretti and Dauchot, 2005) and numerical simulations (e.g., Hamilton et al., 1995; Schmiegel and Eckhardt, 1997; Eckhardt and Mersmann, 1999) clearly indicate transition to turbulence in linearly stable shear flows at Reynolds numbers larger than certain critical values. Because such flows exhibit linear stability at all Reynolds numbers, the transition Reynolds number is determined by non-linear effects and the transition to turbulence in this case is called subcritical. Earlier understanding of a subcritical transition was associated with non-linear instabilities in shear flows (e.g., Orszag and Patera, 1980, 1983; Bayly et al., 1988; Lerner and Knobloch, 1988; Dubrulle and Zahn, 1991; Dubrulle, 1993). A key factor in this scenario is the existence of quasi-steady, two-dimensional, long-wavelength, non-linear perturbations in the mean flow, which are spatially periodic in the streamwise direction (see e.g., Zahn et al., 1974). These finite-amplitude solutions, in turn, modify the basic flow profile, thereby making it susceptible to rapid, short-wavelength three-dimensional (secondary) instabilities that eventually lead to turbulence. So, in this case, the turbulent state is energetically sustained by non-linear processes.[7] This general trend of the onset of hydrodynamic turbulence at high Reynolds numbers in laboratory flows prompted Lynden-Bell and Pringle (1974) and Shakura et al. (1978) to simply suggest that the same mechanisms, i.e., non-linear instabilities and huge Reynolds numbers ($\sim 10^{13}$), would naturally produce hydrodynamic turbulence in discs as well, which can then be described with the above-mentioned $\alpha-$prescription (although MHD turbulence can also be accommodated within this parametrisation). Thus, hydrodynamic turbulence, if present in disc flows, must be of a subcritical nature.[8]

This proposition was first challenged and examined in more detail by Balbus et al. (1996) and Hawley et al. (1999), who carried out 3D non-linear simulations of rotating shear flows with the aim of understanding the transition to turbulence in non-magnetised discs. These two studies concluded that the Keplerian profile is quite stable (locally) to non-linear perturbations and shows no sign of a transition to turbulence. The non-linear stability of Keplerian shear flows was attributed to the Coriolis force that appears to have a stabilising effect; the non-rotating shear layers and marginally stable (i.e., nearly constant angular momentum) flows, for which the effect of rotation is small, indeed quickly became turbulent in their simulations.

Longaretti (2002) argued that numerical simulations did not show turbulence in Keplerian flows, because the numerical resolution typically adopted was not high enough to resolve turbulent length-scales. The largest turbulent scales are substantially reduced under the action of the Coriolis force (Bech et al., 1995; Bidokhti and Tritton, 1992), implying that more numerical effort is needed to resolve them and to see a subcritical turbulent transition showing up in simulations. For the same reason, the simulations by Balbus et al. and Hawley et al. might not be able to capture such length-scales because of the relatively low resolution (256^3) adopted.[9] Finally, we note that the non-linear hydrodynamic stability of

---

[7] Unfortunately, such an analysis gives little information on the properties of the turbulent state itself – its location in the parameter space and the efficiency of turbulent transport (see e.g., Lesur and Longaretti, 2005). Below we describe an alternative and more recent ‘bypass’ scenario of subcritical transition in shear flows that tends to better reproduce experimental results.

[8] Global linear, non-axisymmetric instabilities of the type considered by Papaloizou and Pringle (1984, 1985); Narayan et al. (1987); Dubrulle et al. (2005) can also take place in accretion discs despite the axisymmetric stability according to Rayleigh’s criterion. However, these instabilities hinge on the boundary dynamics and therefore seem less relevant to accretion disc turbulence which is a local phenomenon. Moreover, it was demonstrated that the non-linear development of these instabilities does not lead to turbulence, but to a saturated state (Blaes, 1987; Hawley, 1991).

[9] We should note in this regard that recent analogous simulations by Shen et al. (2006) with much higher resolution ($1024^2 \times 256$) did not display any turbulence either; instead the perturbation energy and α decay at late times, see section 2.2.4.

Keplerian rotation was also confirmed in a global disc setup by Godon and Livio (1999a), but their calculations are restricted to two dimensions and therefore of limited relevance to hydrodynamic stability.

Experiments with Rayleigh-stable Taylor-Couette flows characterised by radially decreasing angular velocity and radially increasing specific angular momentum profiles (see e.g., Richard and Zahn, 1999; Richard, 2001; Richard et al., 2001) as well as with plane rotating shear flows under the influence of a stabilising Coriolis force (Bidokhti and Tritton, 1992; Tillmark and Alfredsson, 1996; Longaretti and Dauchot, 2005), nevertheless, display a subcritical transition to turbulence at high enough Reynolds numbers. But boundary dynamics (e.g., secondary flows, such as Eckman circulation near boundaries), which is not shared by disc flows, could possibly play a role in turbulent transition in these experiments. This is clearly in conflict with the results of Balbus et al. and Hawley et al. and also with the recent analogous experiments with Taylor-Couette flow by Ji et al. (2006), carried out at much higher Reynolds numbers (up to millions) than those in previous ones, not showing any turbulent transition for Keplerian-like flows. However, it was found both experimentally (Tillmark and Alfredsson, 1996) and in numerical simulations (Lesur and Longaretti, 2005) that the transition Reynolds number steeply increases when stabilising rotation, especially an anticyclonic one, is imposed (Fig. 2.2). This, with the above-discussed reduced turbulent scales due to rotation, can be considered as the main reason why hydrodynamic turbulence is so elusive in numerical simulations. Just this Rayleigh-stable anticyclonic regime is of principal interest for discs, because Keplerian flow itself and its counterpart in the local shearing box model – azimuthal plane Couette flow with radial shear and anticyclonic rotation – are Rayleigh-stable.

An alternative scenario – a so-called ‘bypass’ concept – of subcritical transition to turbulence in shear flows was developed by the hydrodynamical community in the 1990s (e.g., Trefethen et al., 1993; Gebhardt and Grossmann, 1994; Baggett et al., 1995; Waleffe, 1997; Brosa and Grossmann, 1999; Grossmann, 2000; Schmid, 2000; Schmid and Henningson, 2000; Reshotko, 2001; Waleffe, 2003) and its applicability/relevance to the disc turbulence problem was investigated by several authors (Chagelishvili et al., 2003; Tevzadze et al., 2003; Yecko, 2004; Mukhopadhyay et al., 2005; Afshordi et al., 2005; Longaretti, 2002; Lesur and Longaretti, 2005; Rincon et al., 2007; Tevzadze et al., 2008). This scenario relies on the well-known and important result from hydrodynamical studies that linear operators governing the time-development of perturbations in shear flows are mathematically non-normal, that is, these operators are not commutable with their corresponding hermitian adjoint ones

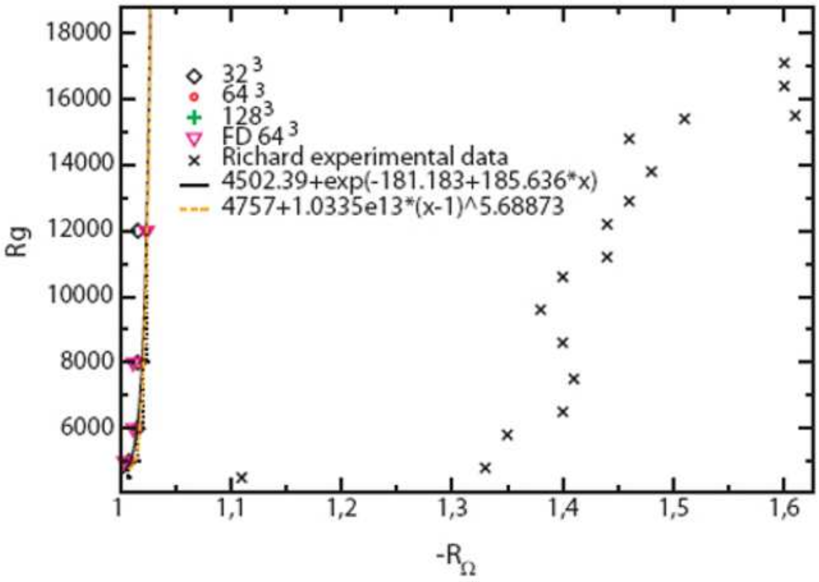

Figure 2.2: Transition Reynolds number R_g versus rotation number $R_\Omega = 2\Omega/S$ on the anticyclonic side as obtained by Lesur and Longaretti (2005), where Ω is the angular velocity of flow rotation and S is the shear parameter. Crosses correspond to the experimental results of Richard (2001), other signs indicate numerically obtained transition Reynolds numbers for different resolutions, solid and dotted lines represent interpolation curves. As clearly seen, the theoretical interpolated curves are substantially steeper than the experimental data. For a Keplerian flow ($R_\Omega = -4/3$, anticyclonic rotation), depending on the method of interpolation, Lesur and Longaretti obtained for R_g values $1.8 \times 10^6, 2 \times 10^7, 1.1 \times 10^{10}$ and 1.3×10^{26}. The first three values are comparable to or smaller than Reynolds numbers typically found in accretion discs ($10^{10} - 10^{16}$), while the fourth one is much greater, implying the absence of turbulence. As mentioned in the text, disagreement with the experimental data of Richard (2001) may be due to secondary flows induced by boundaries in the experiments that tend to undermine flow stability. Thus, there is still uncertainty whether or not purely hydrodynamic turbulence can actually occur in accretion discs.

(Reddy et al., 1993; Trefethen et al., 1993; Grossmann, 2000). As a consequence of this non-normality, the eigenfunctions associated with such an operator are mutually non-orthogonal leading to transient (i.e., during a limited time interval) amplification of perturbations, which is stronger the larger the Reynolds number (e.g., Reddy and Henningson, 1993). Because of the very high Reynolds numbers in accretion discs, it was suggested that the linear transient growth of perturbations could be very effective in discs as well (Ioannou and Kakouris, 2001; Chagelishvili et al., 2003; Tevzadze et al., 2003; Yecko, 2004; Mukhopadhyay et al., 2005; Afshordi et al., 2005; Tevzadze et al., 2008). Specifically, it was shown that the transient amplification factor scales with the Reynolds number as $Re^{2/3}$ implying that it is very powerful. But non-linearity comes into play quickly because of such a strong transient growth. So, in the bypass scenario, the self-sustenance of turbulence is achieved by a subtle balance between non-normality-induced linear transient amplification of perturbations and non-linearity. The main energy supply for turbulence comes from the linear transient growth mechanism – linear amplification of streamwise streaks due to streamwise vortices. Then these strongly amplified streaks become subject to a three-dimensional linear instability, which provides a non-linear feedback re-energising original streamwise vortices and thus makes the whole process self-sustaining (see e.g., Waleffe, 1997; Rincon et al., 2007, for details). Thus, the linear transient amplification of perturbations due to the non-normality plays a key role as a source of energy for turbulence. Non-linearity, not producing any perturbation energy, only redistributes the energy gained via the linear mechanism to ensure self-sustenance of turbulence, or in other words, provide a positive feedback without which only transient growth would decay at large times. The bypass scenario is also able to better explain transition in shear flows at Reynolds numbers considerably smaller than those given by a classical theory of transition based on non-linear secondary instabilities outlined above (see Waleffe, 1997).

The bypass scenario of subcritical transition was originally formulated for non-rotating shear flows. Its possible role and effectiveness for differentially rotating disc flows in the Rayleigh-stable regime were investigated in detail by Lesur and Longaretti (2005) and Rincon et al. (2007) using the 3D shearing box approach (without vertical stratification), where the disc shear flow appears as a simple rotating plane Couette flow (see section 2.3). Effects of both cyclonic and anticyclonic rotation on turbulent transition were considered. Compared with previous simulations, Lesur and Longaretti were able to better control and quantify the effects of numerical resolution on a subcritical transition as well as to achieve higher Reynolds numbers in their study. For the specific range of rotation numbers considered, they found that for both rotation directions, the flow undergoes transition. However, the transition Reynolds numbers are big ($\gtrsim 10^4$) even close to the marginal stability limit (i.e., in the slow rotation regime) and rise steeply (by orders of magnitude) with further increasing the absolute value of rotation number on both cyclonic and anticyclonic sides, being still much steeper on the anticyclonic side relevant to disc flows (Fig. 2.2, a similar qualitative behaviour is seen in the cyclonic regime, see Fig. 4 of Lesur and Longaretti 2005). This implies that rotation in the Rayleigh-stable regime increasingly hinders transition to turbulence, as also demonstrated by the above-mentioned latest experimental work. Although simulations of Lesur and Longaretti cover only a small range of anticyclonic rotation numbers, they made interpolation of the numerically obtained dependence of transition Reynolds numbers on rotation numbers to larger (by absolute value) rotation numbers, including the value $-4/3$ corresponding to Keplerian rotation. Depending on the type of interpolation, corresponding transition Reynolds numbers obtained are similar or several orders of magnitude smaller than those typical of accretion discs (see Fig. 2.2), implying that the onset of hydrodynamic turbulence in discs is in principle possible. But unfortunately, the α parameter associated with such a turbulent state turned out to be too small. The most optimistic/largest value Lesur and Longaretti found, $\alpha_{max} = 3.1 \times 10^{-6}$, is still orders of magnitude smaller than that required in accretion discs. Thus, these results, in principle, do not rule out the possibility that the bypass mechanism can operate in Keplerian accretion discs, but intrinsic stabilising anticyclonic rotation greatly hinders it from being effectively realised. Rincon et al. (2007) also arrived at the same conclusion based on generalising steady, self-sustaining non-linear solutions found by Waleffe (1997, 2003) in non-rotating shear flows to the case of rotating shear flows. Although these authors did not succeed in finding any self-sustaining solution/process in the case of Rayleigh-stable anticyclonic rotation, as opposed to the case of cyclonic one, they expressed hope that such solutions may also exist in the anticyclonic regime under more general conditions (e.g., absence of walls, different boundary conditions).

In summary, to date a generally accepted view does not favour the idea that turbulence in Keplerian discs is of purely hydrodynamical origin, that is, driven by differential character, or shear, of the disc's rotation at high Reynolds numbers. (Although, as we will see throughout the thesis, the effect of velocity shear in various dynamical processes in discs is very important in its own right.) Nevertheless, the problem of hydrodynamic non-linear instability and turbulence in accretion discs is still not completely settled and there is on-going active research in this field to further explore a possible role of other non-magnetic physical factors specific to discs, such as vertical and radial stratification, convection, vortices and spiral density waves in driving angular momentum transport, which are outlined below.

2.2.2 Thermal convection

Thermal convection in the direction normal to the midplane was originally suggested as a possible source of anomalous angular momentum transport in protoplanetary discs long before the MRI (Cameron, 1978; Lin and Papaloizou, 1980, 1985; Ruden and Lin, 1986; Ruden and Pollack, 1991). In these earlier studies, it was pointed out, on the basis of radiative transfer calculations of the disc vertical structure, that under certain conditions a specific dependence of opacity on temperature in discs can give rise to a superadiabatic vertical temperature gradient (i.e., with specific entropy decreasing with height), which leads to convective instability according to Schwarzschild criterion for thermal convection. Recently, Rafikov (2007) has more precisely identified temperature and opacity regimes where the disc is expected to be convective based on the opacity law $\kappa = \kappa_0 P^\alpha T^\beta$ (Bell and Lin, 1994), where T and P are the gas temperature and pressure. Using an analysis similar to Lin and Papaloizou (1980), he showed that in optically thick discs (where one can use a flux-limited radiative transfer approach), under certain assumptions regarding the vertical distribution of the energy dissipation rate (which are met in protoplanetary discs in most cases), Schwarzschild criterion for convective instability can be equivalently formulated in terms of the exponents α and β as

$$\nabla_0 \equiv \frac{1+\alpha}{4-\beta} \geq \nabla_{ad},$$

where $\nabla_{ad} \equiv (d\ln T/d\ln P)_{ad} = 1-1/\gamma = 2/7$ is the adiabatic temperature gradient with the adiabatic index $\gamma = 1.4$ of gas (molecular hydrogen). From this condition it follows that in the low temperature regime the disc is convective for $T \lesssim 150$ K, where the opacity is dominated by ice grains corresponding to $\alpha = 0$ and $\beta \approx 2$ with $\nabla_0 \approx 0.5$. At higher temperatures, when ice grains sublimate, but below the sublimation point of silicate and metal grains that are responsible for the opacity in this regime, we have $\alpha = 0$ and $\beta \approx 0.5$ that also results in convective instability, though marginal, $\nabla_0 \approx \nabla_{ad}$. Sub-millimeter observations of protoplanetary discs demonstrate that the opacity due to dust grains at low temperatures ($\lesssim 100$ K) is well described by $\alpha = 0$ and $\beta \approx 1$ (Kitamura et al., 2002). In this case, $\nabla_0 \approx 1/3$ and the disc is again convective. At very high temperatures [$T \approx (1.5-5)\times 10^3$ K], when all dust grains sublimate, opacity becomes pressure dependent. In this case, $\alpha \approx 2/3$ and $\beta \approx 7/3$ for the regime of molecular opacity, so that $\nabla_0 \approx 1$ and the disc is convective. Thus, protoplanetary discs can be convectively unstable for a broad temperature range. Then corresponding effective α-viscosity was estimated in the above-mentioned studies, using a phenomenological mixing-length prescription, and evolutionary models of discs were constructed based on the implicit premise that angular momentum transport due to convection is outward, or the corresponding α is positive. The estimated evolutionary time-scales of these models seemed to be consistent with disc lifetimes. However, there had been no rigorous proof of the main assumption that vertical convection actually transports angular momentum outwards. So, subsequent research in this direction was devoted to clarification of the transport properties of convection.

Ruden et al. (1988) carried out a linear analysis of axisymmetric convective modes in discs. Although the linear axisymmetric modes do not produce torques themselves, both the radial and vertical extent of the unstable convective modes appeared comparable to the disc scale height and could result, as suggested by these authors, in an efficient transport of angular momentum either in the case of non-axisymmetric convection or through non-linear dissipation of axisymmetric vertical convection (e.g., by some sort of secondary instabilities of finite-amplitude axisymmetric convective rolls). Ruden et al. also estimated, from the radial wavelengths and growth rates of the axisymmetric convective modes, an expected value of $\alpha \simeq 10^{-3} - 10^{-2}$. However, the non-linear development of these axisymmetric (two-dimensional) convective modes turned out to lead to inward (i.e., directed towards the central star), rather than outward, transport of angular momentum (Kley et al., 1993; Rüdiger et al., 2002), contrary to what was originally thought. This is obviously not what is required for the accretion of matter from the disc onto the central star. Thus, angular momentum transport due to convection, if any, can only be due to non-axisymmetric convective motions.

To more fully explore the problem of convective transport, following studies focused on the dynamics of non-axisymmetric convection. Korycansky (1992) and Ryu and Goodman (1992) carried out a linear analysis of non-axisymmetric convective modes in Keplerian discs using the shearing box approximation.[10] The main result of their studies is that unstable non-axisymmetric convective modes eventually become tightly trailing, almost axisymmetric, due to sweeping by background Keplerian shear that, in turn, leads predominantly to an inward transport of angular momentum (negative effective viscosity). Ryu and Goodman also gave a simple explanation for inward transport by non-linear axisymmetric convective rolls mentioned above. Because of non-linear dissipation, any two neighbouring axisymmetric convective rolls try to equate their angular momenta, which in the case of a Keplerian profile leads to inward transport (see also Balbus, 2000). Later Lin et al. (1993) argued that the main reason for the inward transport is that the shearing box model adopted by these authors had a radially uniform (i.e., no radial structure) equilibrium state and that the shearing box boundary conditions themselves enforce, eventually, tightly trailing perturbations, which tend to transport angular momentum inwards. Lin et al. further demonstrated that the radially stratified background can support localised, in the radial direction, packets of the convective mode, which are able to provide an outward transport of angular momentum. But the initial conditions adopted in their linear numerical calculations were somewhat artificial. So, further non-linear study of non-axisymmetric convection was needed in order to decide the direction of convective transport.

---

[10] In contrast to axisymmetric convection, Keplerian shear/non-normality introduces new effects in the dynamics of non-axisymmetric convection, which have not been considered by these authors and are the main subject of analysis in chapter 4.

Earlier numerical simulations by Cabot and Pollack (1992) considered non-axisymmetric convection in Keplerian discs in the local shearing box approximation under the assumption of incompressibility. In these simulations, as in Kley et al. (1993), low Reynolds numbers (high artificial viscosities) were imposed. Their results showed that with increasing rotation velocity Reynolds stresses change sign from positive to negative, because of the convective flow becoming nearly axisymmetric at rapid rotation, which from previous 2D simulations is known to be associated with inward transport of angular momentum. Although, based on this result Cabot and Pollack hinted that convection in discs might not be characterised by outward transport, their study still left broadly the question of convective transport open.

Subsequent more general, non-linear 3D compressible simulations of convection in Keplerian discs by Cabot (1996) and Stone and Balbus (1996), performed again in the local shearing box approximation, did not yield outward transport either. They found that initially non-axisymmetric convective motions soon get sheared and become nearly axisymmetric (i.e., with no azimuthal dependence), so that azimuthal pressure gradients (pressure-strain correlations) are no longer coherent and cannot sustain long-term angular momentum transport. As a result, the corresponding α fluctuated rapidly in time between relatively large positive and negative values, but when averaged over several orbital periods, turned out to be very small and negative ($\approx -4.2 \times 10^{-5}$), implying inward transport. This led to vertical convection being regarded as an inefficient mechanism for driving angular momentum transport and disc's secular evolution. However, Klahr et al. (1999), using global disc simulations, demonstrated that convective flow can in fact be non-axisymmetric (i.e., have an azimuthal structure). They pointed out that the small azimuthal extent of the computational domains (neglect of disc curvature because of the use of shearing box approximation) and low Reynolds numbers used in the simulations of Cabot (1996) and Stone and Balbus (1996), together with shear, had a tendency to wipe out all azimuthal variations that, resulting in near axisymmetry, gives inward transport. But Klahr et al. did not calculate α associated with such non-axisymmetric 3D convective motions to see whether it is positive or negative.

Interest in the convective transport has revived recently, because of high-resolution simulations with improved numerical techniques becoming increasingly affordable. Lesur and Ogilvie (2010) have performed 3D simulations of *incompressible* convection in the shearing box and found, contrary to previous results, that at high enough Rayleigh (Reynolds) numbers ($\gtrsim 10^6$), the corresponding α becomes positive, though still remains small ($\sim 10^{-4}$).[11] This

---

[11] On the basis of analysis presented in chapter 4, we will later discuss the way of enhancing values of convective α.

was demonstrated to be a consequence of the non-axisymmetric structure of convection, which is established at large Rayleigh numbers and plays a central role in the sense that it gives rise to an appreciable pressure-strain correlation tensor that, in turn, changes the direction of transport to outward. Past simulations mentioned above, being at low Rayleigh (Reynolds) numbers, had thus been unable to capture the non-axisymmetry of convective flow. Analogous simulations by Käpylä et al. (2010) also concluded that due to non-axisymmetry, convection can transport angular momentum outwards. However, the parameter range considered by these authors is somewhat different to that of Lesur and Ogilvie (2010).

In summary, although the possibility that vertical convection can be one of the mechanisms providing angular momentum transport in non-magnetised discs has gained some evidence recently and the initial results seem encouraging, they should be regarded with caution and considered as tentative, requiring further investigation.

2.2.3 Gravitational instability and spiral density waves

Gravitational instability is caused by disc self-gravity and typically develops on dynamical time-scales. It mainly results in the formation of density enhancements – spiral density waves (SDWs) – that by exerting hydrodynamic and gravitational torques on the disc transport mass inwards and angular momentum outwards. However, the angular momentum transport due to self-gravitating SDWs may not always be described within the framework of an α-prescription, since this is a local description by nature, whereas, depending on disc parameters, the wavelength of SDWs can be comparable to disc radius (i.e., SDWs extending over the whole disc) in which case the transport has a global rather than a local character (see below). Disc self-gravity is also believed to play an important role in the planet formation process: under certain conditions, the non-linear development of gravitational instability, if sufficiently violent, can cause the fragmentation of the disc into bound clumps, that could contract to form planets. Self-gravity dominates angular momentum transport and the disc's secular evolution at an early stage of the disc lifetime, when it is still massive enough (Class I phase), though the influence of self-gravity can sometimes extend into a later T Tauri phase (Hartmann et al., 2006; Kratter et al., 2008; Vorobyov and Basu, 2008). Self-gravity can also play an important role in the dynamics of the 'dead zone'. A mass accumulation in the 'dead zone' due to the accretion from the active zone can activate gravitational instability that, in turn, can subsequently trigger the MRI within it, leading to an enhanced accretion rate. This mechanism is believed to explain the FU Orionis-type outbursts phenomenon (Gammie, 1999; Armitage et al., 2001; Zhu et al., 2009), that is, episodic rises in the accretion rate and therefore in optical light of a young stellar object by 5 mag or more on a relatively short time-scale of $1-10$ yr and slow decays during about 100 yr (Herbig, 1977).

The basic physics underlying gravitational instability in discs is similar to that of the classical Jeans instability of a homogeneous fluid (see, e.g., Binney and Tremaine, 1987). Gravitational forces always tend to enhance any overdensity in a runaway manner by attracting more and more matter (gravitational collapse), whereas pressure forces oppose this. As a result, there exists a certain characteristic length – the Jeans length. If the initial size of an overdense region is larger than the Jeans length, pressure forces are unable to oppose gravitational forces and gravitational instability/collapse ensues. On the other hand, if the size is less than the Jeans length, pressure forces are strong enough to suppress gravitational forces and the perturbation propagates as a sound wave, but with phase speed reduced by self-gravity. In the disc case, one should, in addition, take into consideration rotation and the flattened geometry. A simple and instructive mathematical analysis of the gravitational instability can be done if we assume perturbations (density waves) to have radial wavelengths much smaller than the disc radius (tight-winding approximation) and to be axisymmetric. In this case, the local dispersion relation for density waves in razor-thin (2D) self-gravitating discs is easily obtained (e.g. Goldreich and Tremaine, 1978; Binney and Tremaine, 1987; Lodato, 2007)

$$\omega^2 = c_s^2 k^2 - 2\pi G \Sigma k + \kappa^2, \tag{2.5}$$

where ω and k are, respectively, the frequency and radial wavenumber of a density wave that satisfies $kr \gg 1$ in the tight-winding limit, c_s is the local sound speed in the disc, Σ is the surface density, κ is the epicyclic frequency and G is the gravitational constant. When $\omega^2 > 0$, density waves are oscillatory and freely propagate through the disc. On the other hand, when $\omega^2 < 0$, density waves are unstable and grow exponentially. We can identify stabilising and destabilising terms on the right hand side of Eq. (2.5). The first term, associated with pressure forces, and the third term, related to disc rotation, have a stabilising effect, while the negative second term due to self-gravity is potentially destabilising. Thermal pressure and rotation stabilise the disc, respectively, at short wavelengths (large k) and at long wavelengths (small k) when they both dominate over the self-gravity. The latter can be significant at intermediate wavelengths. To find these wavelengths, we notice that Eq. (2.5) is a simple quadratic form in k, so it is easy to show that if

$$Q \equiv \frac{c_s \kappa}{\pi G \Sigma} < 1,$$

there exists a range of wavenumbers for which the disc is unstable ($\omega^2 < 0$) due to self-gravity.

The non-dimensional parameter Q, first introduced by Toomre (1964) to characterise the strength of self-gravity in disc geometry, plays a central role in the theory of self-gravitating discs. A larger disc surface density and/or smaller sound speed give smaller Q and therefore the disc is more prone to the influence of self-gravity (gravitational instability). The sound speed is, in turn, proportional to temperature, so cooling and heating effects are crucial for the development of gravitational instability, especially in the non-linear regime (see below). It follows from Eq. (2.5) that for $Q < 1$ the largest growth rate corresponds to the minimum of the quadratic form attained at the wavenumber

$$k_J = \frac{\pi G \Sigma}{c_s^2},$$

which is referred to as a Jeans wavenumber in the disc case. It can be said that a density wave with this wavenumber is most subject to the influence of self-gravity, that is, it is most unstable at $Q < 1$. In the opposite case of $Q > 1$, the quadratic form is positive definite, implying that self-gravity cannot overcome stabilising pressure and rotation and the disc is gravitationally stable against axisymmetric perturbations at all wavenumbers. In this case, the propagation speed of density waves is reduced by self-gravity compared to that in non-self-gravitating discs.

The situation is somewhat different for non-axisymmetric, or spiral-like perturbations (i.e., SDWs), which can in fact become gravitationally unstable for larger than unity values of Q. In other words, non-axisymmetric SDWs are more unstable than axisymmetric density waves. Various numerical studies show that, depending on the details of the disc model adopted, SDWs start to emerge typically at $Q \lesssim Q_{cr} = 1.5 - 2$ in the form of multi-armed spirals in the disc surface density and grow on the time-scale of an orbital period in the linear regime (see, e.g., review by Durisen et al., 2007, and references therein). The likely mechanism responsible for the growth of SDWs in discs is the 'swing' amplification that occurs as the wave changes (swings) from leading to trailing due to Keplerian shear (e.g., Goldreich and Lynden-Bell, 1965b; Goldreich and Tremaine, 1978; Toomre, 1981; Sellwood and Carlberg, 1984; Binney and Tremaine, 1987; Kim and Ostriker, 2001; Vorobyov and Basu, 2007, swing amplification is also described in detail in chapter 3). So, the amplified SDWs are trailing and therefore lead, as mentioned above, to efficient transport of angular momentum outwards and mass inwards via hydrodynamical and gravitational stresses (Lynden-Bell and Kalnajs, 1972; Papaloizou and Savonije, 1991; Laughlin and Bodenheimer, 1994).

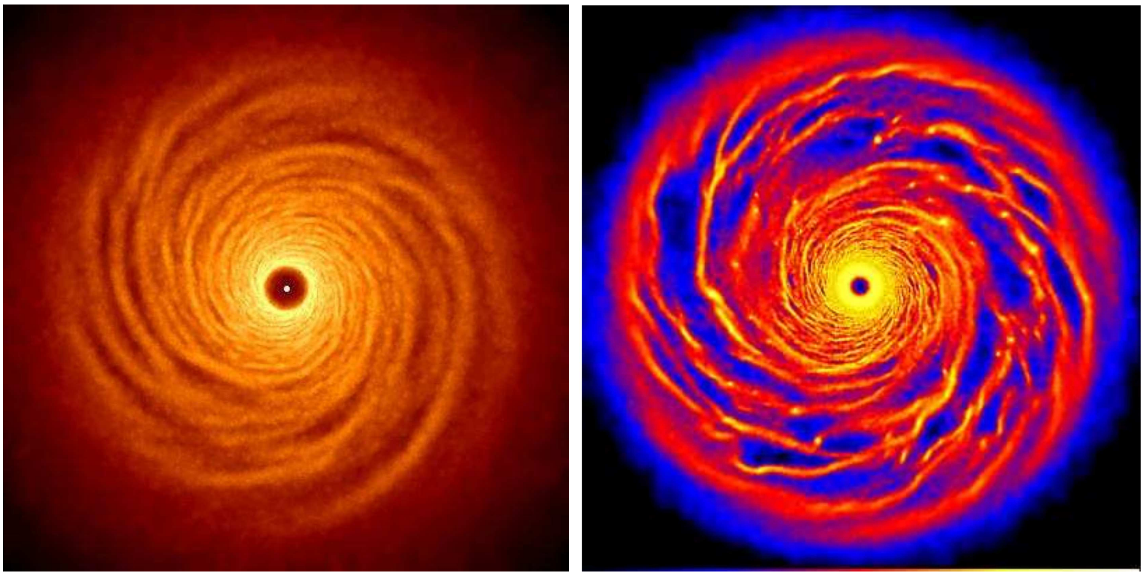

Figure 2.3: Surface density structure of a self-gravitating disc with the mass $M_{disc} = 0.1 M_\star$ in a self-regulated quasi-steady state for the cooling time $\Omega_0 \tau_c = 7.5$ (left), which displays a prominent spiral pattern, and of a fragmenting disc with the same mass for $\Omega_0 \tau_c = 3$ (right). Adapted from Lodato and Rice (2004) and Lodato (2007).

Non-linear evolution of gravitational instability

The subsequent non-linear development of the gravitational instability is largely determined by cooling processes in the disc. Depending on the intensity of cooling, gravitational instability may either cause disc fragmentation or saturate at a non-linear amplitude corresponding to a quasi-steady, self-regulated state. In the beginning, SDWs grow linearly and then, in the non-linear regime, steepen into shocks that produce strong localised heating. The other part of heating comes from compression as well as from the work done by gravity on fluid elements as they sink into the central gravitational potential as a result of inward mass transport. If the disc does not radiate this heat, it will eventually raise the disc temperature, and therefore Q, above the critical value Q_{cr} that will switch off the gravitational instability. But the cooling counteracts heating by reducing the temperature and hence Q that promotes the instability. The outcome from this competition of cooling and heating depends on whether a balance can be reached between these two processes (Bertin and Lodato, 2001). If the cooling is not fast (see below), such a balance is reached and the disc settles into a quasi-steady state where all variables (velocity, density, temperature, etc.) fluctuate around constant average values and Q remains close to the marginal stability threshold, $Q \approx Q_{cr}$. In this case, a nice multi-armed spiral structure due to non-linear SDWs is seen on the disc face that lasts for many orbital periods (Fig. 2.3). Heating and cooling rates are on average the same and the system is in equilibrium. If suddenly the disc is cooled further, Q will decrease and gravitational instability will develop, heating the disc back to marginal stability. On the hand, if the disc is heated more, Q will rise and gravitational instability will be switched off. Nothing preventing the disc from cooling, it will eventually cool down and again switch on gravitational instability, which through generating heat, will bring the disc back to marginal stability. Thus, for slow to moderate cooling, the heating rate due to gravitational instability can adjust to the cooling rate resulting in a thermally regulated, quasi-steady state. The saturated amplitude of SDWs in this state is controlled by the cooling rate (Cossins et al., 2009). By contrast, for sufficiently fast cooling, the balance between heating and cooling is not possible, pressure forces cannot counteract self-gravity and, as a consequence, the disc fragments into several bound clumps instead of reaching a quasi-steady state (Fig. 2.3).

The above-described behaviour of self-gravitating discs was established in earlier numerical simulations by considering a simple/idealised cooling prescription in which the cooling rate per unit volume is given by $\dot{\varepsilon} = -\varepsilon/\tau_c$, where ε is the internal energy per unit volume and τ_c is the constant cooling time (e.g., Gammie, 2001; Rice et al., 2003; Mejía et al., 2005; Rice et al., 2005). These simulations showed that in order for the disc not to fragment and settle into thermally balanced quasi-steady state, the cooling time/rate should satisfy the condition

$$\Omega_0 \tau_c \gtrsim \beta, \tag{2.6}$$

where an exact value of β depends on disc thermodynamics (adiabatic index) and is typically $\beta \simeq 3 - 12$. Condition (2.6) implies that the cooling time must be larger than the dynamical/orbital time in order for self-regulation to set in. In the opposite case $\Omega_0 \tau_c \lesssim \beta$, when the cooling time is shorter than the dynamical time-scale – the characteristic time of gravitational instability, the heat generation rate due to self-gravitating SDWs cannot compensate for the faster cooling that causes further decrease of Q and strengthens the gravitational instability (gas compression in spiral shocks). Eventually, as mentioned above, denser parts of the SDWs break up into clumps with gravitational binding energy exceeding the clump's internal and rotational energies.

Although a simple cooling law allows us to understand the basic dynamical behaviour of self-gravitating discs, for a rigorous and complete analysis one must consider more realistic cooling through radiative (and possibly also convective) transport of dissipated energy from the disc interior to its surface and eventual loss into space. Because discs contain large regions of small and large optical depth, this represents a complex 3D radiative transfer problem. (For example, there can be complications due to a lack of proper knowledge of dust composition, grain size distribution, grain growth and settling, all that affect opacity.) In contrast to a constant cooling time, the radiative cooling time can decrease with increasing temperature (see Durisen et al., 2007). Therefore, short radiative cooling times correspond to high temperatures at which the disc is more stable and less susceptible to fragmentation. So, one can no longer view short cooling times as a necessary criterion for fragmentation. A detailed analytical study of the radiative cooling process by Rafikov (2005, 2007), which also takes into account the possibility of vertical convective heat transport in the disc interior, indicates that the necessary conditions for disc fragmentation due to self-gravity – short ($\sim \Omega_0^{-1}$) effective radiative cooling times and at the same time lower temperatures, so that $Q \lesssim Q_{cr}$ – can occur only at large radii ($\gtrsim$ 100 AU) from the central star for typical parameters of protoplanetary discs; at smaller radii radiative cooling times appear too long, so that fragmentation cannot occur. Contrary to claims by other authors (Boss, 2004; Mayer et al., 2007, see also below), Rafikov (2007) also demonstrated that even convection, when present, does not lead to a significant reduction in the radiative cooling time at smaller radii to permit fragmentation. Vertical convection can only efficiently transport/deposit heat from the midplane to the upper radiative atmosphere, which is optically thinner than the interior. But an eventual loss of heat into space should occur by means of radiative transport through this atmosphere and hence cannot be faster than the atmospheric temperature permits. In the convective case, the temperature in the radiative atmosphere, and therefore the radiative cooling rate, are limited by the adiabatic temperature gradient (at a given midplane temperature). So, the insufficient cooling rate from the upper radiative atmosphere creates a bottleneck preventing the disc from cooling fast enough for fragmentation. In particular, these arguments also suggest that gas giant planet formation by disc instability is unlikely inside about 100 AU, as also confirmed by a body of numerical simulations described below.

Numerical simulations explicitly incorporating radiative cooling bear out the above analytical results on disc fragmentation and further specify, depending on the disc model, a radius beyond which fragmentation is possible. However, the results of different groups still differ. Specifically, a number of 3D global simulations of discs with the typical masses $M_{disc} \lesssim 0.1M_\star$ around $M_\star = 0.5-1M_\odot$ stars and with outer radii a few tens of AU, employing various radiative algorithms,[12] unanimously demonstrate that within about 70 AU the discs generally do

---

[12] These are radiative diffusion approximation in the optically thick disc interior coupled with different treatments of radiative transfer at the interface between this optically thick region and optically thin upper atmosphere, from which radiative losses actually occur. Some of these algorithms take a bolder approach and solve ray equation, approximating the disc's vertical structure as a plane parallel atmosphere.

not radiatively cool fast enough to fragment (e.g., Nelson et al., 2000; Boley et al., 2006; Cai et al., 2006; Boley et al., 2007; Stamatellos and Whitworth, 2008; Boley, 2009; Forgan et al., 2009; Cai et al., 2010; Cossins et al., 2010). Interestingly, in this case, the non-fragmentation criterion (2.6) holds, but for the effective radiative cooling times in the subsequent quasi-steady state rather than for initial cooling times, which can significantly differ from the former (Johnson and Gammie, 2003; Cai et al., 2006). So, in the case of radiative cooling, it is difficult to determine whether a disc will fragment or not due to gravitational instability based on its initial cooling time. However, Johnson and Gammie (2003) and Cossins et al. (2010) noted that the disc can still be susceptible to fragmentation if its temperature at some location is near or falls into the opacity gap – a temperature range at which opacity becomes very small due to dust sublimation and therefore radiative cooling becomes very efficient. But the opacity gap occurs at temperatures $T \sim 10^3$ K that correspond to very small radii in the immediate vicinity of the central star, where Q is high and therefore fragmentation due to small values of opacity can be avoided (see also Rafikov, 2007). On the other hand, Boss (2002, 2004) and Mayer et al. (2007) found fragmentation in their simulations for similar disc parameters. The masses of resulting bound clumps were typically a few Jupiter masses. Because a purely radiative cooling is inefficient, these authors attributed the fragmentation to vertical convection that develops in their radiative disc models and apparently provides fast enough cooling. By contrast, in the above-mentioned radiative hydrodynamic simulations by other authors, vertical convection was either not observed or, if observed, it carried only about ten percent of vertical heat flux insufficient for fragmentation (Boley et al., 2006, 2007). Convection in simulations of Boss (2004) and Mayer et al. (2007) could have also been mistaken for spiral shock bores, or hydraulic jumps, also seen in some simulations of self-gravitating discs (e.g., Boley and Durisen, 2006, see also chapter 5). Such 3D shocks are characterised by upwellings like those in the presence of convection.

Comparisons of employed numerical methods as well as radiative algorithms and thermodynamics were made by these authors in order to reconcile different outcomes of their simulations (Cai et al., 2006; Boley et al., 2007; Boss, 2007, 2008; Cai et al., 2010, see also review by Durisen et al., 2007). These discrepancies boil down to the following issues: different treatments of the boundary between the optically thick and thin regions, which strongly influences radiative cooling; properties of the different equations of state used; different opacity laws. As a result of such detailed comparisons, despite some controversy, currently the prevailing view in the literature is that disc fragmentation due to gravitational instability is an unlikely occurrence within radii of a few tens of AU. However, this does not completely rule out the possibility of fragmentation, which can, in principle, occur at radii larger than $70-100$ AU primarily perhaps during an early stage of evolution, when discs are still massive and extended (e.g., Vorobyov and Basu, 2006; Clarke, 2009; Boley, 2009; Boley et al., 2010; Rice et al., 2010; Vorobyov and Basu, 2010). At these early phases, discs are still accreting material from their parent cloud that was shown to facilitate fragmentation. In fact, as suggested recently by Dodson-Robinson et al. (2009), gravitational instability may well be responsible for the formation of recently detected high planetary-mass objects orbiting at large separations from their host stars (Kalas et al., 2008; Marois et al., 2008), although there has been some criticism of this scenario (Kratter et al., 2010).

Angular momentum transport by gravitational instability

The question of whether angular momentum transport by non-axisymmetric SDWs in self-gravitating discs is a local or global phenomenon and if it can be modelled by a simple α-prescription (i.e., by a single transport parameter α) was also investigated by means of numerical simulations (e.g., Laughlin and Bodenheimer, 1994; Laughlin and Rozyczka, 1996; Gammie, 2001; Lodato and Rice, 2004, 2005; Mejía et al., 2005; Boley et al., 2006; Cossins et al., 2009, see also review by Lodato, 2007). The general conclusion from these studies is as follows. For light discs with masses $M_{disc}/M_\star \lesssim 0.25$ in a quasi-steady self-regulated state, angular momentum transport induced by SDWs is of a local nature and, to a good approximation, is described in terms of the α-prescription. The locality of transport is attributable to the fact that in low mass discs, many SDW modes with higher azimuthal wavenumbers m are excited and are responsible for most of the transport (Gammie, 2001; Lodato and Rice, 2004; Lodato, 2007). So, the contribution of the non-local part due to low-m modes to the total transport appears to be small compared to the local one. In this case, the corresponding α, consisting of hydrodynamical and gravitational stresses, can be found from the energy balance condition (Gammie, 2001; Lodato, 2007, see also chapter 3)

$$\alpha = \frac{1}{\gamma(\gamma-1)\Omega_0\tau_c}\left|\frac{d\ \ln\ \Omega_0}{d\ \ln\ r}\right|^{-2}, \tag{2.7}$$

which simply states that the energy dissipation due to self-gravity, represented by α-viscosity, locally matches cooling. This relation agrees well with the total stress computed directly from simulations using the same constant cooling time. Equation (2.7) also holds reasonably well in the presence of more realistic radiative cooling. In this case, the corresponding cooling time is found as a ratio of the vertically averaged internal energy to the actual radiative cooling rate (Boley et al., 2006, these authors found α of order 0.01). Alternatively, when the effective cooling time is somewhat difficult to define (e.g., in the case of isothermal or polytropic equation of states), the viscous α in the self-regulated state can be deduced by comparing the secular evolution of the surface density as given by simulations to that predicted from the viscous evolution equation (2.3) and fitting the value of the viscosity to get an agreement between these two evolutions (Laughlin and Bodenheimer, 1994; Laughlin and Rozyczka, 1996). Such a comparison was also made for radiatively cooled discs by Boley et al. (2006) giving about the same value as in the first case. By contrast, for more massive discs with $M_{disc}/M_\star \gtrsim 0.25$ the local approximation breaks down due to the domination of low-m SDW modes in the total stress and consequently transport, being of a global nature, can no longer be described within the α-formalism. In these heavy discs, gravitational instability occurs as recurring transient episodes and is stronger than that in lighter discs, so that quasi-steady state/self-regulation is not quite reached (Lodato and Rice, 2005; Lodato, 2007). In this case, the gravitational torques can be very large ($\alpha \approx 0.5$) and cause fast redistribution of disc gas.

It follows from Eq. (2.7) that a minimum value of cooling time preventing fragmentation corresponds to a maximum value of α. Based on this, condition (2.6) can be interpreted in a different way, namely, in terms of how much angular momentum can be transported by SDWs in a quasi-steady state (Rice et al., 2005). It was demonstrated that depending on disc thermodynamics (adiabatic index), the fragmentation threshold β is different, but interestingly the maximum α that can be provided by self-gravity in a quasi-steady state appears to be independent of the adiabatic index and has approximately a constant value $\alpha_{max} \approx 0.06$. The values of α exceeding α_{max} cannot be sustained by the disc, because they correspond to short cooling times and hence to stronger torques, which self-gravity is no longer able to produce. As a result, a quasi-steady state cannot be achieved and the disc undergoes fragmentation.

In summary, after the MRI, gravitational instabilities are considered, as the second most likely candidate driving angular momentum transport in protoplanetary discs, especially at early times. Moreover, in certain cases (e.g., in episodic accretion events and outbursts) these two agents can also operate in discs in concert (e.g., Menou and Quataert, 2001; Fromang, 2005; Zhu et al., 2009).

2.2.4 Vortices

Vortices in protoplanetary discs have recently attracted much interest both because they can represent aggregation regions of solids for the formation of planetesimals and more generally for the understanding of accretion process and the basic problem of angular momentum transport in magnetically inactive discs. The formation and evolution of coherent vortex structures in a strongly sheared Keplerian flow was studied through 2D and 3D numerical simulations. In thin (2D) discs, both the barotropic case, where the potential vorticity (PV) – a basic quantity characterising vortices – is conserved, and the baroclinic case, where the generation of PV is possible due to inhomogeneities in the temperature (entropy) distribution in the disc, were considered. We emphasise that all these simulations of vortices that we also mention below have been performed ignoring disc self-gravity. The effects of self-gravity on the vortex dynamics will be addressed in chapter 3, but prior to that we describe the basic properties of vortex evolution in *non-self-gravitating* discs gathered from these simulations.

It was demonstrated both in 2D global (Bracco et al., 1999; Godon and Livio, 1999a,b, 2000; Davis et al., 2000; Li et al., 2001; Davis, 2002; Bodo et al., 2007; Klahr and Bodenheimer, 2003; Petersen et al., 2007b) and in the local shearing sheet simulations (Umurhan and Regev, 2004; Barranco and Marcus, 2005; Johnson and Gammie, 2005b; Shen et al., 2006; Lithwick, 2007; Oishi and Mac Low, 2009; Lesur and Papaloizou, 2010) that only anticyclonic vortices (circulating in the same sense as the background shear) can survive in discs for hundreds of orbital periods and can be thought of as long-lived stable structures. On the other hand, cyclonic vortices (circulating opposite to the shear) get sheared out by the differential rotation of the disc and quickly disappear. Simulations are often initiated either with random PV perturbations, containing positive (cyclonic) and negative (anticyclonic) values in equal portions, or with an imposed single vortex. The former shows that initial small scale anticyclonic regions with negative PV at first get sheared into strips, which then wrap up into small scale vortices (Fig. 2.4) because of the specific instability discussed below. These small vortices gradually grow in size via merging into each other due to inverse cascade of energy specific to 2D flows until eventually their size becomes of order the disc scale height (Johnson and Gammie, 2005b); growth beyond this scale is restricted by compressibility effects.[13] This limitation on the vortex size due to compressibility was also confirmed by Bodo et al. (2007) through detailed 2D simulations of a single vortex in a Keplerian disc. In these simulations, a vortex with an initial length-scale exceeding the disc scale height undergoes non-linear adjustment during which it decreases in size, radiating excess energy in the form of SDWs [14] (shocks) and finally acquires length-scales comparable to the disc scale height. These final scales are independent of the initial vortex size and are determined by the disc

---

[13] This implies that, although the 2D analysis is marginally valid, still a fuller 3D treatment would be more appropriate in this case.

[14] In the linear regime, SDWs in non-self-gravitating discs are described by the same dispersion relation (2.5), but without the negative self-gravity term (e.g., Balbus, 2003).

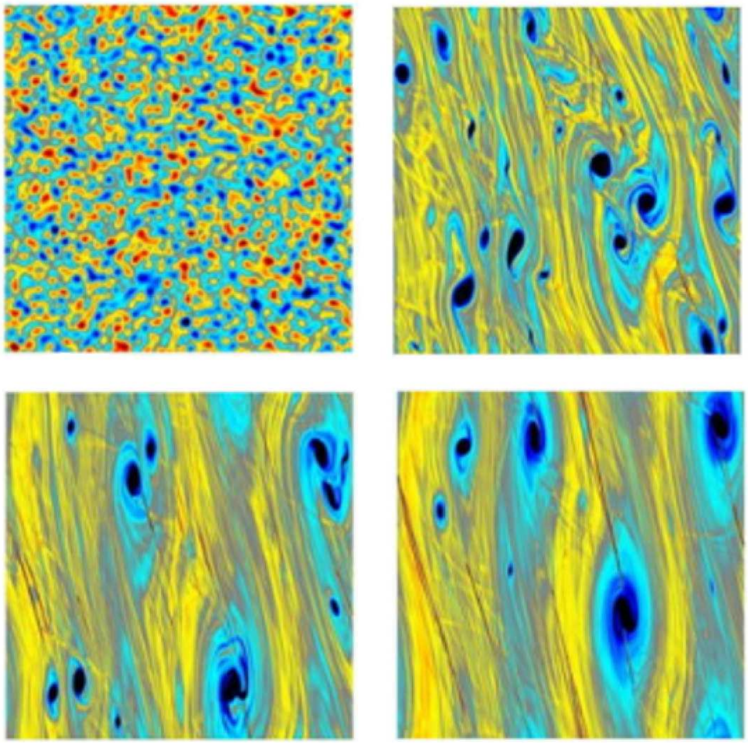

Figure 2.4: Formation and evolution of vortices in a non-self-gravitating disc from the initial random distribution of PV. The initial conditions are shown in the upper left panel and the other panels trace evolution in the lexicographic order. This simulation is performed in the shearing sheet (i.e., is two-dimensional). Blue regions correspond to negative PV (anticyclonic vortices) and sheared out yellow strips to positive PV (cyclonic) regions. We see that initial small-scale anticyclonic vortices merge into each other with time and form larger vortices. Adapted from Johnson and Gammie (2005b).

properties (sound speed and the scale height). By contrast, the initial small scale cyclonic regions instead of wrapping up into distinct vortices, get strongly sheared into strips and remain so during the entire course of evolution.

The emission of SDWs by surviving anticyclonic vortices during the adjustment process is due to the background Keplerian shear that provides coupling between these two modes/types of perturbations (see also section 2.4), when the horizontal length-scale of vortices is equal to or larger than the disc scale height (Li et al., 2001; Davis, 2002; Johnson and Gammie, 2005a,b; Bodo et al., 2005; Mamatsashvili and Chagelishvili, 2007; Heinemann and Papaloizou, 2009a). This implies that one can employ the thin-disc approximation, or the 2D treatment for studying vortices, but at the same time must also take into account effects of compressibility, which play an important role at such length-scale. The emitted SDWs steepen into shocks afterwards and thus vortices appear to generate shocks. The significance of this phenomenon is that these shocks of SDWs accompanying vortices tend to greatly enhance outward transport of angular momentum due solely to vortices. It was shown that the value of $\alpha \sim 10^{-3} - 10^{-2}$, which is associated mostly with shocks of SDWs (i.e., with compressibility), is about two orders of magnitude larger than that provided solely by vortices in the absence of compressibility (Johnson and Gammie, 2005b; Shen et al., 2006; Oishi and Mac Low, 2009). However, if there is no mechanism maintaining vorticity, this value of angular momentum transport, although appropriate for discs, still slowly decays with time as $t^{-1/2}$ due to shock dissipation. By contrast, similar incompressible 2D simulations by Umurhan and Regev (2004) yielded non-decaying coherent anticyclonic vortices (providing extremely small transport). But incompressible 2D simulations are less realistic, because they exhibit the growth of vortex size beyond the scale height – the regime where compressibility is important.

The basic underlying mechanism/instability responsible for the development of vortices as well as the necessary criterion were identified by Lithwick (2007) for incompressible shear flows. He interpreted it as a non-linear Kelvin-Helmholtz instability of an axisymmetric vortex strip with vorticity of the same sign as the background vorticity (for incompressible flows vorticity and PV are equivalent); for a vortex strip with the opposite sign of vorticity no such instability is observed. This readily explains why cyclonic vortices are unable to survive. This mechanism of vortex formation should be extended to compressible shear flows as well. Note that in this scenario one needs PV perturbations (e.g., vortex strips or random type) initially present in a disc. Therefore, there should exist some mechanism that generates PV in discs and one of the possibilities is the described below baroclinic and Rossby wave instabilities due to the radial gradient of temperature (entropy) in discs.

The situation in the more realistic 3D case is more complicated as far as vortex survival is concerned. Shen et al. (2006) and Lithwick (2009) carried out 3D non-linear simulations of vortices in the unstratified shearing box with Keplerian shear. Shen et al. (2006) initially imposed homogeneous in the vertical direction, but random in the horizontal plane, vortex columns. Subsequent evolution showed that these vortex columns get strongly deformed and eventually destroyed by the elliptical instability (Kerswell, 2002) if their vertical extent exceeds the horizontal one. The elliptical instability is a parametric instability arising when a multiple of the vortex turnover frequency becomes equal to the inertial wave frequency, which leads to resonance destroying 3D vortices (see also Lesur and Papaloizou, 2009). In this case, the corresponding kinetic energy and α undergo rapid (faster than that in 2D) decay with time, not resulting in any self-sustaining turbulent state.

In an analogous study, Lithwick (2009) concentrated on the dynamics of a single vortex column alone in order to more fully understand the vortex behaviour in 3D. He demonstrated both analytically and by means of (incompressible) simulations that in order for a vortex to survive, its azimuthal length-scale must be larger than the disc scale height. When this criterion is violated, as mentioned above, the vortex resonantly couples with inertial waves and therefore decays because of the elliptical instability. It was also shown that the ratio of the azimuthal length-scale of the vortex to its radial one (the vortex aspect-ratio) is inversely proportional to its vorticity. So, for a fixed radial size of the vortex, weaker vortices have larger azimuthal extent/aspect-ratio (i.e., are more azimuthally elongated) and can more easily survive than stronger ones that are more circular and therefore subject to destruction. Based on these results, Lithwick suggested that the vortices got destroyed in the simulations of Shen et al. (2006), because the corresponding vorticity was strong.

Lesur and Papaloizou (2009) carried out a more detailed analysis of the elliptical instability of a 3D vortex embedded in a Keplerian disc both for unstratified and stratified cases. In the unstratified case, it was found that vortices with aspect-ratios of the order of unity (i.e., strong vortices) are most subject to the elliptical instability, with a growth time comparable to the orbital/dynamical time in agreement with the results of Shen et al. (2006) and Lithwick (2009). On the other hand, for relatively weak elongated vortices with larger aspect-ratios, the growth time of the elliptical instability turned out to be very long, of the order of a hundred disc orbital times. This is also consistent with the result of Lithwick (2009) that weaker vortices tend to survive longer. Nevertheless, they are still unstable, whereas the simulations of Lithwick (2009) yielded stable vortices with large aspect-ratios. The reason for this discrepancy can be different numerical dissipations in the codes used by these authors, because, as Lesur and Papaloizou pointed out, the development of the elliptical instability for weak vortices in simulations is sensitive to numerical resolution.

In vertically stratified discs, Lesur and Papaloizou (2009) also found that the influence of vertical stratification is important mainly for weak vortices (i.e., at large aspect-ratios). For such vortices, the growth rate of the elliptical instability decreases significantly with increasing stratification. Thus, stratification considerably weakens the elliptical instability, but does not suppress it completely. Barranco and Marcus (2005) focused on the long-term non-linear development of the elliptical instability of a single columnar vortex through incompressible simulations in the shearing box with isothermal vertical structure. Subsequent evolution showed that this tall vortex with several scale heights is destroyed due to the elliptical instability. As a result, first a shorter columnar vortex forms with height equal to 1-2 disc scale heights. Then, this secondary vortex becomes unstable due to the same mechanism and splits into two vortices located at 1-3 scale heights above and below the midplane. These off-midplane vortices are able to survive for hundreds of orbital periods. Interestingly, the mechanism that creates these off-midplane vortices does not seem to necessarily 'need' the vorticity of the short vortex. Most likely the necessary vorticity for off-midplane vortices comes from baroclinic effects as inertia-gravity waves, generated during the destruction of the short vortex, travel upwards and downwards from the midplane through the vertically stratified atmosphere. Thus, vertical stratification is essential for vortex survival. However, the value of $\alpha \sim 10^{-5}$ reported in these incompressible simulations is very small, perhaps due to the absence of compressibility.

Sources of vorticity in protoplanetary discs can be radial inhomogeneities of temperature (entropy), which through Rossby wave and/or baroclinic instabilities lead to the generation of PV and hence vortical structures. Lovelace et al. (1999) and Li et al. (2000) showed, using a linear stability analysis, that local sharp changes ('bumps' or 'jumps') in the radial distribution of entropy/temperature in the disc can give rise to the Rossby wave instability. In discs, such variations in the radial structure can arise, for example, in the following way (Lovelace et al., 1999): as mass accumulates in the disc by accretion, the density and therefore opacity rises resulting in the trapping of heat and in the increase in temperature/entropy at this location. Depending on the heating, opacity, and mass accretion rate, there will be a radius where a local gradient (a local extremum) of entropy is sufficient to trigger the Rossby wave instability. The conditions for the Rossby wave instability may also occur near the borders of a 'dead zone', because the latter is characterised by a smaller accretion rate than the rest of a disc (see e.g. Lyra et al., 2009). Li et al. (2001) further demonstrated vortex formation by following the non-linear development of the Rossby wave instability in 2D numerical simulations. It was also shown that these vortices, as expected, generate SDWs and hence transport angular momentum outwards with corresponding α in the range $10^{-4} - 10^{-2}$.

Klahr and Bodenheimer (2003) and Petersen et al. (2007a,b), on the other hand, paid attention to the fact that in discs, the global distribution of entropy/temperature with radius is declining (negative radial entropy gradient) and can give rise to the global baroclinic instability, also generating vorticity in the disc flow (see also Klahr, 2004; Johnson and Gammie, 2005a, 2006; Tevzadze et al., 2010). However, it has a relatively small growth rate, of the order of a hundred orbital periods. These authors followed the baroclinic instability in the non-linear regime and found that it leads to the emergence of anticyclonic vortices, which are sustained for long times. The outward angular momentum transport, as before, is mainly due to shocks of SDWs generated by these vortices, giving $\alpha \sim 10^{-3}$ in these simulations. Lesur and Papaloizou (2010) further clarified the nature of the baroclinic instability in discs by showing that it is in fact subcritical, that is, occurring only in the non-linear regime (at finite amplitudes)[15], when both the radial entropy gradient is weakly unstable according to Schwarzschild criterion (for radial buoyancy) and the cooling time is short (or equivalently, thermal diffusivity is strong). Lesur and Papaloizou also found that in 3D, the baroclinic instability develops similarly as in 2D. The vortices produced are overall stable (i.e., are not destroyed) against the elliptical instability, which induces only vertical structures (motions) localised inside the vortex core. In other words, thanks to baroclinicity, vortices can survive in 3D despite the elliptical instability.

Anticyclonic vortices also have an important property of trapping dust particles (e.g., Barge and Sommeria, 1995; Chavanis, 2000; Godon and Livio, 2000; Johansen et al., 2004; Inaba and Barge, 2006; Klahr and Bodenheimer, 2006; Lyra et al., 2009). It was shown that a smooth, sufficiently long-lived vortex is indeed able to effectively trap dust particles in its core via exerting aerodynamic drag on dust grains, possibly accelerating planetesimal formation.

In summary, vortices are regarded as an important candidate in the angular momentum transport process in non-magnetised discs and possible sites of solid aggregation. However, their detailed 3D behaviour (survival) is still a subject of future investigations. As we have seen in the above cases, the majority of angular momentum transport induced by vortices is actually related to SDWs emitted by these vortices. As noted above, this vortex-wave coupling occurs due to the Keplerian shear and will be discussed further in section 2.4, because one of the goals of this thesis is to properly understand and characterise this phenomenon when other factors (e.g., self-gravity) are included.

2.3 Method of analysis – the local shearing sheet/box approach

All the above-described dynamical processes in discs have characteristic length-scales much smaller than the disc's radial size. To study such small-scale processes, Goldreich and Lynden-Bell (1965b) devised a local approximation – shearing sheet/box formalism ('sheet' and 'box' refer to the 2D and 3D cases, respectively), which has been employed extensively since then

---

[15] This is one of the reasons why the baroclinic instability and resulting vortices were not observed at small initial amplitudes of perturbations in the related simulations by Johnson and Gammie (2006).

for studies of various processes in discs (see also Hawley et al., 1995). In other words, the shearing sheet/box approach is suitable for analysing perturbations (instabilities) whose wavelengths are small compared with the distance to the disc centre. From the viewpoint of numerical simulations of various processes in discs, the shearing box/sheet approximation has an advantage over global disc models in that it allows much higher numerical resolution. Because we also adopt this formalism as a framework in the thesis, in this section we give its derivation following Goldreich and Lynden-Bell's paper.

The essence of this local approximation is as follows. Let us first write the basic hydrodynamic equations in cylindrical coordinate frame (r, ϕ, z) with the origin at the disc centre and rotating with an angular velocity Ω equal to the disc's angular velocity, $\Omega_0(r)$, at some fiducial radius r_0 from the centre, $\Omega = \Omega_0(r_0)$ (we remind that $\Omega_0(r) \propto r^{-3/2}$ corresponding to the disc's Keplerian differential rotation). The disc midplane lies in the (r, ϕ)-plane. In this rotating frame, the equation of motion takes the form

$$\frac{\partial \mathbf{U}}{\partial t} + (\mathbf{U}\cdot\nabla)\mathbf{U} + 2\Omega\mathbf{e}_z \times \mathbf{U} - \Omega^2\mathbf{r} = -\frac{1}{\rho}\nabla p - \nabla\psi_c - \nabla\psi, \tag{2.8}$$

where $\mathbf{U}$ is the gas velocity relative to the rotating frame, $\mathbf{e}_z$ is the unit vector along the vertical z-axis, $\mathbf{r} = r\mathbf{e}_r$ is the radius vector from the rotation axis with $\mathbf{e}_r$ being the radial unit vector, ρ is the gas density, p is the gas pressure, ψ_c and ψ are, respectively, the gravitational potentials of the central star and disc self-gravity. Further, we decompose the total velocity as

$$\mathbf{U} = \mathbf{u_0} + \mathbf{u},$$

where $\mathbf{u_0} = r(\Omega_0(r) - \Omega)\mathbf{e}_\phi$ is the equilibrium rotational velocity of the disc in the rotating frame and $\mathbf{u}$ is the perturbed velocity. Substituting this into Eq. (2.8) and using the identity $\mathbf{e}_r = \mathbf{e}_\phi \times \mathbf{e}_z$, we obtain

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u_0}\cdot\nabla)\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u_0} + 2\Omega\mathbf{e}_z \times \mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} = \left(\frac{u_0^2}{r} + 2\Omega u_0 + \Omega^2 r\right)\mathbf{e}_r - \frac{1}{\rho}\nabla p - \nabla\psi_c - \nabla\psi. \tag{2.9}$$

In the axisymmetric equilibrium state, $\mathbf{u} = 0$ and only the gravitational potential of the central star, $\psi_c = -GM_\star(r^2 + z^2)^{-1/2} \approx -GM_\star/r$, dominates in the radial direction (since the disc is thin, $H \ll r$, and relatively light, $M_{disc} \ll M_\star$) resulting in Keplerian rotation, so that

$$\left(\frac{u_0^2}{r} + 2\Omega u_0 + \Omega^2 r\right) = \frac{\partial \psi_c}{\partial r}.$$

Using this in Eq. (2.9) we get

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u_0}\cdot\nabla)\mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u_0} + 2\Omega\mathbf{e}_z \times \mathbf{u} + (\mathbf{u}\cdot\nabla)\mathbf{u} = -\frac{1}{\rho}\nabla p - \frac{\partial \psi_c}{\partial z}\mathbf{e}_z - \nabla\psi. \tag{2.10}$$

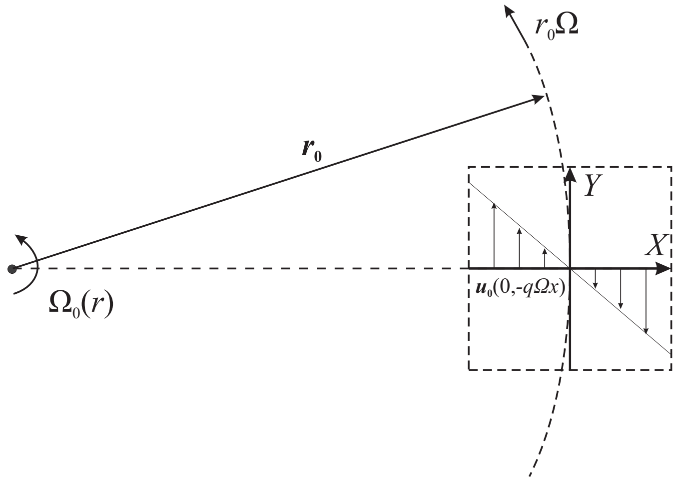

Figure 2.5: Schematic illustration of the local coordinate frame (X, Y, Z) (face-on view), which co-rotates with the disc matter at the distance r_0 with velocity $r_0\Omega$, where $\Omega = \Omega_0(r_0)$. In this local frame, the disc's Keplerian differential rotation appears as a parallel azimuthal shear flow $\mathbf{u_0} = (0, -q\Omega x, 0)$.

In this rotating cylindrical coordinate frame, we want to analyse small-scale perturbations. For this purpose, we now concentrate on a small patch of the disc in the vicinity of some point $(r_0, \phi_0, 0)$ located at the fiducial radius r_0 and set up a new local Cartesian reference frame centred at this point with the $x-$, $y-$ and $z-$axes pointing, respectively, towards the radial, azimuthal and vertical directions, $(x = r - r_0, y = r_0(\phi - \phi_0), z)$ (Fig. 2.5). Since this frame is local, $x/r_0, y/r_0, z/r_0 \ll 1$ and we can expand $\mathbf{u_0}$ in powers of x/r_0 about the origin

$$\mathbf{u_0} = r_0^2\Omega_0'(r_0)\left[\left(0, \frac{x}{r_0}, 0\right) + O\left(\frac{x^2}{r_0^2}\right)\right], \tag{2.11}$$

where prime over Ω_0 denotes derivative with respect to radius. A zero order term is absent, because by our choice $\Omega = \Omega_0(r_0)$ and therefore $\mathbf{u_0}(r_0) = 0$. $r_0\Omega_0'(r_0) = -q\Omega_0(r_0) = -q\Omega$, where the parameter q, characterising the shear of differential rotation, is $q = 1.5$ for a Keplerian profile (because we restrict ourselves to Keplerian rotation throughout the thesis, we will use this value and notation for the shear parameter from now on). Thus, in the local frame, Keplerian rotation of the disc manifests itself as a parallel azimuthal flow with a constant velocity shear q in the radial direction, $\mathbf{u_0} = (0, -q\Omega x, 0)$.

Substituting expansion (2.11) into Eq. (2.10), which is in cylindrical coordinates, we finally obtain the momentum equation in the rotating local frame

$$\frac{\partial \mathbf{u}}{\partial t} - q\Omega x\frac{\partial \mathbf{u}}{\partial y} + (\mathbf{u}\cdot\nabla)\mathbf{u} - q\Omega u_x\mathbf{e}_y + 2\Omega\mathbf{e}_z \times \mathbf{u} + \Omega^2 z\mathbf{e}_z = -\frac{1}{\rho}\nabla p - \nabla\psi, \tag{2.12}$$

where in deriving this equation, the terms of order x/r_0 have been neglected while those of order $x/l \sim O(1)$, where l is the typical length-scale of perturbations (satisfying $l \ll r_0$, small-scale perturbations), have been kept. As a result, the gradient operator ∇ is simplified from cylindrical to cartesian form, $\nabla = (\partial/\partial x, \partial/\partial y, \partial/\partial z)$. Accordingly, Poisson's equation for self-gravity in the local frame takes the form

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\psi = 4\pi G\rho. \tag{2.13}$$

In a similar manner, we can derive the continuity equation in the local frame using expansion (2.11),

$$\frac{\partial \rho}{\partial t} - q\Omega x\frac{\partial \rho}{\partial y} + \nabla \cdot (\rho \mathbf{u}) = 0. \tag{2.14}$$

Thus, the local approximation is an expansion of basic equations with respect to the small parameter l/r_0. Equations (2.12-2.14) correspond to zero order in this parameter and therefore curvature terms due to cylindrical geometry (which sometimes unnecessarily complicate dynamics) are neglected in these equations, while terms related to velocity shear, which are essential for the perturbation dynamics in differentially rotating discs, are retained. Dynamical equations (2.12) and (2.14) together with Poisson's equation (2.13) constitute a full set of shearing box equations.

We also note that in this local approximation, the unperturbed density ρ_0 can be assumed to be approximately independent of the radial coordinate x (there is no y-dependence as well, because of axisymmetric equilibrium). To show this, we write

$$\rho_0(r, z) = \rho_0(r_0, z) + x\rho_0'(r_0, z) + O(x^2/r_0^2)$$

and because the equilibrium density varies on length-scales comparable to the disc radius, $\rho_0' \sim \rho_0/r_0$, to zero order in l/r_0, in which the above equations have been derived, the equilibrium density does not vary with x to this order within the local frame. Similarly, the equilibrium pressure p_0 and ψ_0, also varying on a large, disc radius scale, appear independent of x (and of y) in the local frame.

In certain situations, the vertical variation of quantities are not important and they can be assumed to be independent of z (e.g., when the horizontal length-scale of perturbations, l, is larger than the disc scale height but still much smaller than r_0, $H \lesssim l \ll r_0$, see also chapter 5). In this case, the 3D density and pressure in Eqs. (2.12-2.14) are replaced with the vertically integrated 2D surface density $\Sigma = \int_{-H}^{H} \rho dz$ and pressure $P = \int_{-H}^{H} p dz$. In this way, we obtain the 2D shearing sheet equations

$$\frac{\partial \mathbf{u}}{\partial t} - q\Omega x\frac{\partial \mathbf{u}}{\partial y} + (\mathbf{u}\cdot\nabla)\mathbf{u} - q\Omega u_x \mathbf{e}_y + 2\Omega \mathbf{e}_z \times \mathbf{u} = -\frac{1}{\Sigma}\nabla P - \nabla\psi, \tag{2.15}$$

$$\frac{\partial \Sigma}{\partial t} - q\Omega x\frac{\partial \Sigma}{\partial y} + \nabla \cdot (\Sigma \mathbf{u}) = 0. \tag{2.16}$$

and Poisson's equation

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\psi = 4\pi G\delta(z)\Sigma, \tag{2.17}$$

where the operator $\nabla = (\partial/\partial x, \partial/\partial y)$ is now 2D and $\delta(z)$ is Dirac's delta function. These shearing box (2.12-2.14) and shearing sheet (2.15-2.17) equations will be the basis of our subsequent analysis and will be used throughout the thesis.

2.4 Objectives and outline of the thesis

As mentioned in section 2.2.1, the operators governing the linear dynamics of perturbations in flows with inhomogeneous kinematics are non-normal due to the shear of the mean velocity profile and therefore the corresponding eigenfunctions are non-orthogonal and strongly interfere (e.g., Reddy et al., 1993; Trefethen et al., 1993). So, to form a complete picture of the temporal evolution of perturbations in shear flows, one should fully analyse the superposition/interference of the eigenfunctions, which is a rather daunting task. The non-normality of shear flows and its dynamical consequences were well understood and precisely characterised by the hydrodynamic community in the 1990s (see Schmid and Henningson, 2000). In the standard modal analysis, i.e., spectral expansion of perturbations in time and then examination of eigenfrequencies and eigenfunctions, commonly employed in hydrodynamics (Drazin and Reid, 1981), the main focus is only on the flow stability at asymptotically large times, while no attention is directed to any particular initial value or intermediate/finite-time aspect of the perturbation dynamics. In this way, the need to compute the effects of eigenfunction interference, necessary for understanding transient phenomena, is avoided in the modal treatment. Instead, in the 1990s, with revealing the non-normal nature of shear flow operators, emphasis was shifted from the analysis of flow stability at asymptotically large times to the study of finite-time (transient) behaviour. Accordingly, new mathematical methods – the so-called non-modal approach – had been developed (see e.g., Trefethen et al., 1993; Schmid and Henningson, 2000) to correctly account for the perturbation dynamics resulting from the non-normality of shear flows. The non-modal approach leads to an initial value problem enabling us to trace a full temporal evolution of perturbations and in that respect it is advantageous over the modal approach, which leads to a boundary value problem and describes perturbation behaviour (stability) only at large times. It was demonstrated that as a consequence of the non-orthogonality of eigenfunctions, a strong linear growth of perturbations can occur at intermediate times in asymptotically/spectrally stable (according to Reyleigh's 1880 criterion, that is, without an inflection point in the velocity profile) hydrodynamic shear flows (e.g., Gustavsson, 1991; Butler and Farrell, 1992; Reddy and Henningson, 1993; Farrell and Ioannou, 1993, 2000). So, in this case the non-normality determines energy exchange (extraction) between the mean flow and perturbations. As noted above, such a transient growth plays a key role in a subcritical transition to turbulence in spectrally stable shear flows.

Another important phenomenon stemming from the non-normality is the coupling in the linear regime among different perturbation modes existing in shear flows (e.g., Chagelishvili et al., 1996, 1997; Farrell and Ioannou, 2000; Vanneste and Yavneh, 2004; Tevzadze, 2006; Bodo et al., 2001; Vanneste, 2008; Bakas and Farrell, 2009). Modes with different time- and length-scales, that would evolve independently in the absence of shear, interact and exchange energy with each other during a limited time interval of their evolution when the background flow velocity has some shear. In other words, shear couples linear perturbation modes and leads to their mutual transformations. This process is also called a linear conversion of modes and is a general phenomenon intrinsic to flows with inhomogeneous kinematics. Traditionally, coupling among perturbation modes is attributed to non-linear effects, but in the presence of background shear, it manifests itself even in the linear regime. In other words, mode coupling in shear flows is a first order effect and plays an important role in the subsequent non-linear development of perturbations and largely defines the characteristics of the resulting non-linear state (turbulence). Like the transient amplification phenomenon, it is also associated with the non-normality and hence, having a finite-time character, cannot be captured within the framework of the modal approach, which looks only into asymptotic dynamics of perturbations. Thus, for a proper understanding of energy exchange processes in the non-linear regime, it is necessary to first analyse all possible couplings and energy exchange channels among modes within the linear theory.

In section 2.3, we have demonstrated mathematically that the dynamics of small-scale perturbations in differentially rotating discs is governed by equations similar to those in parallel shear flows. Therefore, we should expect that similar dynamical processes take place in discs as well (partly we have already seen this in section 2.2.1). In other words, differentially rotating discs are likely to share many properties with hydrodynamical (laboratory) shear flows regarding dynamical behaviour of perturbations that, in turn, allows us to apply a vast knowledge accumulated from studies of the latter to understanding disc flow dynamics (see e.g., Richard and Zahn, 1999; Longaretti, 2002). Indeed, as outlined above, the concepts of non-normality-induced transient amplification of perturbations and based on it the 'bypass' scenario of subcritical turbulent transition have already been invoked in the disc turbulence problem. Recently, it has also been demonstrated that another manifestation of the non-normality – linear mode coupling phenomenon – can take place in astrophysical discs as well in the presence of various physical factors specific to discs: radial and vertical stratification, self-gravity, MRI-turbulence, etc. (Tevzadze et al., 2003; Bodo et al., 2005; Johnson and Gammie, 2005a,b; Bodo et al., 2007; Mamatsashvili and Chagelishvili, 2007; Tevzadze et al., 2008; Mamatsashvili and Rice, 2009; Heinemann and Papaloizou, 2009a,b; Tevzadze et al., 2010; Shtemler et al., 2010). In this case, the above-described means of angular momentum transport – SDWs, vortices and convection, which actually represent different types/modes of perturbations existing in discs – are expected to become dynamically coupled due to shear/non-normality associated with the disc's differential rotation. This, in turn, allows us to shed a new light on the dynamics of these modes and, in particular, on their capability to transport angular momentum. Specifically, as already mentioned in section 2.2.4, it was shown that because of the shear-induced linear mode coupling phenomenon, vortices can excite SDWs in discs. As evident from sections 2.2.3 and 2.2.4, the role SDWs play in the dynamics of both self-gravitating and non-self-gravitating discs cannot be overestimated. Significantly, SDWs are able to enhance the angular momentum transport rate, as we have seen in the case of vortices. In addition, it was also suggested that gravitational forces due to stochastic density perturbations associated with SDWs in a turbulent disc flow, may be important in the migration of low-mass planets (Nelson and Papaloizou, 2004; Nelson, 2005). SDWs also play a role in the dynamics and angular momentum transport in dead zones (Fleming and Stone, 2003; Oishi et al., 2007; Oishi and Mac Low, 2009). So, it is crucial to identify and analyse the generation mechanisms of SDWs, which is one of the goals of the present work. Namely, we will show later that under certain conditions, vertical convection in discs can also excite SDWs due to the same shear-induced mode coupling phenomenon.

Thus, in this thesis, we are motivated by the recent progress and developments in the dynamical behaviour of perturbation modes – SDWs, vortices and convection – that can be responsible for angular momentum transport in non-magnetised, neutral (regions of) discs, and also by an identification of an important role played by shear of the disc's differential rotation in mode dynamics. We further investigate some new effects brought about by shear of disc flow and its self-gravity in this dynamical picture that have not been considered and analysed in previous studies (the Introduction at the beginning of each chapter broaches the specific dynamical processes that we intend to study). As mentioned above, the modal approach, which has been as widely employed in analysing (non-axisymmetric) wave dynamics in discs (e.g., Papaloizou and Pringle, 1985; Narayan et al., 1987; Lin et al., 1990; Lubow and

Ogilvie, 1998; Ogilvie and Lubow, 1999; Li et al., 2003) as in hydrodynamics, does not permit the full study of finite-time phenomena originating from the non-normality/shear of disc flow. For this reason, when analysing dynamical processes in discs, we will use a different technique – the method of shearing waves (Kelvin modes), a special type of the non-modal approach, which is also often employed in disc analyses (e.g., Goldreich and Lynden-Bell, 1965b; Goldreich and Tremaine, 1978; Nakagawa and Sekiya, 1992; Johnson and Gammie, 2005a, see chapters 3 and 4 for details) and is ideally suited for characterising non-normality-induced perturbation dynamics in constant shear flows. The remainder of the thesis is organised as follows.

In the next chapter we study the linear and non-linear dynamics of vortices in razor-thin (2D) self-gravitating discs through analytical and numerical analysis. First, in the linear regime, we investigate the growth of the vortical mode under the combined effects of self-gravity and shear. We also quantify the efficiency of shear-induced SDW generation by the vortical mode at different strengths of self-gravity (i.e., at different Toomre Q). Then in the non-linear regime, using numerical simulations, we follow the process of vortex (PV) development in time again under the action of self-gravity and shear, starting with a random distribution of PV. We also discuss the implications of this for the process of planetesimal formation via trapping of dust particles by vortices.

In chapter 4, we consider non-self-gravitating 3D discs with a superadiabatic vertical stratification. In such discs, we investigate in detail another case of shear-induced linear mode coupling phenomenon – generation of SDWs by vertical convective motions. In other words, we study here new effects of compressibility in light of the recent findings on the dynamical (transport) properties of incompressible convection in protoplanetary discs. We also discuss the angular momentum transport by generated SDWs.

In chapter 5, we investigate axisymmetric modes in vertically stratified self-gravitating discs. Although axisymmetric modes do not display shear-induced couplings, their analysis allows us to understand how gravitational instability develops in 3D discs compared to that in 2D ones and what type of gas motion is induced by the most unstable mode. Based on these findings, we discuss the specific behaviour of 3D perturbations involving distortions of disc surfaces, as has been seen in some numerical simulations of self-gravitating discs.

In the final chapter 6, we give a summary of the present work and future directions.

Chapter 3

Dynamics of vortices in self-gravitating discs

3.1 Introduction

As discussed in chapter 2, the dynamical activities of two basic types/modes of perturbations – spiral density waves (SDWs) and vortices – are a subject of active research in protoplanetary discs, as these two modes can drive outward angular momentum transport necessary for the secular evolution of neutral discs, where the MRI cannot operate. However, each mode is mostly analysed for different settings. Because they are excited by gravitational instability modally (i.e., due to the exponential instability), SDWs are the central perturbation types and therefore the main agents responsible for angular momentum transport in self-gravitating discs. However, SDWs can also be active in non-self-gravitating discs if there is some mechanism driving them. On the other hand, vortices are commonly studied in non-self-gravitating discs. As outlined in sections 2.2.4 and 2.4, the latest developments have revealed that in non-self-gravitating discs vortices can be one of the possible sources of SDWs because of their shear-induced linear coupling with the latter (Bodo et al., 2005; Johnson and Gammie, 2005a; Heinemann and Papaloizou, 2009a). This coupling is asymmetric: vortices with non-zero potential vorticity (PV) are able to excite SDWs characterised by zero PV, but not vice versa because PV is a conserved quantity (in the linear regime, in the absence of baroclinicity). Thus, this asymmetric coupling lends significance to vortices as a participant in SDW and shock manifestations in discs, which, in turn, as we have seen above, are responsible for the bulk of outward angular momentum transport. We emphasise once more that this linear coupling between vortices and SDWs induced by shear/non-normality is a special manifestation of a more general linear mode coupling phenomenon inevitably taking place in flows with inhomogeneous velocity profile, or shear (Chagelishvili et al., 1997) and, particularly, in Keplerian discs as well, which are a special case of shear flow. (Another manifestation – linear coupling between convection and SDWs – is considered in chapter 4.) This noteworthy finding, in turn, calls for revisiting self-gravitating disc dynamics with a particular emphasis to be put on the possible role of vortical perturbations in the overall dynamical picture together with SDWs.

All the studies on vortex dynamics mentioned in section 2.2.4 miss out an important aspect of protoplanetary discs – their self-gravity. As discussed in section 2.2.3, effective cooling times in discs are typically too long to cause fragmentation under the action of their own self-gravity (Boley et al., 2006, 2007; Rafikov, 2007). As a consequence, balance is established between heating due to gravitational instability and cooling (self-regulation mechanism, Bertin and Lodato 2001). Discs are expected to stay in this self-regulated, quasi-steady, gravitoturbulent state for a long time. In this state, Toomre's parameter Q hovers on the margin of gravitational instability. A general dynamical picture, as described in section 2.2.3, is that spiral structure develops in a disc and transports angular momentum outwards through gravitational and hydrodynamical stresses, thereby allowing matter to accrete onto the central star (e.g., Laughlin and Bodenheimer, 1994; Boley et al., 2006; Lodato and Rice, 2004; Lodato, 2007). The angular momentum transport in this case is usually attributed to SDWs, which are thought of as the only perturbation type present in the disc. In other words, almost all studies on self-gravitating disc dynamics concentrate on the dynamical activity (with the possibility of amplification) of SDWs and leave another class of perturbations – vortices – out of consideration. As discussed above, the latter plays an important role in non-self-gravitating discs and it seems natural to look for them and analyse their dynamics in self-gravitating discs too. Indeed, as we demonstrate in this chapter, in the linear regime, the coupling between vortices and SDWs is even more efficient in the presence of self-gravity. Besides, vortices can exhibit comparable or even larger (transient) growth due to self-gravity than that of SDWs. So, the omission of the vortical mode and its growth may result in an incomplete dynamical picture of self-gravitating discs. In perspective, such a study of vortex dynamics in the presence of self-gravity will allow us to see if the same mechanism of planetesimal formation – dust particle trapping by vortices – can also be at work in self-gravitating discs, or in other words, if the process of planetesimal formation could be started at the early stages of disc lifetime, when self-gravity was still important (i.e., during the embedded and early T Tauri phases).

To study the (non-linear) dynamics of vortices in self-gravitating discs, one must examine the behaviour of the PV field – the basic quantity characterising vortices. To date, no systematic investigation of PV behaviour, similar to that done for non-self-gravitating discs, has been carried out for self-gravitating ones. However, we should mention two relevant works by Adams and Watkins (1995) and Wada et al. (2002).

Adams and Watkins (1995) investigated the dynamics of a single vortex in a self-gravitating Keplerian disc in the quasi-geostrophic limit using the local shearing sheet approximation. The quasi-geostrophic approximation implies that the characteristic time-scale of a problem is much larger than the orbital period. The vortex considered in their paper is in geostrophic balance and, therefore, remains steady for many rotation periods. The gas motion inside the vortex is subsonic as well. In this case, the effect of self-gravity is only to make the effective length scale (Rossby radius) of the vortex larger than that in the non-self-gravitating case. The quasi-geostrophic approximation does not permit consideration of the most important aspect of dynamics – effects of compressibility (SDWs and shocks), which are intertwined with vortices and have typical time-scales of the order of orbital (shear) time. In other words, these relatively fast motions associated with compressibility are filtered out in the quasi-geostrophic approximation.

Wada et al. (2002) investigated the properties of the gravitoturbulent state in the interstellar medium of galaxies. The vorticity (but not PV) field in this state was indeed calculated in their paper, which has a rather complicated structure. However, the authors did not discuss in detail the properties of the vorticity field and its relation to the density and pressure fields. They only pointed out that negative and positive vorticity regions are associated with dense filaments seen in the density field. In that paper the main emphasis is put on analysing the spectral properties of gravitoturbulence. Their study clearly demonstrated that vortical perturbations are as important as spiral density waves in the formation of spectra of the resulting gravitoturbulent state.

In this chapter, following the approach of Gammie (2001) and Johnson and Gammie (2005a,b), we study the specific properties of vortex dynamics in self-gravitating discs by means of analytical and numerical analysis. We adopt the local 2D model described in section 2.3 without invoking the quasi-geostrophic approximation, thereby allowing for compressibility effects. In this respect, our analysis is more general than that of Adams and Watkins

(1995). We start by outlining the main features of vortex evolution in the linear regime – its transient amplification and coupling with SDWs. As is usually done in the shearing sheet, perturbations are decomposed into spatial Fourier harmonics, or shearing waves, with a time-dependent radial wavenumber. We demonstrate that an initially imposed shearing wave of the vortical mode characterised by non-zero PV, evolving in the flow, undergoes transient amplification and generates in this process a corresponding shearing wave of SDW mode with zero PV. Previous related studies in the shearing-sheet with self-gravity by Goldreich and Lynden-Bell (1965b); Goldreich and Tremaine (1978); Toomre (1981) concentrated only on the dynamics of perturbations with zero PV (that is, on SDWs), leaving vortical perturbations out of consideration. After analysing the dynamical behaviour of each shearing wave of the vortical mode, we turn to a more general problem of the evolution of more realistic initial conditions – an ensemble of vortical mode shearing waves distributed randomly in the wavenumber plane. In the real physical plane, this translates into the random perturbations of velocity components and, hence of PV, imposed initially on the background Keplerian flow of the disc. We then trace the development of structures out of this chaotic field. One of the main goals here is to investigate the dynamical picture of PV evolution in the state of quasi-steady gravitoturbulence and how this picture differs from that occurring in non-self-gravitating discs (given the same chaotic type of initial conditions) studied previously by Johnson and Gammie (2005b); Shen et al. (2006). This, in turn, allows us to draw important conclusions about the effects of self-gravity on the formation and evolution of vortices. We will see that in the presence of self-gravity, vortex evolution is generally not as smooth and regular as it is in the non-self-gravitating case.

The plan of this chapter is as follows.[1] The mathematical formalism of the problem is described in section 3.2. The linear theory of vortical perturbations in the presence of self-gravity is presented in section 3.3. Non-linear evolution of vortices is presented in section 3.4 and a summary and discussions are given in section 3.5.

3.2 Physical Model and Equations

To study the dynamics of perturbation modes in thin self-gravitating gaseous discs, we employ the 2D local shearing sheet model. As noted in section 2.3, in this model, the unperturbed Keplerian differential rotation of the disc manifests itself as a parallel azimuthal flow with

---

[1] This chapter is based on my published two papers with the collaborators: Mamatsashvili and Chagelishvili (2007) and Mamatsashvili and Rice (2009).

a constant velocity shear in the radial direction, $\mathbf{u_0} = (0, -q\Omega x)$, in the local Cartesian reference frame (we recall that $q = 1.5$ for Keplerian shear considered here). The unperturbed background surface density Σ_0 and two-dimensional pressure P_0 corresponding to this shear flow are assumed to be spatially constant. So, we again give here the momentum equation (2.15) componentwise, continuity equation (2.16), and Poisson's equation (2.17) of this local model, which are used throughout this chapter:

$$\frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma \mathbf{u}) - q\Omega x \frac{\partial \Sigma}{\partial y} = 0, \tag{3.1}$$

$$\frac{\partial u_x}{\partial t} + (\mathbf{u} \cdot \nabla) u_x - q\Omega x \frac{\partial u_x}{\partial y} = -\frac{1}{\Sigma}\frac{\partial P}{\partial x} + 2\Omega u_y - \frac{\partial \psi}{\partial x}, \tag{3.2}$$

$$\frac{\partial u_y}{\partial t} + (\mathbf{u} \cdot \nabla)\, u_y - q\Omega x \frac{\partial u_y}{\partial y} = -\frac{1}{\Sigma}\frac{\partial P}{\partial y} + (q-2)\Omega u_x - \frac{\partial \psi}{\partial y}. \tag{3.3}$$

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\psi = 4\pi G(\Sigma - \Sigma_0)\delta(z). \tag{3.4}$$

Here again $\mathbf{u} = (u_x, u_y), P, \Sigma$ and ψ are, respectively, the perturbed velocity relative to the background parallel shear flow $\mathbf{u_0}$, the two-dimensional pressure, the surface density and the gravitational potential of the gas sheet. Since Eqs. (3.1-3.3) are written for perturbed velocities, only the gravitational potential due to the perturbed surface density $\Sigma - \Sigma_0$ is used (see Eq. 3.4). In these dynamical equations, the gradients of the gravitational potential are taken at $z = 0$, i.e., where the shearing sheet is located, because only this quantity depends on the vertical coordinate z. The equation of state is

$$P = (\gamma - 1)U,$$

where U and γ are the two-dimensional internal energy density and adiabatic index of gas, respectively. The two-dimensional adiabatic index γ is obtained by mapping the usual three-dimensional adiabatic index, assuming hydrostatic equilibrium in the vertical z-direction. Without loss of generality, following Gammie (2001), here we adopt the value $\gamma = 2$ (different values of this index do not qualitatively change results). The adiabatic sound speed is defined as $c_s^2 = \gamma P/\Sigma = \gamma(\gamma - 1)U/\Sigma$.

A central quantity of this study is the vertical component of potential vorticity, referred to as PV for short:

$$I \equiv \frac{\mathbf{e}_z \cdot \nabla \times \mathbf{u} + (2-q)\Omega}{\Sigma} = \frac{1}{\Sigma}\left(\frac{\partial u_y}{\partial x} - \frac{\partial u_x}{\partial y} + (2-q)\Omega\right).$$

In the unperturbed state, where there is only the background shear flow, $\mathbf{u_0}$, with the constant equilibrium surface density Σ_0, PV is equal to $I_0 = (2-q)\Omega/\Sigma_0$. The PV will play an important role in the subsequent analysis, as it generally characterises the formation of coherent structures (vortices) in a disc flow (e.g., Godon and Livio, 1999a,b; Johnson and Gammie, 2005b; Bodo et al., 2007). Using Eqs. (3.1-3.3), after some algebra, one can show that the evolution of PV is governed by the following equation

$$\left(\frac{\partial}{\partial t}+\mathbf{u}\cdot\nabla-q\Omega x\frac{\partial}{\partial y}\right)I=\frac{1}{\Sigma^3}\left(\frac{\partial\Sigma}{\partial x}\frac{\partial P}{\partial y}-\frac{\partial\Sigma}{\partial y}\frac{\partial P}{\partial x}\right).$$

This equation describes the advection of PV along the trajectories of fluid elements (Lagrangean derivative inside the brackets on the left hand side) and its change due to the non-linear baroclinic term on the right hand side. In the present case, the pressure and surface density are not related by any (e.g., isentropic, isothermal or polytropic) constraint, so this baroclinic term is, in general, non-zero and, therefore, PV is not conserved. However, in the linear approximation it vanishes and PV is conserved making it possible to classify modes into vortical and wave types (see below). Also note that the gravitational potential, as it should be, does not explicitly enter into this equation; self-gravity only influences PV evolution through surface density, pressure and velocity fields.

The evolution of internal energy density is governed by the equation

$$\frac{\partial U}{\partial t}+\nabla\cdot(U\mathbf{u})-q\Omega x\frac{\partial U}{\partial y}=-P\nabla\cdot\mathbf{u}-\frac{U}{\tau_c},\tag{3.5}$$

where the first term on the right hand side is the compressional heating term and the second term describes cooling of the disc according to a simple cooling law prescription with a constant characteristic time τ_c (e.g., Gammie, 2001; Rice et al., 2003). Here we choose the value of τ_c, such that the disc does not fragment and settles into a saturated quasi-steady state, where all quantities fluctuate around constant averaged values. Namely, we take $\tau_c = 20\Omega^{-1}$, which means a non-fragmenting disc according to criterion (2.6). We refer to this quasi-steady state as gravitoturbulence. In the present study, we concentrate on examining the specific properties of PV evolution in such a gravitoturbulent state. Strictly speaking, where fragmentation conditions are concerned, more realistic (radiative) cooling laws, with τ_c being a function of Σ, U, Ω rather than a constant, are necessary (e.g., Johnson and Gammie, 2003; Boley et al., 2006). We have seen in section 2.2.3 that most simulations of self-gravitating discs with such a realistic cooling, however, indicate that in this case cooling is usually not sufficient to cause fragmentation over most of the disc except possibly for the outer regions, therefore, a quasi-steady state is more likely. We do not include any artificial heating terms in the internal energy equation; heating is solely due to the compressional term and to shocks produced during the development of gravitational instability.[2]

---

[2] In our numerical code employed to solve Eqs. (3.1-3.5), shocks are captured, and therefore shock-heating

Before proceeding to the main analysis, we introduce normalised quantities which are more convenient to work with. As mentioned above, in the unperturbed state the background surface density Σ_0, pressure P_0 and internal energy U_0 are all spatially constant. In the unperturbed state the sound speed is $c_{s0}^2 = \gamma P_0/\Sigma_0 = \gamma(\gamma-1)U_0/\Sigma_0$. So, we switch to the following non-dimensional quantities:

$$t \to \Omega t, \quad (x, y, z) \to \left(\frac{x\Omega}{c_{s0}}, \frac{y\Omega}{c_{s0}}, \frac{z\Omega}{c_{s0}}\right), \quad (u_x, u_y, c_s) \to \left(\frac{u_x}{c_{s0}}, \frac{u_y}{c_{s0}}, \frac{c_s}{c_{s0}}\right),$$

$$\Sigma \to \frac{\Sigma}{\Sigma_0}, \quad (P, U) \to \left(\frac{P}{\Sigma_0 c_{s0}^2}, \frac{U}{\Sigma_0 c_{s0}^2}\right), \quad \psi \to \frac{\psi}{c_{s0}^2}, \quad I \to \frac{I\Sigma_0}{\Omega}, \quad \tau_c \to \Omega\tau_c.$$

Note that c_{s0}/Ω is actually the scale height of the disc. Therefore, distances are normalised by the disc scale height. The local Mach number is defined as $M = \sqrt{u_x^2 + u_y^2}/c_s$. Toomre's parameter, a measure of disc self-gravity, is $Q = c_s\Omega/\pi G\Sigma$, as defined in section 2.2.3 (we recall that $\kappa = \Omega$ for Keplerian rotation). These non-dimensional quantities will be used from now on throughout this chapter.

3.3 Linear dynamics of vortical mode

To understand the specific dynamics of vortices in the presence of self-gravity, we start with the linear analysis and then consider the full non-linear dynamics in the next section. Linearising Eqs. (3.1-3.4) about the unperturbed state and taking into account the above normalisation, we get:

$$\left(\frac{\partial}{\partial t} - qx\frac{\partial}{\partial y}\right)\sigma + \frac{\partial u_x}{\partial x} + \frac{\partial u_y}{\partial y} = 0, \tag{3.6}$$

$$\left(\frac{\partial}{\partial t} - qx\frac{\partial}{\partial y}\right)u_x = 2u_y - \frac{\partial}{\partial x}(\sigma + \psi), \tag{3.7}$$

$$\left(\frac{\partial}{\partial t} - qx\frac{\partial}{\partial y}\right)u_y = (q-2)u_x - \frac{\partial}{\partial y}(\sigma + \psi) \tag{3.8}$$

and the normalised Poisson's equation

$$\Delta\psi = \frac{4}{Q}\sigma\delta(z), \tag{3.9}$$

where $\sigma \equiv \Sigma - 1$ is the non-dimensional perturbed surface density. To keep the linear analysis simple, without losing the main features of the mode dynamics, in Eqs. (3.7) and (3.8), we have used a polytropic equation of state relating the non-dimensional perturbed pressure, $P - 1/\gamma$, and the perturbed surface density by $P - 1/\gamma = \sigma$, instead of a more general Eq. (3.5) for the internal energy with cooling.

---

is taken into account in energy equation (3.5), through artificial (von Neumann-Richtmyer) viscosity (see Appendix).

As usually done in the shearing sheet (e.g., Goldreich and Lynden-Bell, 1965b; Goldreich and Tremaine, 1978; Nakagawa and Sekiya, 1992; Johnson and Gammie, 2005a), we introduce shearing waves, or spatial Fourier harmonics (SFHs) of perturbations with time-dependent amplitudes and phases:

$$F(\mathbf{r},t) = F(t)\exp[\mathrm{i}K_x(t)x + \mathrm{i}K_y y], \quad K_x(t) = K_x(0) + qK_y t, \tag{3.10}$$

where $F \equiv (u_x, u_y, \sigma, \psi_{z=0})$ and K_x, K_y are the wavenumbers normalised by Ω/c_{s0}. The streamwise/azimuthal wavenumber K_y remains unchanged, while the streamcross/radial wavenumber $K_x(t)$ changes with time at a constant rate qK_y if $K_y \neq 0$ (i.e., for non-axisymmetric perturbations) due to the effect of the shearing background on wave crests. It can be said that in the linear regime, SFHs 'drift' along the K_x-axis in the wavenumber $\mathbf{K}$-plane. In other words, lines of constant phase of each SFH are sheared over by the mean flow in the real plane. So, as time passes, an initially leading SFH (with $K_x(0)/K_y < 0$) eventually becomes trailing (with $K_x(t)/K_y > 0$). This change of the SFH's orientation from leading to trailing is called 'swing' and occurs when $K_x(t) = 0$. The method of decomposition of perturbations into shearing waves, or Kelvin modes, is a special case of the non-modal approach and was originally devised by Lord Kelvin (Thomson, 1887) in order to study transiently growing solutions in inviscid incompressible parallel shear flows. It is widely used today in different applications involving flows with constant shear and greatly helps to grasp finite-time linear phenomena – the transient growth and coupling of perturbation modes – in shear flows, which tend to be overlooked in the standard modal analysis. Recently, it has been mathematically proven by Yoshida (2005) that these Kelvin modes in fact represent the simplest/basic 'elements' of dynamical processes at constant shear rate.

Substituting Eq. (3.10) into Eqs. (3.6-3.9) and making a change $\sigma \to \mathrm{i}\sigma$, $\phi \equiv \mathrm{i}\psi_{z=0}$, we get the system of ordinary differential equations governing the linear dynamics of SFHs of perturbations:

$$\frac{d\sigma}{dt} = K_x(t)u_x + K_y u_y, \tag{3.11}$$

$$\frac{du_x}{dt} - 2u_y = -K_x(t)\left(\sigma + \phi\right), \tag{3.12}$$

$$\frac{du_y}{dt} + (2-q)u_x = -K_y\left(\sigma + \phi\right), \tag{3.13}$$

$$\phi = -\frac{2}{QK(t)}\sigma, \tag{3.14}$$

where $K(t) = (K_x^2(t) + K_y^2)^{1/2}$. Equation (3.14) follows from Poisson's equation (3.9) and is straightforward to derive (see e.g., Nakagawa and Sekiya, 1992).

It can be easily shown that this system possesses an important time invariant

$$K_x(t)u_y - K_y u_x + (2-q)\sigma \equiv \mathcal{I}, \tag{3.15}$$

which is a linearised form of PV for an individual SFH. This time invariant $\mathcal{I}$, in turn, indicates the existence of the vortical/aperiodic mode in the perturbation spectrum. Clarification of the role of this mode in the disc flow dynamics represents one of the goals of our study.

In the calculations below, we use the quadratic form (spectral energy density) for a separate SFH as a measure of its intensity:

$$E(t) \equiv \frac{1}{2}\left(|u_x|^2 + |u_y|^2 + |\sigma|^2\right), \tag{3.16}$$

where the first two terms correspond to the perturbed kinetic energy and the third term to the perturbed potential energy of SFH. Strictly speaking, this is not an exact expression for perturbation energy, since it misses terms corresponding to gravitational energy. Nevertheless, we find this quadratic form convenient for a presented below comparative analysis of transient growth of perturbation modes at different values of Q.

The numerical study of SFHs' dynamics is based on Eqs. (3.11-3.15). However, for a better understanding of the dynamical processes, it is advisable to derive from these equations a single second order inhomogeneous differential equation for ϕ:

$$\frac{d^2\phi}{dt^2} + \omega^2(K_x(t), K_y)\phi = -\frac{4}{QK(t)}\left(1 - \frac{qK_y^2}{K^2(t)}\right)\mathcal{I}, \tag{3.17}$$

where

$$\omega^2(K_x(t), K_y) = 1 + K^2(t) - \frac{2}{Q}K(t) - \frac{4qK_y^2}{K^2(t)} + \frac{3q^2K_y^4}{K^4(t)} \tag{3.18}$$

is the shear-modified frequency of SDWs (see classification below); in the shearless limit ($q = 0$), Eq. (3.18) reduces to the well-known dispersion relation (2.5) for SDWs. All other perturbed quantities are easily expressed in terms of $\phi(t)$ and its time derivative. Notice that due to the time-varying radial wavenumber $K_x(t)$, the frequency ω is also time-dependent resulting, as we will demonstrate below, in the modes of perturbations becoming coupled in the linear theory. For further reference, in Fig. 3.1 we show the unstable ($\omega^2 < 0$) domains in the $\mathbf{K}$-plane for various Q. At $Q < 1$, the unstable domains exist even in the shearless limit, while at $Q \geq 1$, their occurrence is due to the combined action of shear and self-gravity.

3.3.1 Classification of perturbation modes

We can classify perturbation modes involved in Eq. (3.17) (or, equivalently, in Eqs. 3.11-3.14) from the mathematical and physical standpoints separately.

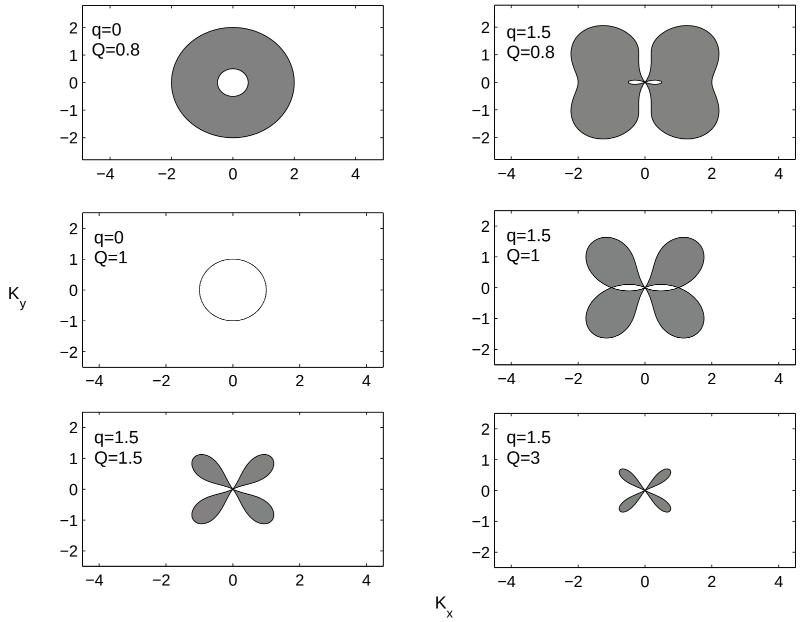

Figure 3.1: Unstable ($\omega^2 < 0$) domains (grey) in the **K**-plane for various values of Q. At $Q < 1$, unstable domains exist even in the shearless limit. At $Q \geq 1$, compared with the case of $q = 0$, shear leads to the emergence of the unstable domains. The lower two panels are given only for $q \neq 0$, since there are no unstable domains in the shearless limit at $Q \geq 1$.

Mathematically, a general solution of Eq. (3.17) can be represented as a sum of two parts: a *general* solution of the corresponding homogeneous equation – an oscillatory with time SDW mode – and a *particular* solution of this inhomogeneous equation. It should be emphasised that the particular solution is not uniquely determined: the sum of a particular solution of the inhomogeneous equation and any particular solution of the corresponding homogeneous equation (i.e. wave mode solution) is also a particular solution of the inhomogeneous equation, that is, a particular solution may comprise any amount of the wave mode.

Physically, Eq. (3.17) describes two different modes/types of perturbations:
(a) The oscillatory (with the time-dependent frequency $\omega(t)$) SDW mode ($\phi^{(\mathrm{w})}$), that is determined by a general solution of the corresponding homogeneous equation and therefore carries zero PV ($\mathcal{I} = 0$);
(b) The vortical mode ($\phi^{(\mathrm{v})}$), originating from the equation inhomogeneity (the right hand side of Eq. 3.17), is associated with the non-oscillatory part of a particular solution of the inhomogeneous equation. In the shearless limit, the vortical mode is independent of time and has zero velocity divergence (i.e., it is incompressible with $\nabla \cdot \mathbf{u} = 0$), but non-zero PV. However, in the presence of shear, it becomes slowly varying with time and can acquire divergence as well (see below). From the above argument, it follows that the correspondence between the

aperiodic vortical mode and the particular solution of the inhomogeneous equation is quite unambiguous: the vortical mode is associated only with that part of the particular solution that does not contain any oscillations. The amplitude of the vortical mode is proportional to $\mathcal{I}$ and goes to zero when $\mathcal{I} = 0$.

In the following, we will keep to the physical standpoint of separation of perturbation modes. Thus, any solution of Eqs. (3.11-3.14) can be expressed as a superposition of oscillatory SDW and non-oscillatory/aperiodic vortical modes:

$$u_x = u_x^{(\mathrm{w})} + u_x^{(\mathrm{v})}, \quad u_y = u_y^{(\mathrm{w})} + u_y^{(\mathrm{v})},$$

$$\sigma = \sigma^{(\mathrm{w})} + \sigma^{(\mathrm{v})}, \quad \phi = \phi^{(\mathrm{w})} + \phi^{(\mathrm{v})},$$

where $u_x^{(\mathrm{w})}$, $u_y^{(\mathrm{w})}$, $\sigma^{(\mathrm{w})}$ and $u_x^{(\mathrm{v})}$, $u_y^{(\mathrm{v})}$, $\sigma^{(\mathrm{v})}$ are found, respectively, from $\phi^{(\mathrm{w})}$ and $\phi^{(\mathrm{v})}$ by mean of Eqs. (3.11), (3.14) and (3.15) (for $u_x^{(\mathrm{w})}$, $u_y^{(\mathrm{w})}$, $\sigma^{(\mathrm{w})}$, we should, obviously, take $\mathcal{I} = 0$ in Eq. 3.15). In fact, the (modified) initial value problem is solved by Eqs. (3.11-3.14) (or, equivalently, by Eq. 3.17). The character of the subsequent dynamics depends on which perturbation mode is inserted initially into Eqs. (3.11-3.14): pure SDW mode, without a mix of aperiodic vortices, or pure aperiodic vortical mode, without a mix of SDWs.

The classification of perturbation modes that is widespread divides them into solenoidal type, with non-zero curl and zero divergence, and divergent (potential) type, with non-zero divergence and zero curl, in the presence of background shear (see e.g., Wada et al., 2002; Johnson and Gammie, 2005b; Shen et al., 2006), by analogy with a similar classification in the case of no shear. However, in the considered Keplerian (i.e., strongly sheared) flow, the situation is fundamentally different: the solenoidal mode (according to this classification) may acquire divergent nature and initially predominantly divergent mode may acquire curl in the course of evolution.

To illustrate this, in Fig. 3.2 we present the time-development of the parameter:

$$\eta = \left| \frac{K_x(t)u_y - K_y u_x}{K_x(t)u_x + K_y u_y} \right|,$$

which represents the ratio of the z-component of curl to divergence, and its inverse value, $1/\eta$, for initially imposed SDW and vortical mode SFHs, respectively. In the case where initially a predominantly divergent, leading SDW mode SFH (with one sign of frequency) is inserted into Eqs. (3.11-3.14), as seen in Fig. 3.2a, it acquires curl at about the time of entering the unstable/non-adiabatic domain in Fig. 3.1 and also in Fig. 3.4 below. Figure 3.2b shows that an initially inserted, leading pure vortical mode SFH (with the form given by Eq. 3.19 below), being mostly solenoidal at the beginning, acquires divergent nature at

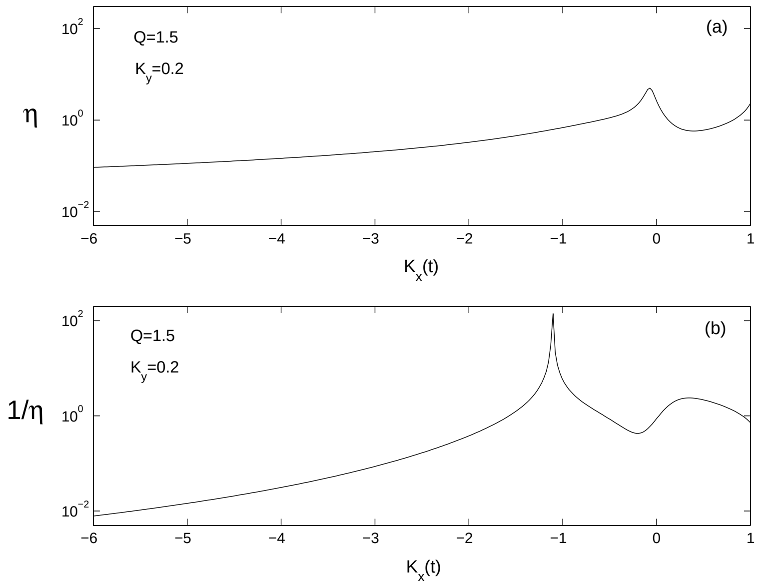

Figure 3.2: Panel (a) shows the evolution of η in the case where a leading pure SDW mode SFH with only one sign (positive) of frequency is inserted initially into Eqs. (3.11-3.14). This wave mode SFH acquires curl at about the time when it starts to enter the unstable (non-adiabatic) domain in Figs. 3.1 and 3.4. Panel (b) shows the evolution of $1/\eta$ for an initially inserted leading pure vortical mode SFH. This vortical mode SFH acquires divergent nature at about the same time. In both panels, $Q = 1.5$ and $K_y = 0.2$.

about the same time. This coupling between the solenoidal and divergent components in the presence of shear is directly related to the coupling of SDWs and vortical modes described below.

Thus, in the Keplerian flow (or in a shear flow in general), truly divergent and solenoidal perturbations in the sense described above, in fact represent some mix of vortical and wave modes and classification of perturbations as solenoidal or divergent may be misleading. So, we prefer the above-performed classification of perturbations into wave and vortical modes according to the value of PV and investigate dynamical processes in terms of the dynamics of these two modes.

3.3.2 Transient growth of the vortical mode and its coupling with the SDW mode – numerical analysis

We now numerically integrate Eqs. (3.11-3.14), imposing initially a leading (with $K_x(0)/K_y < 0$) pure vortical mode SFH without any mix of SDW mode SFH. Such a selection of the vortical mode is possible only far from the unstable/non-adiabatic domains, where $|K_x(t)/K_y| \gg 1$ (i.e., when SFHs are tightly wrapped) and the adiabatic condition with respect to time,

$|d\omega(t)/dt| \ll \omega^2(t)$, is satisfied (Bodo et al., 2005; Tevzadze, 2006; Heinemann and Papaloizou, 2009a). In this case, the tightly leading (with $K_x(t)/K_y \ll -1$) vortical mode is given by

$$\phi^{(\mathrm{v})}(t) = -\frac{4}{QK(t)\omega^2(t)}\left(1 - \frac{qK_y^2}{K^2(t)}\right)\mathcal{I}, \tag{3.19}$$

which is obviously an non-oscillatory, slowly-varying particular solution of Eq. (3.17), valid provided the adiabatic condition holds (i.e., the second derivative of this solution is negligible in Eq. 3.17). In fact, this approximate solution is the first term of a regular perturbation expansion of a slowly-varying with time (because of shear) aperiodic particular solution of Eq. (3.17) in powers of the small quantity ω'/ω^2, with higher-order correction terms to be found by means of an exponential-asymptotic technique described by Vanneste and Yavneh (2004); Bodo et al. (2005); Tevzadze (2006); Heinemann and Papaloizou (2009a). For our subsequent numerical analysis, however, only the first term of such an expansion (i.e., expression 3.19) will do, because we start integration sufficiently far from the non-adiabatic domain. We should note that such an asymptotic expansion method can only yield a non-oscillatory solution – the vortical mode – with higher and higher accuracy in ω'/ω^2, but it is unable to capture SDW generation by the latter (see e.g., Vanneste and Yavneh, 2004; Bakas and Farrell, 2009), so numerical integration is necessary in this case. We use a standard Runge-Kutta scheme (MATLAB ode45 RK implementation) to solve Eqs. (3.11-3.14).

In Fig. 3.3, we present the subsequent time-development of u_x, u_y, σ and $E/E(0)$ corresponding to the initially imposed vortical solution (3.19) at different values of K_y. Figure 3.4 sketches the evolution ('drift') of the associated SFH with $K_y \lesssim 2Q^{-1}$. At the beginning, the vortical mode SFH is located (schematically) at point 1, far from the unstable domains, and then drifts along the K_x-axis in the positive direction ($1 \to 2 \to 3 \to 4 \to 5 \to 6$ in Fig. 3.4). The drift velocity (qK_y) depends linearly on K_y (henceforth, we take the azimuthal wavenumber K_y to be positive without loss of generality). Initially, being in the adiabatic region, the vortical mode SFH gains energy from the mean flow and amplifies algebraically, but retains its aperiodic nature. Then, the dynamics becomes non-adiabatic (i.e., $|d\omega(t)/dt| \sim \omega^2(t)$ and the second derivative in Eq. 3.17 becomes important) when the SFH reaches the first unstable domain where $\omega^2(t) < 0$ (point 2). From this point, a temporal exponential growth and simultaneous excitation of the corresponding SFH of SDW mode take place; at this stage of the evolution, the linear coupling of the vortical and wave mode SFHs is at work.[3] Then,

---

[3] This shear-induced linear phenomenon of conversion of vortices into waves was first revealed and thoroughly described for the simplest shear flow in Chagelishvili et al. (1997) and then for non-self-gravitating Keplerian discs in Bodo et al. (2005) and Heinemann and Papaloizou (2009a).

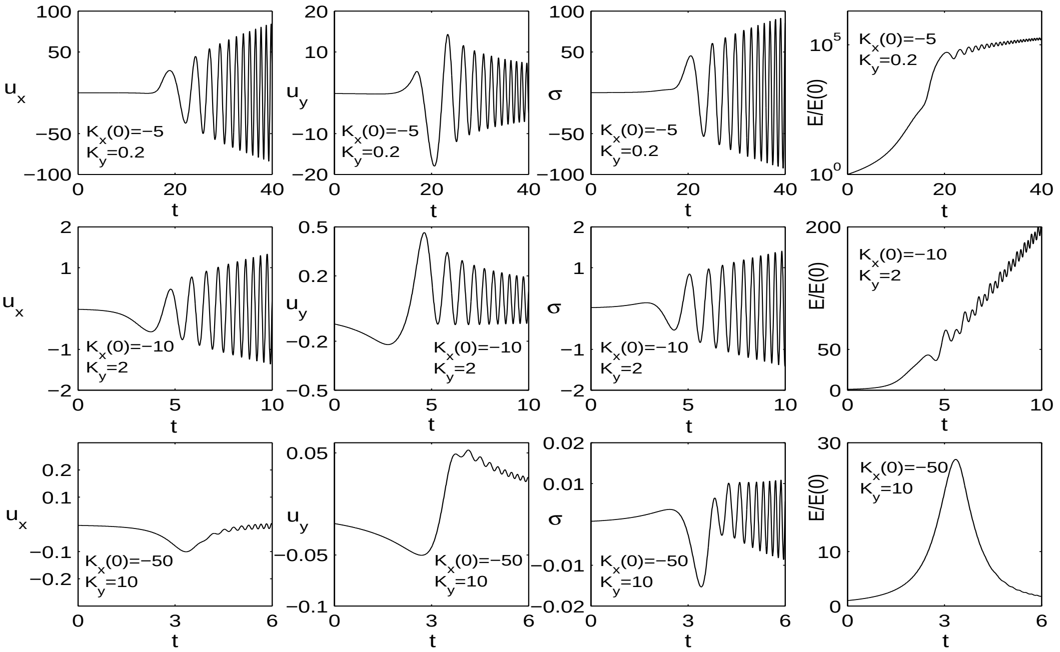

Figure 3.3: Evolution of perturbed quantities pertaining to an initially imposed leading pure vortical mode SFH with various $K_x(0)$ and K_y at fixed $Q = 1.5$. Before reaching the first unstable domain, the vortical mode SFH gains energy from the mean flow and amplifies, but retains its aperiodic nature. In the unstable domains, oscillations (that is, SDWs) begin to emerge. The energy increases monotonically and then, having gone through the unstable domains, asymptotically linearly corresponding to the generated SDW mode SFHs (for $K_yQ < 2$, the energy curves are in logarithmic scale because of large growth factors, so the flat part is actually linear growth). The contribution of the vortical mode energy to the total perturbation energy is negligible at this asymptotic stage; it simply dies down giving way to trailing SDWs. The wave generation is very efficient for $K_yQ < 2$, moderate for $K_yQ \sim 2$ and vanishing for $K_yQ \gg 2$.

the vortical and the generated SDW mode SFHs reach the intermediate stable region (point 3) where $\omega^2(t) > 0$, cross it and move again into the next domain where $\omega^2(t) < 0$ (point 4). Further exponential growth of both the vortical and SDW mode SFHs and, in addition, excitation of another SDW mode SFH by the vortical mode one occur while crossing this second unstable domain (until point 5). On moving away from this domain (point 6), the linear dynamics of the vortical and SDW mode SFHs become decoupled and adiabatic: the energy of the vortical mode SFH dies down and the energy of the wave mode SFHs increases. No further SDW excitation is observed beyond point 5.

Here we have described the SDW generation for $K_y \lesssim 2Q^{-1}$, though it similarly occurs for $K_y \gtrsim 2Q^{-1}$ (see Fig. 3.3) except that the transient amplification of an initially imposed

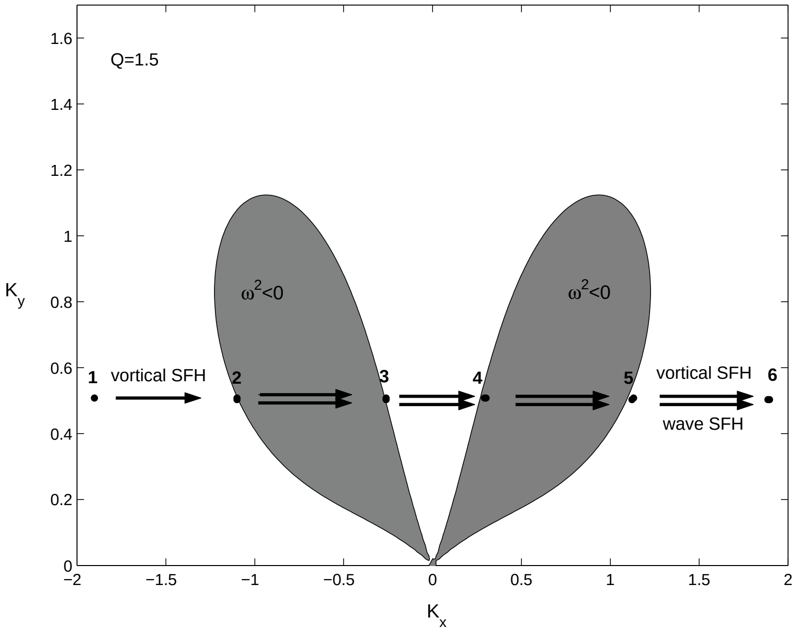

Figure 3.4: Schematic diagram of the SFH evolution in the **K**-plane. A leading vortical mode SFH, located initially at point 1, drifts with time along the K_x-axis in the direction denoted by the arrows $1 \to 2 \to 3 \to 4 \to 5 \to 6$. After crossing point 2, a SDW mode SFH appears. The first stage of the wave excitation takes place from point 2 to point 3; the second, from point 4 to point 5.

vortical mode SFH is algebraic, not exponential, because the unstable domains do not extend to such larger K_y. As a consequence, the intensity of amplification and the amplitudes of generated SDW mode SFHs are several orders of magnitude lower than those for $K_y \lesssim 2Q^{-1}$.

Thus, the linear dynamics of a vortical mode SFH is followed by the generation of the corresponding SDW mode SFHs. These generated SFHs as well as the former vortical mode SFH eventually acquire a (tightly) trailing orientation, since $K_x(t)/K_y > 0$ after leaving the non-adiabatic region (which stretches roughly from point 2 to point 5 in Fig. 3.4). In the non-adiabatic region, the characteristic time-scales of the vortical and SDW mode SFHs are comparable and therefore these modes cannot be quite separated/disentangled. But with moving away from the non-adiabatic region, modes get cleanly separated: the time-scale of the SDW mode SFHs becomes much shorter than that of the vortical mode SFH, as the frequency of waves increases with time (see Eq. 3.18). We can formally divide the energy evolution into two stages: the first stage represents the transient amplification (because of both the non-normality and the unstable domains) of the originally imposed pure vortical mode SFH and excitation (and also subsequent exponential amplification) of the corresponding SFHs of the SDW mode. The second stage represents the algebraic growth of the generated SDW

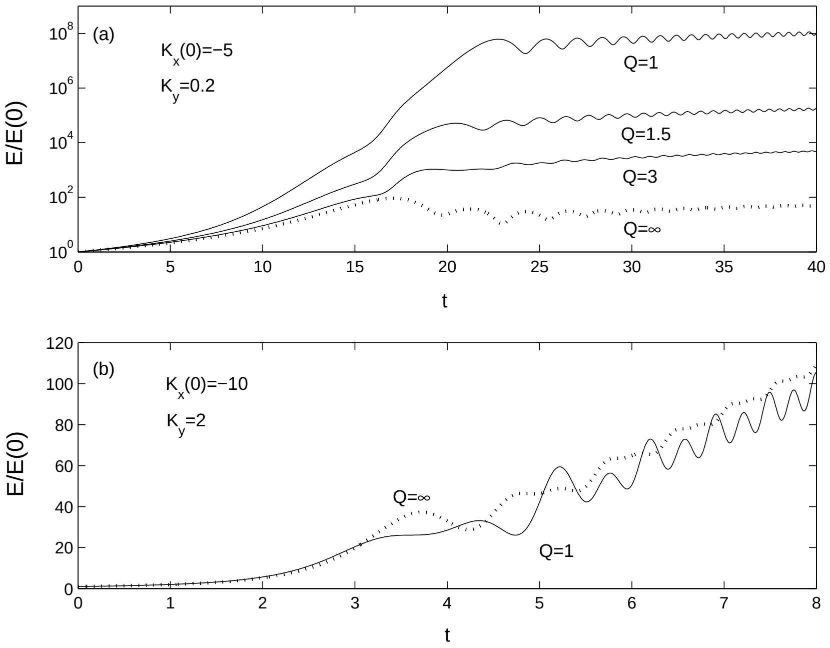

Figure 3.5: Influence of self-gravity on the transient (swing) amplification of a vortical mode SFH (dotted lines correspond to the non-self-gravitating case). This influence is, as expected, largest for $K_yQ \lesssim 2$ (panel a; all four curves are characterised by the common $K_x(0) = -5$, $K_y = 0.2$ and identical initial values of perturbed quantities) and weak for $K_yQ \sim 2$ (panel b, $K_x(0) = -10$, $K_y = 2$).

mode SFHs. The latter exhibit linear amplification at asymptotically large times. In the absence of the wave excitation (i.e., for $K_y \gg 2Q^{-1}$), this second stage describes decreasing energy of the vortical mode SFH. Thus, newly excited trailing SDW mode SFHs in the linear regime extract energy from the mean Keplerian flow in contrast to a trailing vortical mode SFH, which after leaving the non-adiabatic region, gradually returns all its energy to the mean flow. It can be said that vortical mode perturbations act as a mediator between the mean flow and waves; the energy needed for the wave excitation is extracted from the shear and self-gravity with the help of the vortical mode.

The following two subsections are devoted to the behaviour of the vortical mode under self-gravity for various values of the Q parameter.

Effect of self-gravity on the transient (swing) growth of the vortical mode

In Fig. 3.5, we compare the evolution of $E/E(0)$ for the initially imposed vortical mode SFHs in the presence and absence of self-gravity for different values of Q and K_y. It is clear that the growth of the vortical mode SFH continues longer and may be several orders of magnitude stronger than that in the non-self-gravitating case ($Q \to \infty$, dotted lines in Figs. 3.5a and

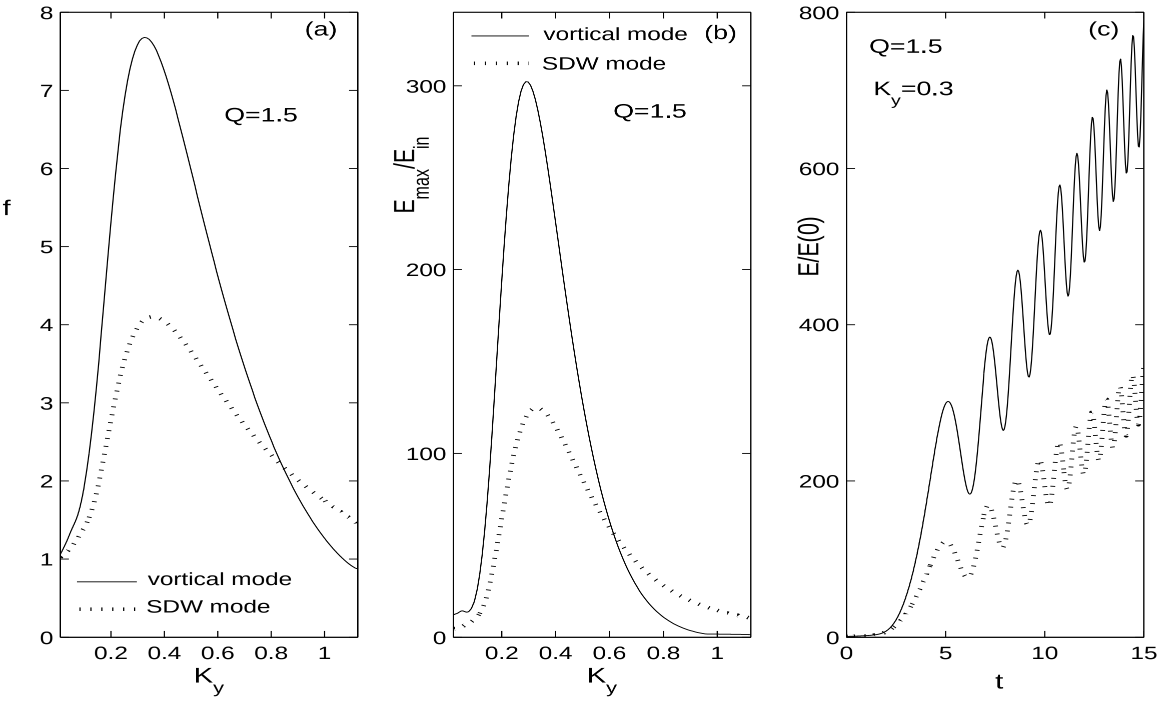

Figure 3.6: Growth factors of density (a) and energy (b) versus K_y. Also shown is the evolution of $E/E(0)$ for the two cases (c). Solid curves in these panels correspond to the initially imposed vortical mode SFH; dotted curves correspond to the initially imposed SDW mode SFH. In all three panels, $Q = 1.5$. In panel (c), $K_y = 0.3$. $t = 0$ corresponds to point 2 in Fig. 3.4.

3.5b). In the latter case, the growth of the vortical mode SFH occurs only at the leading stage ($K_x(t)/K_y < 0$); on becoming trailing ($K_x(t)/K_y > 0$), the SFH give back energy to the mean flow and decay (Chagelishvili et al., 2003). In the case of self-gravity, however, the amplification of SFHs also continues into the trailing stage because of the existence of the unstable domain at $K_x(t)/K_y > 0$ (see Figs. 3.1 and 3.4). From Fig. 3.5, it is seen that at $Q = 1$ and $K_y = 0.2$, the vortical mode SFH grows about 10^6 times stronger than in the non-self-gravitating case; at $Q = 1.5$ and $K_y = 0.2$, it is about 10^4 times stronger; at $Q = 3$ and $K_y = 0.2$, it is about 10^2 times stronger; at $Q = 1$ and $K_y = 2$ (i.e., at $K_yQ \sim 2$), the growth is almost the same as in the non-self-gravitating case. In any case, one can conclude that self-gravity provides a substantial enhancement of the transient growth of vortical mode perturbations, which occurs as SFHs swing from leading to trailing, thereby making them active participants in dynamical processes in self-gravitating discs together with SDWs.

In order to better understand the role of the vortical mode, it is interesting to carry out a comparative analysis of the transient (swing) amplification of the vortical and SDW modes. First, we define a density growth factor f for an SFH initially located at point 1 in Fig. 3.4, as a ratio of the absolute values of $\sigma(t)$ after (at point 5) and before (at point 2) crossing the unstable domains in Fig. 3.4, $f \equiv |\sigma(t'')|/|\sigma(t')|$, where t'' and t' are the time moments corresponding to points 5 and 2, respectively. A similar growth factor was used by Kim and Ostriker (2001) to quantify swing amplification, but they took its logarithm. In Fig. 3.6a, we present this parameter computed separately for the initially imposed vortical and SDW mode SFHs as a function of the normalised azimuthal wavenumber K_y at $Q = 1.5$. The initially imposed SDW mode SFH has a certain (positive) sign of frequency. As seen in this figure, in the dynamically important regions (i.e., for such values of K_y, at which both modes experience maximal transient growth), the growth factor for the vortical mode SFH is almost two times larger than that for the SDW mode one. An analogous comparison is made in Fig. 3.6b. Here we display the ratio of the maximum value achieved by the energy $E(t)$ during transient amplification in the unstable domains to its initial value on entering these domains (i.e., at point 2), computed separately for the initially imposed vortical and SDW mode SFHs as a function of K_y, similar to what is done in Fig. 3.6a. But now, as distinct from the first case, for the wave mode SFH we choose initial conditions at point 2 in such a way as to obtain the largest possible amplification of the wave energy during the transient growth (swing) phase for its fixed initial value (i.e., we take the transiently most unstable wave mode SFH). The situation is similar to the above situation: the energy amplification factor for the vortical mode SFH is more than two times greater than the largest possible energy amplification factor for the SDW mode SFH. In Fig. 3.6c, we present the parallel evolution of the energies of the initially (at point 2) imposed vortical and maximally amplified wave mode SFHs for $K_y = 0.3$, at which the energy growth factors of both modes during swing phase are the largest (see Fig. 3.6b). Both SFHs start with the same energy. Figure 3.6c shows that the energy corresponding to the initially imposed vortical mode SFH remains about two times larger than that corresponding to the SDW mode SFH at all times.

From Fig. 3.6, it is evident that the vortical mode prevails over the SDW mode in two ways: in the transient amplification and in wave generation. (By wave generation for the initially imposed SDW mode SFH, we mean wave amplification as a result of the over-reflection mechanism; for details, see Nakagawa and Sekiya 1992). The latter follows from the asymptotic stage at large times, when both energy curves become linear (see Figs. 3.3 and 3.6c) with inclinations proportional to the square of the amplitudes of generated SDW mode SFHs after crossing the unstable domains. We see that the energy of SDW mode SFHs excited by the initial vortical mode SFH at this asymptotic stage is about two times larger than that generated by the initial SDW mode SFH. This again confirms that in self-gravitating discs, vortical perturbations play as important a role as SDWs.

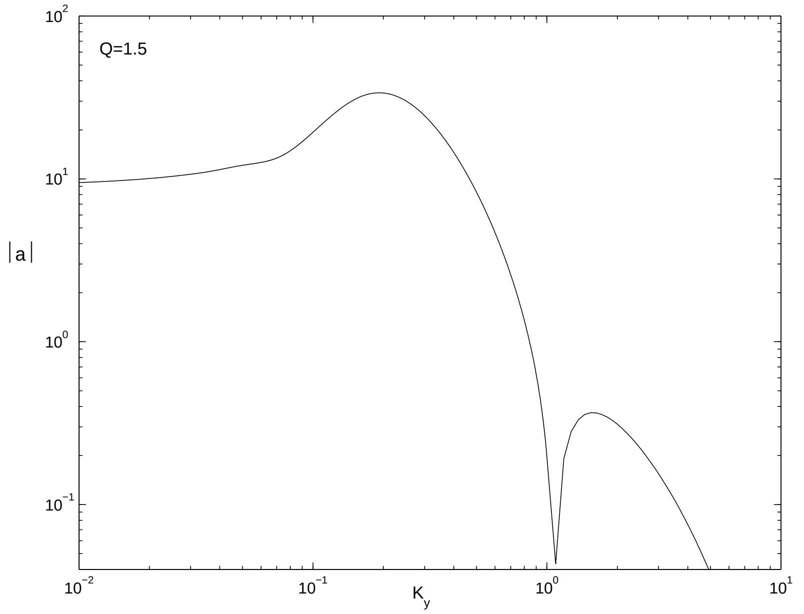

Figure 3.7: The amplitude $|a|$ of trailing SDW mode SFHs generated by a leading vortical mode SFH versus K_y at $Q = 1.5$.

Mechanisms of SDW generation

There are two mechanisms of SDW generation (amplification) in the disc flow considered here. The first is a direct and well-known over-reflection mechanism (Nakagawa and Sekiya, 1992): by inserting a leading SDW mode SFH into Eqs. (3.11-3.14), we get the energy behaviour represented by the dotted curve in Fig. 3.6c. The curve describes the energy growth in the unstable domains that is followed by a linear growth of the total energy of the resulting over-reflected and (over)-transmitted trailing SDW mode SFHs at large times.

Another way of generating trailing SDW mode SFH is by means of the vortical mode SFH, as described above. A leading vortical mode SFH can effectively excite trailing SDW mode SFHs due to the shear-induced mode coupling phenomenon. Figure 3.6 shows that this second way of SDW generation is about two times more effective than the former.

We now calculate the amplitudes of SDWs generated as a result of the mode coupling. (More precisely, we calculate the amplitudes for the gravitational potential perturbations; the amplitudes for other quantities can be found easily afterwards). We initially insert a tightly leading (i.e., in the adiabatic regime) vortical mode SFH, which is given by expression (3.19), into Eq. (3.17). Then after passing through the non-adiabatic region, in the tightly trailing phase at $t \to \infty$, $K_x(t)/K_y \gg 1$, this solution goes over to the superposition of the former

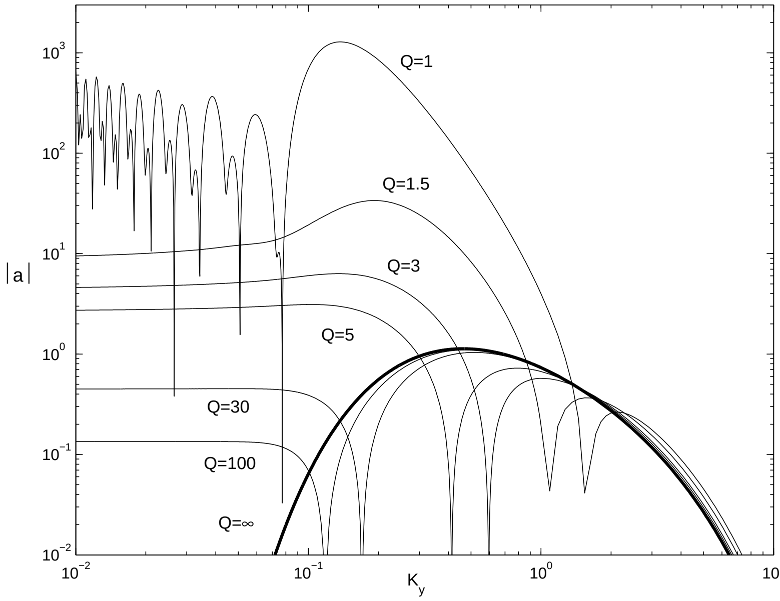

Figure 3.8: Same as in Fig. 3.7, but for different Q, including the non-self-gravitating case ($Q \to \infty$).

vortical mode SFH and excited by it SDW mode SFHs both with trailing orientation:

$$\phi(t) = \phi^{(\mathrm{v})}(t) + \phi^{(\mathrm{w})}(t) = -\frac{4}{QK(t)\omega^2(t)}\left(1 - \frac{qK_y^2}{K^2(t)}\right)\mathcal{I} + \\ + \frac{a}{Q\sqrt{\omega(t)}}\exp\left(-i\int^t \omega(t')dt'\right) + \frac{a^*}{Q\sqrt{\omega(t)}}\exp\left(i\int^t \omega(t')dt'\right),$$

where a and a^* are the amplitudes of generated SDW mode SFHs. The latter are represented by usual WKBJ solutions of the homogeneous part of Eq. (3.17) and come in complex conjugate pairs with different signs of frequencies and, hence, propagating in the opposite directions. Figure 3.7 plots the numerically obtained $|a|$ as a function of K_y at $Q = 1.5$. Figure 3.8 shows the same for different values of Q. In both figures $\mathcal{I}$ is set to unity.

Let us analyse the curves in Figs. 3.7 and 3.8. We clearly see that the amplitude $|a|$, and therefore the wave generation efficiency, decrease with increasing K_y. The maximum value of $|a|$ is achieved for smaller $K_y \sim O(0.1)$ (i.e., at horizontal length-scales of perturbations comparable to or larger than the disc scale height), as at such K_y a SFH drifts slowly in the **K**-plane. Consequently, it slowly crosses the unstable domains and has more time for transient growth. The dips in these figures are because of the crossing of two wave excitation domains by the SFH (see Fig. 3.4) and thus a result of the existence of two, more or less independent, stages of the wave excitation; the resulting (after leaving both unstable domains) wave mode SFH is a superposition of SFHs generated at these stages. At $Q = 1.5$, this interference is destructive near $K_y = 1$ and results in the dip there (Fig. 3.7). As can be seen from Fig. 3.8, the dip occurs at different values of K_y for different Q. At small Q, the number of dips increases (see the curve for $Q = 1$ in Fig. 3.8), as destructive interference happens at several values of K_y. In the non-self-gravitating case ($Q \to \infty$), there is no dip, as in this case there is just one stage of the wave excitation and, therefore, the interference phenomenon is absent. In this connection we would like to mention that, using an exponential-asymptotic technique, Heinemann and Papaloizou (2009a) also found that in non-self-gravitating discs, the amplitude of SDWs generated by the vortical mode behaves with K_y similarly to that represented by our curve with $Q \to \infty$ in Fig. 3.8. Figure 3.8 also demonstrates that the intensity of wave generation increases with decreasing Q, that is, self-gravity greatly enhances non-normality/shear-induced coupling of the SDW and vortical modes.

3.4 Non-linear dynamics of vortices

Having established that in self-gravitating discs vortical perturbations can be dynamically as important as SDWs, in this section we consider a more general case – the non-linear dynamics of vortices, their formation and evolution, in the presence of self-gravity. We start with a random distribution of PV in the real (x, y)-plane and trace the subsequent evolution via numerical simulations. We no longer restrict ourselves to a polytropic equation of state and return to a more general energy equation (3.5).

Our computational domain in the (x, y) plane is a rectangle $-L_x/2 \leq x \leq L_x/2, -L_y/2 \leq y \leq L_y/2$ of size $L_x \times L_y$ divided into $N_x \times N_y$ grid cells. The numerical resolution is therefore $\triangle x \times \triangle y = L_x/N_x \times L_y/N_y$. We will assume that $L_x = L_y \equiv L$ and $N_x = N_y \equiv N$. For the fiducial model presented below we take $N = 1024$, though we also ran a lower resolution ($N = 512$) model that converges to the fiducial model. As a main numerical tool to solve the governing equations (3.1-3.5) within this domain, we use a version of the ZEUS code (Stone and Norman, 1992), which is more suited to the shearing sheet (Gammie, 2001; Johnson and Gammie, 2003, 2005b). ZEUS evolves these equations on a staggered mesh in a time-explicit, operator-split, finite-difference fashion. As mentioned above, the code uses artificial viscosity to capture shocks. A detailed description of this version of the ZEUS code as well as of the method for solving Poisson's equation (3.4) for razor-thin discs are given in Appendix.

As with most works employing shearing sheet approximation, we adopt periodic boundary conditions in the y-direction and shearing periodic in the x-direction, that is, the x-boundaries are initially periodic but as time goes by they shear with respect to each other becoming again periodic when $t_m = mL_y/(qL_x) = m/q$, with $m = 1, 2, ...$ (Hawley et al., 1995). So, for each quantity we can write:

$$f(x,y,t) = f(x,y+L,t) \qquad (y \ \ boundary)$$

$$f(x,y,t) = f(x+L,y-qLt,t) \qquad (x \ \ boundary),$$

where $f \equiv (u_x, u_y, P, \Sigma, U)$ (we remind that time is normalised by Ω^{-1}). The shearing in the x-boundary condition, or shift in the y-direction by an amount $-qLt$ in the 'ghost zones' as well as advection of the basic quantities by the mean Keplerian shear flow are done here by means of the FARGO scheme (see Masset, 2000, and also Appendix for details), as distinct from the original ZEUS algorithm.

For further use, we define the autocorrelation function for PV as

$$R_I(x,y) = \frac{1}{L_xL_y}\int \delta I(x',y')\delta I(x+x',y+y')dx'dy',$$

where the integration is over the entire rectangular simulation domain and $\delta I = I - I_0$. The autocorrelation function characterises emerging coherent structures in a flow; its length scale can be identified with the characteristic scale of such structures. An analogous function was used by Gammie (2001) to analyse density structures in order to establish locality of angular momentum transport and by Johnson and Gammie (2005b) to characterise coherent vortices in the non-self-gravitating shearing sheet.

3.4.1 Total energy equation and α parameter

In this section we derive a relationship between the total energy and hydrodynamic and gravitational stresses. This relationship is important as it clearly shows the role of shear in the energy exchange between the background Keplerian flow and perturbations. The total energy density of perturbed (from Keplerian shear flow) gas motion is defined as

$$E = \frac{1}{2}\Sigma\mathbf{u}^2 + U + \frac{1}{2}(\Sigma - 1)\psi,$$

where the first term is the kinetic energy density, the second term is the internal energy density and the third term is the gravitational energy density. This form of energy density is a generalisation of the form (3.16) defined for linear perturbations (for each SFH). Using basic Eqs. (3.1-3.5) (in the non-dimensional form) and the shearing sheet boundary conditions, one can arrive at the evolution equation for the average total energy

$$\frac{d\langle E\rangle}{dt} = q\left\langle \Sigma u_x u_y + \frac{Q}{4}\int_{-\infty}^{+\infty}\frac{\partial\psi}{\partial x}\frac{\partial\psi}{\partial y}dz\right\rangle - \frac{\langle U\rangle}{\tau_c},$$

where the angle brackets denote averages defined as $\langle f\rangle = \int f dxdy/L_xL_y$ with the integration being done over the entire simulation domain. The first and second terms inside the angle brackets on the right hand side are the Reynolds and gravitational stresses, respectively. As is clear from this equation, these two terms are responsible for the energy exchange between the background Keplerian flow and perturbations, which is in fact possible due to the differential nature (shear) of Keplerian rotation. In the absence of background shear ($q = 0$), the total energy of perturbations changes only due to cooling. Incidentally, the sum in the angle brackets is also proportional to the flux of angular momentum density (Balbus and Hawley, 1998; Gammie, 2001; Balbus, 2003). Since the cooling time chosen ($\tau_c = 20$) is not short enough, the system, as mentioned above, settles into a quasi-steady, gravitoturbulent state after a couple of orbital periods, in which $d\langle E\rangle/dt = 0$ (similarly for every averaged quantity). Therefore, at these times we can write

$$q\left\langle \Sigma u_x u_y + \frac{Q}{4}\int_{-\infty}^{+\infty}\frac{\partial\psi}{\partial x}\frac{\partial\psi}{\partial y}dz\right\rangle = \frac{\langle U\rangle}{\tau_c},$$

or in terms of the well-known α viscosity parameter introduced in chapter 2

$$\alpha \equiv \frac{1}{q\langle\Sigma c_s^2\rangle}\left\langle \Sigma u_x u_y + \frac{Q}{4}\int_{-\infty}^{+\infty}\frac{\partial\psi}{\partial x}\frac{\partial\psi}{\partial y}dz\right\rangle$$

giving

$$\alpha = \frac{1}{q^2\gamma(\gamma-1)\tau_c}. \tag{3.20}$$

This relation is very important, as it shows that *if* a quasi-steady state can be achieved, then the angular momentum transport rate is determined solely by the cooling time for a given rotation rate and, therefore, should be constant with time. (The same formula for α was derived by Gammie 2001 in a different way.) It is a local analog of the well-know result that the surface brightness of the disc does not depend on the details of the angular momentum transport mechanism in a steady state (e.g., Pringle, 1981; Balbus, 2003). Again, the basic premise here is the possibility of the existence of a quasi-steady gravitoturbulent state, which is able to feed itself, or be self-sustaining, at the expense of background shear flow energy extracted by Reynolds and gravitational stresses. In the presence of disc self-gravity and cooling this state is easily reached because of the self-regulation mechanism (Bertin and Lodato, 2001). However, in 2D models of thin non-self-gravitating neutral discs, steady outward angular momentum transport is hard to achieve. In this case, as described in section 2.2.4, typically the disc is unstable to vortex formation and develops well-organised vortices, which although able to transport angular momentum outwards mostly via excitation of compressible motions (SDWs and shocks), are still slowly decaying on the time-scale of several hundred orbital periods. Thus, the role of self-gravity and cooling is crucial for the maintenance of gravitoturbulence.

3.4.2 Initial conditions

Following Godon and Livio (1999b); Johnson and Gammie (2005b); Shen et al. (2006), we take the initial conditions consisting of random u_x and u_y perturbations superimposed on the background Keplerian flow. The surface density and internal energy are not perturbed initially and thus are spatially uniform with values 1 and $1/\gamma(\gamma-1)$, respectively. To generate these initial conditions, for each velocity component at each point in the $\mathbf{K}$-plane we create a Gaussian random field with standard deviation, or amplitude of power spectrum, given by the Kolmogorov power law $|u_{x,y}(K_x, K_y)|^2 \propto K^{-8/3}$, in the range $K_{min} \leq K \leq K_{max}$. The limits of this range are $K_{min} = 2\pi/L$ and $K_{max} = (N/n_g)K_{min}$, where n_g is the number of grid cells contained within the smallest wavelength $2\pi/K_{max}$ for which we choose $n_g = 16$ when $N = 1024$. Outside this wavenumber range the power spectrum is set to zero. This random field is then transformed back into the real (x, y) plane. For the amplitude of velocity perturbations $\sigma = \langle \mathbf{u}^2 \rangle^{1/2}$, we initially take $\sigma = 0.6$. In the simulations presented below, we start with $Q_0 = c_{s0}\Omega/\pi G\Sigma_0 = 1.5$. Figure 3.9 shows these initial conditions in terms of PV in the real plane. From this figure we can see that in the real plane the initial PV field has a chaotic character with equal contributions from negative and positive values. Although we start with a Kolmogorov power spectrum, a general dynamical picture emerging in the quasi-steady state is almost independent on the specific initial conditions. These random velocity perturbations are meant to mimic residual turbulence in a disc resulting from the collapse of a molecular cloud core (see e.g., Godon and Livio, 2000).

Note that we do not impose any constraints on the initial fields u_x and u_y, such as the requirement of incompressibility ($\nabla \cdot \mathbf{u} = 0$) adopted by Johnson and Gammie (2005b) and Shen et al. (2006) in order to pick out only solenoidal perturbations (which are not necessarily vortical in the presence of background shear); so these two velocity components are not correlated. We found that the mere presence of negative PV values in the initial conditions is sufficient for the development of vortices, even though they may also contain some fraction of SDW perturbations. As shown in section 3.3.1, in shear flows, initially divergent-free vortical perturbations can become divergent and initially potential (curl-free) wave perturbations can become rotational (with non-zero velocity curl) due to the background velocity shear. But wave perturbations never acquire PV if it is not present initially (of course, here we assume that there are no baroclinic terms in the PV equation in the linear approximation, or

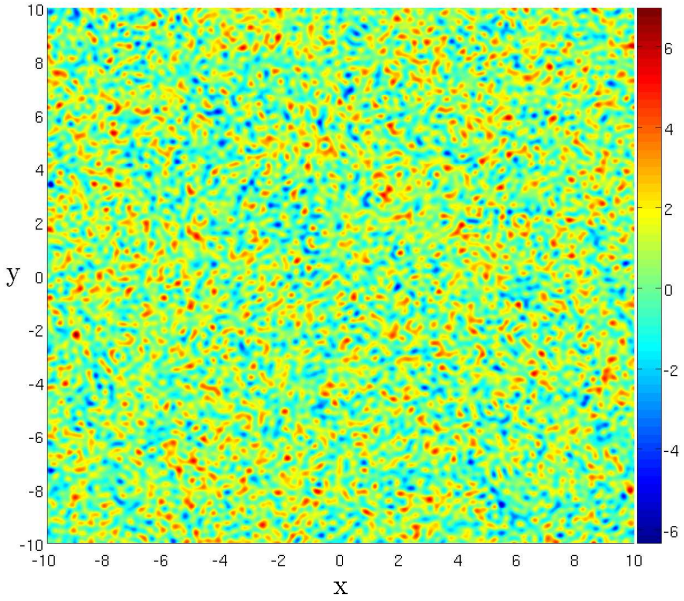

Figure 3.9: Initial (at $t = 0$) distribution of PV in the real coordinate plane corresponding to a Kolmogorov spectrum in the $\mathbf{K}$-plane for the fiducial model with $L = 20$. Both negative (blue and light blue) and positive (yellow and red) values of PV are present.

equivalently the unperturbed surface density and pressure are spatially uniform). Thus, as pointed out before, truly vortical perturbations should be characterised by non-zero PV even though they may develop divergence in the course of evolution. It is the PV that determines the vortex formation and it must be present initially. By choosing an initial perturbed velocity field with non-zero PV, we ensure the presence of vortical perturbations at the beginning. In other words, the class of initial conditions capable of causing vortex formation is broader; they must be characterised only by non-zero negative PV and may not necessarily satisfy the incompressibility requirement, that is, they can comprise a fraction of wave perturbations as well. This alleviates the problem of vorticity injection. Initial conditions resulting from the collapse of a molecular cloud into a disc, where PV is everywhere zero would be somewhat contrived and therefore less likely.

3.4.3 Non-linear evolution

In this section we present the non-linear evolution of the fiducial model with size $L = 20$, starting with the initial conditions described in the previous section. We examine in detail the development of structures in the PV, surface density and pressure/internal energy fields. The size $L = 20$ corresponds to a minimum wavenumber $K_{y\ min} = 2\pi/L = 0.314$ in the

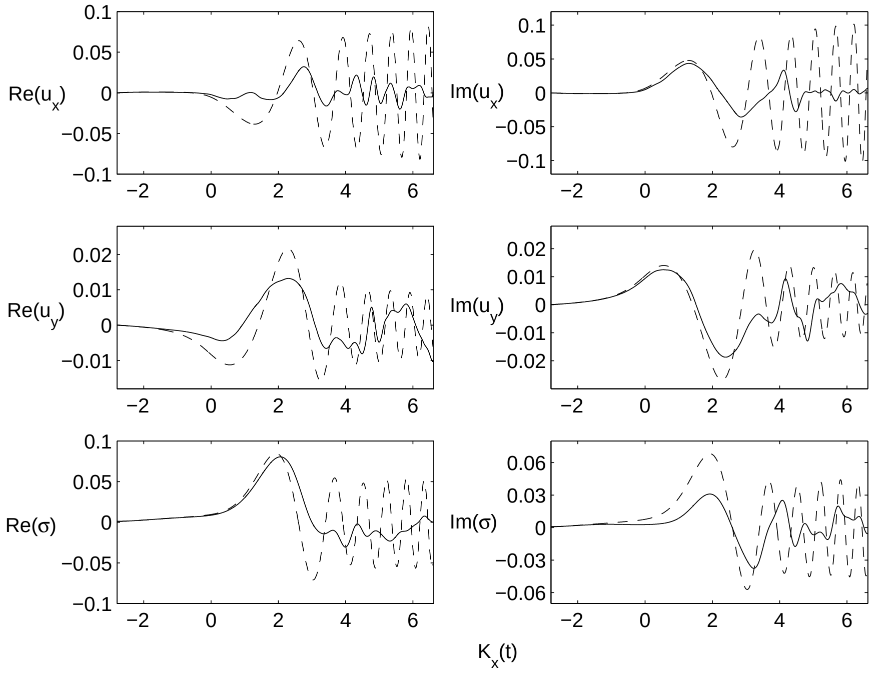

Figure 3.10: Non-linear evolution of the real and imaginary parts of the perturbed velocity and surface density amplitudes (solid lines) for a single vortical mode SFH with $K_x(0) = -18\pi/L,\ K_y = 4\pi/L$ ($L = 20$), which has been selected from the initial spectrum in the wavenumber plane, during its 'drift' along the K_x-axis from leading to trailing. For comparison, we also show with dashed lines the time-development of these quantities for the same (vortical) initial values computed with linearised Eqs. (3.1-3.5). We see that a qualitative picture of the temporal evolution of an individual SFH in the non-linear regime is similar to that in the linear regime (see Fig. 3.3): the vortical mode SFH initially undergoes transient-exponential growth and then in the vicinity of $K_x(t) = 0$ gives rise to oscillations, indicating the excitation of the SDW mode SFH. However, in contrast to the linear case, SDWs are damped after some time in the trailing phase, because they turn into shocks and thus are dissipated. A similar shock-damping of SDW mode SFHs generated by the vortical mode SFH, but in a MRI-turbulent disc shear flow, was also observed by Heinemann and Papaloizou (2009b).

computation domain for which linear swing growth is largest as a function of K_y at initially chosen $Q = 1.5$ (see Fig. 3.6). Thus, we ensure that from the outset our computational domain comprises those length-scales for which the effect of self-gravity is expected to be important.

In the linear regime in section 3.3, we demonstrated that in the presence of both strong Keplerian shear and self-gravity, the main mechanism responsible for the growth of initial perturbations is swing amplification instead of the pure Jeans instability. Swing amplification has a transient-exponential character because of the drift of perturbation SFHs through the unstable domains in the **K**-plane (see Fig. 3.1) that are brought about by the combined

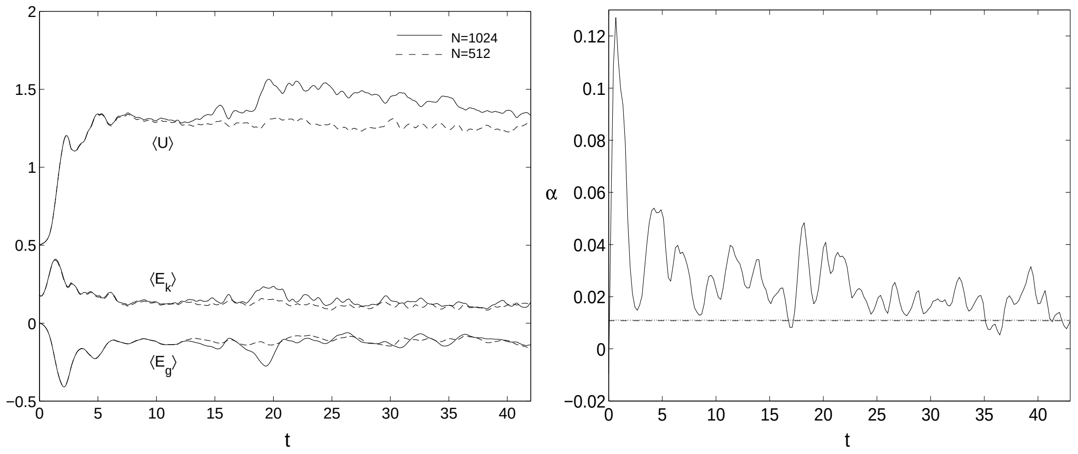

Figure 3.11: Evolution of the average kinetic, internal and gravitational energies (left) and the α parameter (right). During the burst phase, they undergo rapid transient amplification until $t \approx 2-2.5$ and then remain on average constant in the following quasi-steady gravitoturbulent phase, which sets in at about $t = 5$. The evolution of the same quantities for a lower resolution ($N = 512$) run are similar. We see that the time-average α settles down approximately to the value given by Eq. (3.20) (dashed line), which is 0.011 at $\tau_c = 20$ and $\gamma = 2$ adopted here.

action of shear and self-gravity (see also, e.g., Goldreich and Lynden-Bell, 1965b; Goldreich and Tremaine, 1978; Toomre, 1981; Kim and Ostriker, 2001). To illustrate the workings of the swing amplification of SFHs in the non-linear regime, following Heinemann and Papaloizou (2009b), we have picked out a single (vortical mode) SFH from the perturbation spectrum in the **K**-plane and traced its evolution as it swings from leading to trailing (Fig. 3.10). We see from this figure that the swing growth proceeds qualitatively similarly as in the linear regime, i.e., occurs around $K_x(t) = 0$ when the radial wavenumber changes sign, but is modified by non-linear effects mostly in the trailing phase, when SDW generation occurs (see below). As a result of the swing amplification of individual SFHs in the **K**-plane, at the beginning of evolution, the average values of various quantities undergo rapid growth over a relatively short time interval (Figs. 3.11 and 3.12; see also Gammie 2001). The evolution of the same quantities for a lower resolution $N = 512$ run are almost similar. The small difference between these two runs may be attributed to the specificity of PV, surface density and internal energy fields at different resolutions (see also Shen et al., 2006). We call this interval the burst phase. During swing amplification, the initial velocity perturbations induce strong surface density perturbations in the form of shocks with a density contrast of a factor of about 100. These shocks might originate from SDWs generated by swing-amplified vortical mode SFHs contained in the initial spectrum. Indeed, as is seen from Fig. 3.10, the vortex-wave coupling for individual SFHs in the non-linear state occurs qualitatively similarly to that in the linear

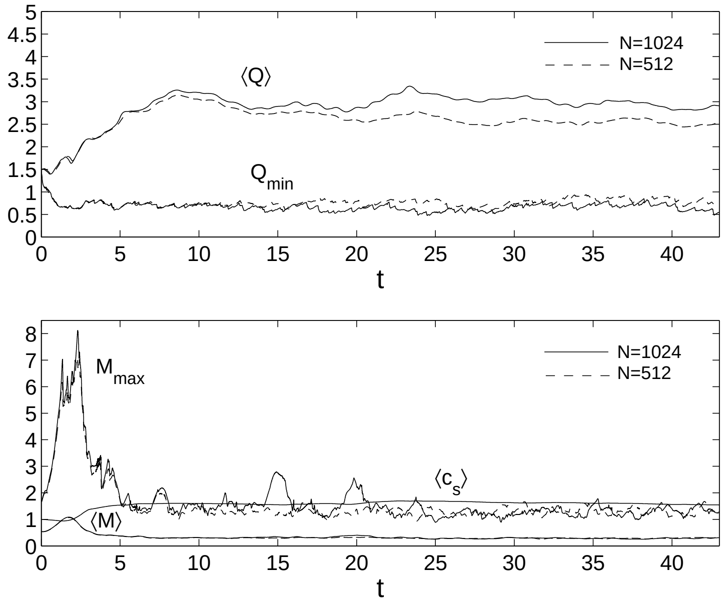

Figure 3.12: Evolution of $\langle Q\rangle$ and minimum Q_{min} (upper plot) and the average Mach number $\langle M\rangle$, maximum M_{max} and $\langle c_s\rangle$ (lower plot). $\langle Q\rangle$ initially rises due to strong shock heating in the burst phase, but then levels off at about 3 in the subsequent gravitoturbulent phase. Q_{min} remains constant and low 0.6-0.7 implying that there are always unstable regions associated with negative PV (vortices). $\langle M\rangle$ initially rises but then settles down to smaller values $\simeq 0.3$, so that gas motion is on average subsonic. However, M_{max} remains of order unity, showing that there is still some sonic motion. The average sound speed $\langle c_s\rangle$ also reaches a constant value in the gravitoturbulent phase.

regime described in section 3.3, except that generated SDW mode SFHs are damped after some time in the trailing phase due to shock formation (see also Heinemann and Papaloizou, 2009b). But, because of non-linearity, new leading vortical mode SFHs are re-generated, maintaining a non-decaying (at the expense of background shear flow) shock structure in the density field during an entire course of evolution (see also Fig. 3.13). For this reason, these shocks of SDWs have trailing orientation, because swing amplification always tends to produce trailing structures. The gas motion is also supersonic at this stage. It can be seen from Fig. 3.12 that the average and maximum Mach numbers reach largest values 1.1 and $7-8$ respectively. Shock regions are the sources of intense heating of the gas together with compressional heating and therefore the internal energy undergoes jumps in the shocks. Eventually, strong shocks heat the disc up, but cooling compensates for heating, and as a result a quasi-steady gravitoturbulent state is reached at about $t=5$, where the average kinetic, internal and gravitational energies as well as $\langle Q\rangle$, Q_{min}, the average Mach number $\langle M\rangle$, the maximum Mach number M_{max}, $\langle c_s\rangle$ and angular momentum transport parameter α fluctuate around constant in time values (Figs. 3.11 and 3.12). At these times, α inferred from the simulation is close to, but still slightly larger than the value given by Eq. (3.20) because of energy dissipation/loss on grid scales due to numerical diffusion, which has not been taken into account when deriving Eq. (3.20) (see also Gammie, 2001). Now $\langle M \rangle$ fluctuates around 0.3, so that on average the gas motion is subsonic, however, M_{max} fluctuates around 1.24 meaning that the motion is still sonic mostly in the vicinity of shocks. $\langle Q \rangle$ fluctuates around 3, but we will see below that the $Q(x, y, t)$ field is very inhomogeneous containing values as small as 0.6-0.7 associated with negative PV regions (anticyclonic vortices).

In Fig. 3.13 we trace the parallel evolution of PV, surface density, pressure (the same as internal energy for $\gamma = 2$) and Q fields in order to identify correlations between them and to compare them with those in non-self-gravitating discs. We emphasise from the beginning that the evolution is not as smooth and regular as it is in the non-self-gravitating case. As we will see below, in self-gravitating discs, vortices display more complex behaviour compared with that in non-self-gravitating ones, where small-scale vortices once formed, gradually smoothly merge into each other and form larger vortices.

During the burst phase, initially small-scale positive and negative PV regions get strongly sheared into strips, but negative PV (anticyclonic) regions start to wrap up into vortex-like structures towards the end of the burst phase due to non-linear Kelvin-Helmholtz instability (Lithwick, 2007). The positive PV regions remain sheared into strips showing no signs of vortex formation during the entire course of evolution. Thus, only anticyclonic regions are able to survive in shear flows by taking the form of vortices. As pointed out in section 2.2.4, this is a quite general result confirmed in many hydrodynamical as well as astrophysical simulations. As evident from Fig. 3.13, in the subsequent quasi-steady state, the overall dynamical picture of PV evolution is still irregular and varies on the dynamical time-scale. For this reason, we use the term ‘vortex’ in a broader sense, meaning generally a negative PV region and not only a vortex with a well-defined shape commonly occurring in non-self-gravitating discs. As demonstrated below, in a quasi-steady state, these negative PV regions are associated with underdense/underpressure and overdense/overpressure regions superimposed on the shocks that, for instance, may play a role in trapping dust particles.

Small-scale anticyclonic vortices initiated at the end of the burst phase continue to grow further in size in the quasi-steady state which is reached, as mentioned above, at about $t = 5$. Figure 3.13 presents a typical dynamical picture of such a gravitoturbulent state, which is quasi-stationary and remains unchanged (on average) with time during the entire course of evolution. Vortices with a large (by absolute value) negative central value of PV (blue dots in the PV field) have more or less vortex-like shape and in the density field correspond to the

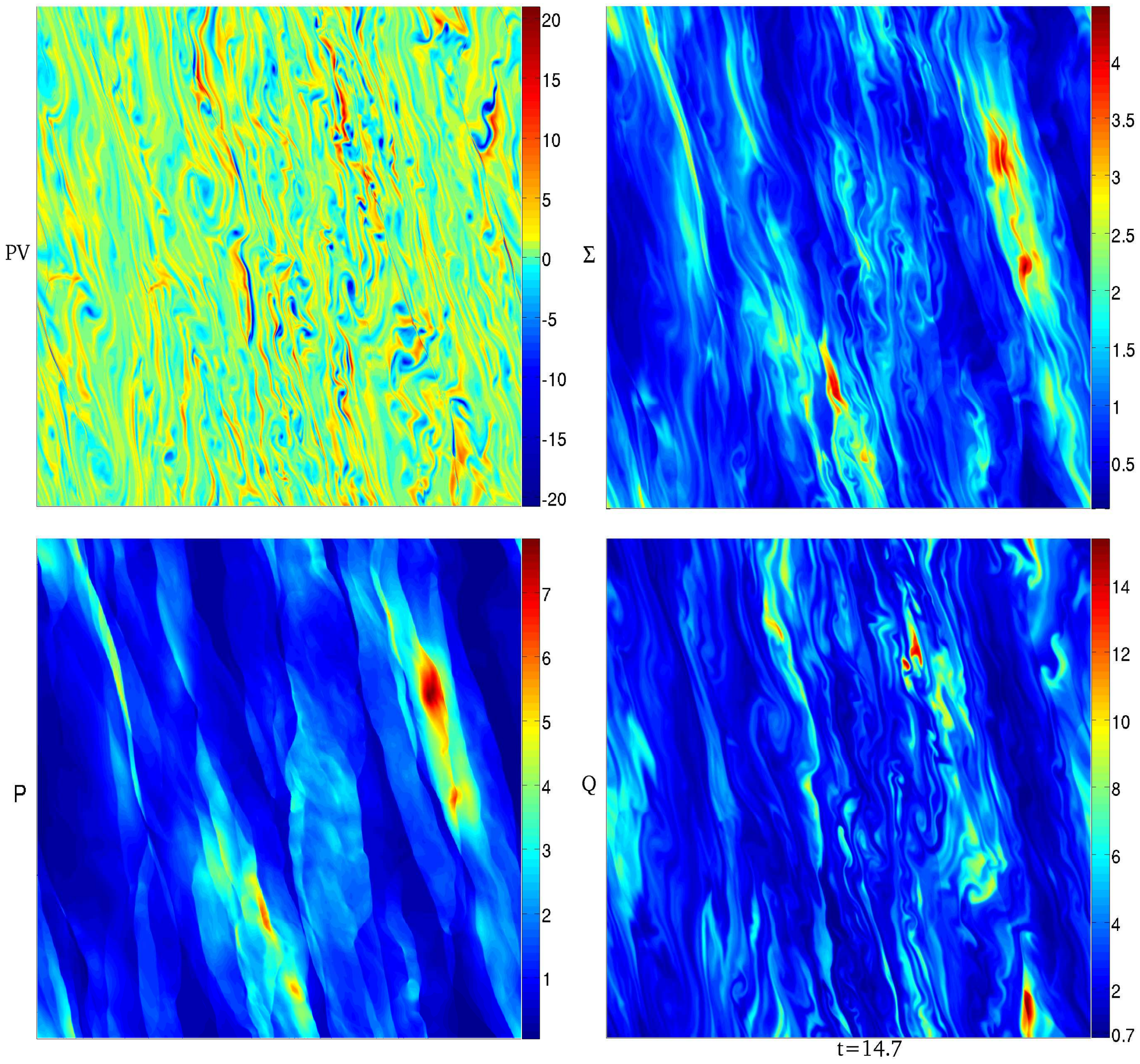

Figure 3.13: Evolution of PV, surface density, pressure/internal energy and Q parameter for the fiducial model in the gravitoturbulent state. Coordinate axes are same as in Fig. 3.9. Snapshots are at times $t = 14.7$ (this panel) and $t = 29.3, 44$ (next two panels). The PV field consists of vortices with different sizes and strengths (blue and light blue regions). Vortices with sizes smaller than the local Jeans scale have higher (by absolute value) negative PV centres (blue dots). They also emit SDWs during evolution that turn into shocks. In the surface density field, these vortices correspond to underdense central regions surrounded by overdense regions marking SDW emission places that are characterised by smaller (by absolute value) than the central values of PV. In the Q field, the underdense and overdense regions give rise to high and low Q values, respectively, but the influence of self-gravity on such small-scale vortices is not strong. The other type of vortices are somewhat larger having scales comparable to the local Jeans scale. They are more diffuse with no clear shape (light blue regions in PV field) and are characterised by smaller (by absolute value) negative PV, but stronger overdense regions (yellow and red in the surface density field) with even smaller Q than those in the previous case. For this reason, such vortices are in the process of being sheared and destroyed. In the pressure field, shocks and only the overpressure regions corresponding to stronger overdensities are noticeable (yellow and red). Exact values of these four quantities can be inferred from colour bars.

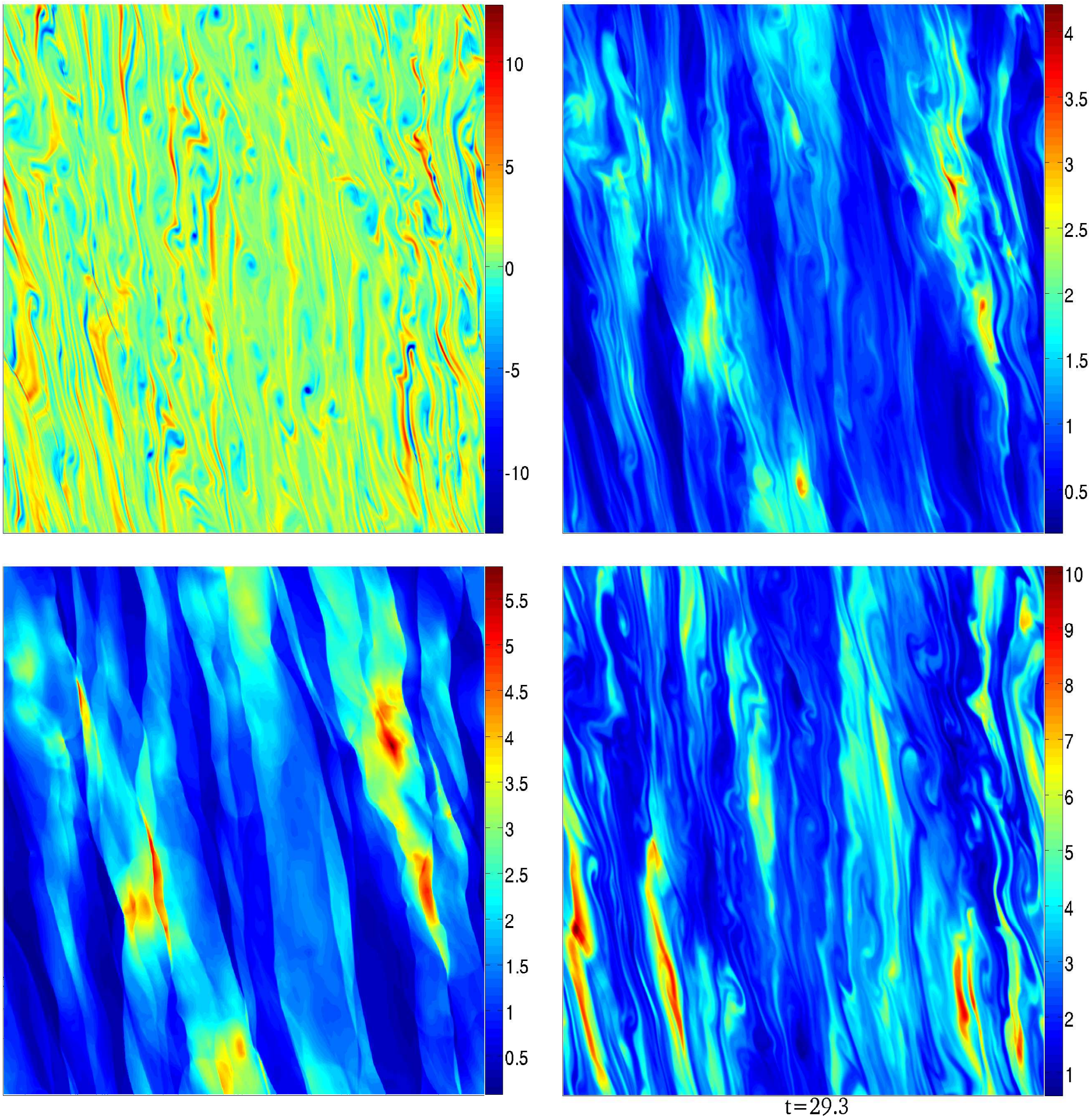

Figure 3.13 (continued): Snapshot at $t = 29.3$. The panels correspond to PV, Σ, P and Q fields in the same order as above.

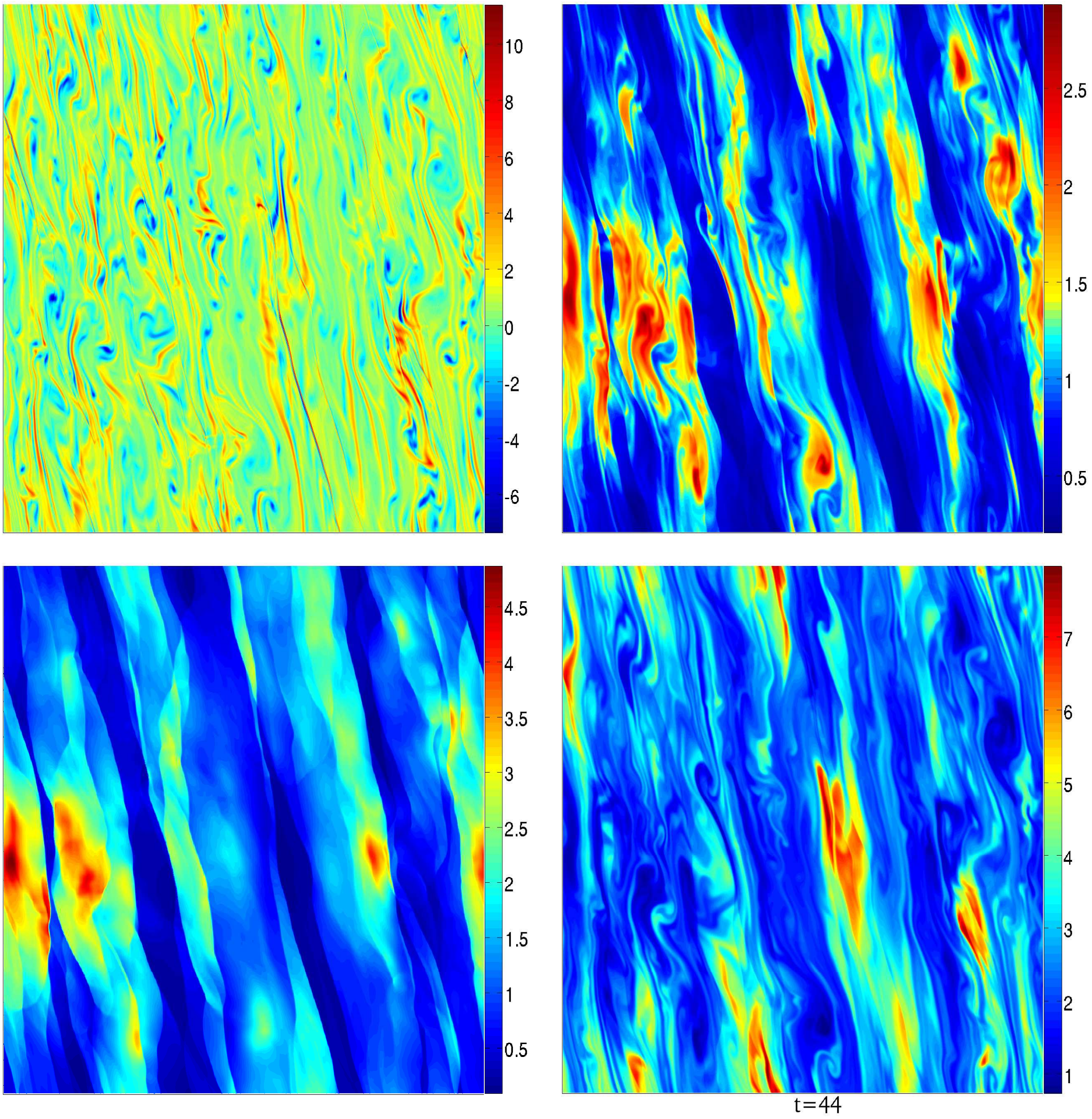

Figure 3.13 (continued): Snapshot at $t = 44$

central underdense regions surrounded by the higher density regions. Because of the vortex-wave coupling phenomenon, the total motion in the vicinity of such negative PV regions is a complex mixture of vortical and compressive (wave) motions giving rise to this type of density structure. Accordingly, in these enhanced density regions associated with compressive motions, PV is smaller by absolute value than that in the vortex centre. Propagating away from each negative PV region, these compressive motions (i.e., SDWs) turn into shocks afterwards that are clearly seen in the surface density and pressure fields in Fig. 3.13 (in fact, the shock structure in these fields is a net result of non-linear superposition/mixture of shock waves emitted by such individual vortices). We have shown in section 3.3.1 that because of shear, the divergence and the curl of perturbations become linked and change with time during the evolution (see Fig. 3.2). So, where angular momentum transport is concerned, it would be more precise to attribute it to the overall gas motion with non-zero PV and divergence rather than to compressive/wave or to vortical perturbations only (otherwise, to divergent or to solenoidal perturbations, as classified in section 3.3.1, only). A similar dynamical picture was also observed in the evolution/adjustment of a single vortex in a compressible non-self-gravitating Keplerian disc by Bodo et al. (2007). The vortex initially produces saddle-like density structure, much like observed in our simulations, with an underdense region at the location of the vortex centre surrounded by an overdense region with accompanying shock waves.

With time the effect of self-gravity on each vortex comes into play because of the growing length-scale of vortices, and at the same time of the increasing surface density/overdensity associated with them. The vortex growth in size is a consequence of the inverse – towards larger scales – cascade of energy characteristic of 2D flows as clearly demonstrated in other simulations (e.g., Godon and Livio, 1999a,b; Umurhan and Regev, 2004). The amplification of the vortex surface density is again due to the swing mechanism. For the same reason that in self-gravitating discs perturbations experience considerably larger swing growth than in non-self-gravitating discs (see Fig. 3.5), the traces of vortices in the surface density field – the overdense regions around underdense regions – become more and more noticeable on the background of density variations related to shocks. From Fig. 3.13 it is seen that overdensities are also characterised by small values of Q. In the pressure/internal energy field we see shocks and only overpressure regions. The stronger overdensities (yellow and red in Fig. 3.13) associated with the final stage of vortex life (see below) clearly correlate with these overpressure regions. However, it is still hard to find as good a correspondence/similar features between PV and pressure/internal energy fields as it is for the surface density field.

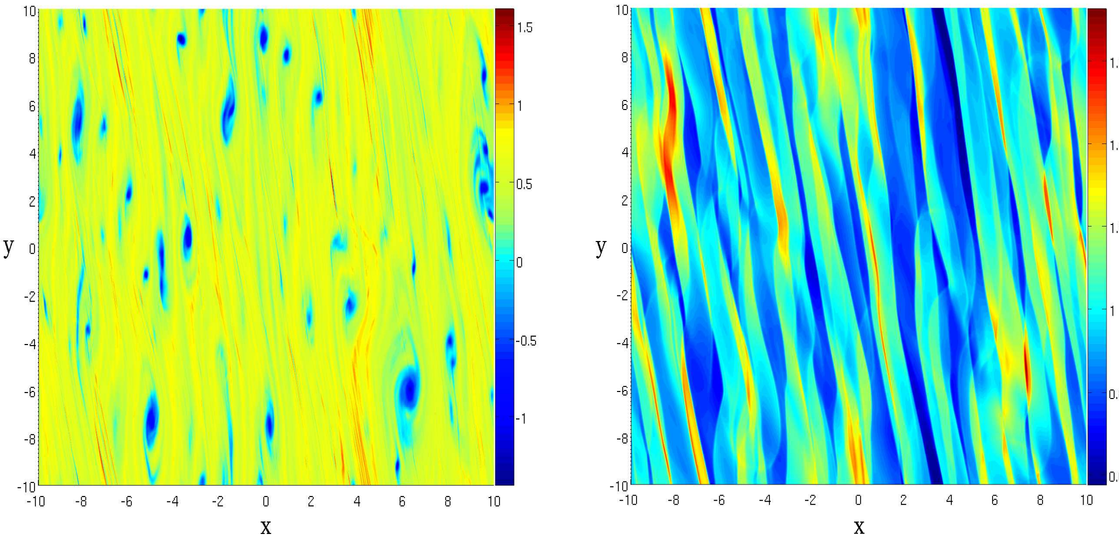

Figure 3.14: Typical PV (left) and the corresponding surface density (right) fields at $t = 44$ for the isothermal fiducial model with self-gravity switched off. The sound speed is taken to be equal to the average sound speed in the gravitoturbulent state (see Fig. 3.12). It is harder to see traces of individual vortices in the surface density field; only shocks shed by these vortices are visible. Vortices are organised into well-shaped larger scale structures as opposed to what is observed in the self-gravitating case in Fig. 3.13.

Nevertheless, in self-gravitating discs one can identify distinct features in the surface density and pressure/internal energy fields connected with vortices. By contrast, in analogous non-self-gravitating isothermal simulations the situation is different. As mentioned above, similar density structures (i.e., with overdense and underdense regions) were also observed in numerical simulations of a single vortex before it settled into an equilibrium configuration, where the anticyclonic vortex gives rise only to a single overdense region (Bodo et al., 2007). However, in other simulations of vortex formation from an initial random distribution of PV (Johnson and Gammie, 2005b; Shen et al., 2006), which are not limited to only a single vortex, overdense regions associated with individual vortices are hard to identify in the surface density field; there were only variations in the surface density due to the shocks shed by these vortices (see also Fig. 3.14). In our case instead, swing amplification due to self-gravity makes overdense and underdense regions clearly noticeable on the background of shocks.

Since $Q \leq 1$ at the location of the stronger overdense regions, self-gravity becomes dominant in the dynamics of vortices and appears to shear and deform them over about $1 - 3\Omega^{-1}$, and more importantly to inhibit their further growth in size. In other words, self-gravity opposes an inverse cascade of energy to smaller wavenumbers and scatters it to higher wavenumbers (see also Lyra et al., 2009). This fact is also reflected in the autocorrelation function of PV shown below (Fig. 3.15). The vortices can only grow to a size comparable to the local Jeans scale $\lambda_J \sim Qc_s$ (in non-dimensional form). When the vortex size approaches λ_J, it has no clear centre with high (by absolute value) negative PV region surrounded by smaller (by absolute value) PV region and, therefore, corresponds only to the stronger overdense region (yellow and red regions in the surface density field in Fig. 3.13) with no underdense centre. At this time, the shape of the vortex is more irregular, sheared and diffuse, characterised by about 10 times smaller (by absolute value) PV than that of growing vortices. From the pressure/internal energy field it is clear that in these stronger overdense regions the internal energy is higher as well and it is expected that the corresponding Q will eventually rise and switch off self-gravity/gravitational instability at that location. Then the overdense region will quickly disperse, or get sheared away. After that sheared PV regions (strips) with higher Q, in turn, can start another cycle of vortex formation by undergoing the same non-linear Kelvin-Helmholtz instability and again wrap up into vortices. Thus, in self-gravitating discs vortices have a recurring nature – each vortex forms, grows to a size of the order of the local Jeans scale and after that gets sheared out and destroyed by the combined action of self-gravity and Keplerian shear. Each cycle typically lasts for about two orbital periods or less (one orbital period= $2\pi/\Omega$). The cooling ensures that the minimum Q_{min} is kept low ($0.6 - 0.7$) and nearly constant with time, so that self-gravity continues to be active.

Above we have described the behaviour of some individual vortices only. From Fig. 3.13 we see that the PV field contains vortices with different sizes that never get organised into distinct coherent structures as they do in the non-self-gravitating case in Fig. 3.14. Vortices evolve differently – some of them are at the end of evolution having already grown to the Jeans scale sizes and are characterised by clearly identifiable overdense regions with low Q and smaller PV, but others have yet to go through this phase and, therefore, may still have underdense central regions surrounded by higher density regions and correspond to higher Q. Thus, regions of high- and low-Q coexist throughout the disc at all times. A similar gravitoturbulent state was observed in the global disc simulations by Wada et al. (2002) in the context of ISM turbulence. However, these authors did not measure PV corresponding to high- and low-Q regions. In general, the overall dynamical picture, as mentioned above, is very irregular and it is hard to keep track of individual vortices for several orbital periods, because they are short-lived structures. Figure 3.14 shows the stark contrast between the non-self-gravitating and the self-gravitating cases. In this figure, we present a snapshot of the same fiducial model, but without self-gravity and with constant sound speed (isothermal equation

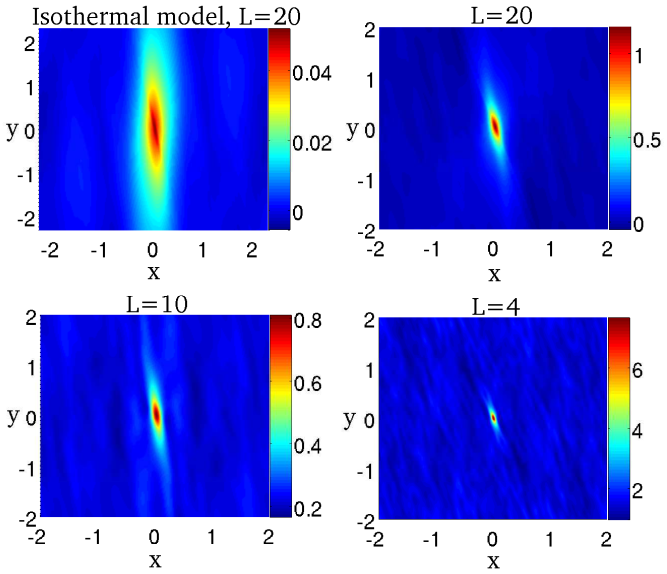

Figure 3.15: Autocorrelation functions of PV for four models: isothermal non-self-gravitating with $L = 20$ (Fig. 3.14) and three self-gravitating with $L = 20, 10, 4$ in the gravitoturbulent state. For the two models with the same $L = 20$ the correlation length is largest for the isothermal non-self-gravitating model and greatly decreases in the presence of self-gravity. It also decreases with decreasing L.

of state) equal to the average sound speed of the fiducial model in the gravitoturbulent state (see Fig. 3.12). In this case, the dynamical picture is identical to that described by other authors (e.g., Umurhan and Regev, 2004; Johnson and Gammie, 2005b; Shen et al., 2006). Now it is easier to trace each vortex and to see how they merge into larger vortices. In the self-gravitating case instead, irregular and chaotic phases of vortex evolution are quasi-steady, or equivalently, self-sustained during the entire course of evolution, whereas in the non-self-gravitating case, though vortices are well-organised and regular, they gradually decay – the corresponding kinetic energy falls off due to shock dissipation.

We also carried out simulations for two other models with sizes $L = 4$ and 10 and with the same type of initial conditions. The dynamical picture of vortex evolution described in detail for the fiducial model remains qualitatively the same. There are differences only in saturation times, and (average) values of various quantities in the gravitoturbulent state. In Fig. 3.15 we compare the autocorrelation functions of PV for these models with that of the non-self-gravitating isothermal model plotted in Fig. 3.14. The extent of the autocorrelation function, or the correlation length, remains on average constant with time in the gravitoturbulent state and does not depend on the spectrum of initial conditions. For the given model size $L = 20$, the correlation length is smaller than that in the non-self-gravitating case, implying that vortices are less coherent in the gravitoturbulent state (we remind the reader that the sound speed of the isothermal model, determining the correlation length for this model, is equal to the average sound speed of the fiducial self-gravitating model). This again can be explained by the tendency of self-gravity to oppose inverse cascade of energy to larger scales by scattering it to smaller scales and thereby broadening the spectrum of the PV autocorrelation function (see also Lyra et al., 2009). The correlation length appears to decrease with decreasing domain size, whereas in the isothermal case it does not depend on the domain size and is determined by the disc scale height (Johnson and Gammie, 2005b). In any case, correlation length is always smaller than the model size justifying the local treatment of vortices.

3.5 Summary and Discussions

In this chapter, we have studied the dynamics of vortical perturbations in self-gravitating discs with Keplerian differential rotation in the local shearing sheet approximation. To gain a first understanding of the effects of self-gravity on the vortical mode, we started with a linear regime. As commonly done in the shearing sheet, the linear dynamics has been investigated by decomposing perturbations into shearing waves, or spatial fourier harmonics (SFHs). Then perturbations were classified into two basic types/modes – spiral density waves (SDWs) characterised by zero potential vorticity (PV), which in fact are commonly focused on and thought to be the only perturbation type present in self-gravitating discs, and vortices, which are main carriers of non-zero PV. We concentrated on a vortical mode SFH and traced its dynamical evolution. It was shown that self-gravity considerably alters the dynamics of the vortical mode – its transient growth may be several orders of magnitude stronger than in the non-self-gravitating case and two to three times larger than the transient (swing) growth of the SDW mode. This suggests that in self-gravitating discs, vortical perturbations can be as important as SDWs, so that one must necessarily consider both these modes, not only SDWs, to get a proper and complete understanding of disc dynamics. We also analysed the influence of self-gravity on the recently found phenomenon of linear coupling between the vortical and SDW modes in non-self-gravitating discs caused by the differential character, or shear of disc flow (Bodo et al., 2005; Heinemann and Papaloizou, 2009a). This coupling is asymmetric: the vortical mode having non-zero PV is able to excite the SDW mode with zero PV, but not vice versa, because of the conservation of PV in the linear limit. We found that the efficiency of SDW generation by the vortical mode increases with increasing the degree of self-gravity. In other words, self-gravity, enhancing vortex-wave coupling, adds weight to the vortical perturbations as one of the key factors in determining the dynamical behaviour of SDWs and shocks in self-gravitating accretion discs.

Having analysed the linear evolution of each vortical mode SFH, we then studied the specific properties of vortex, that is, PV evolution under the action of self-gravity through numerical simulations, starting with a spectrum of vortical mode SFHs (with some mix of SDW mode SFHs) in the wavenumber plane. Since we are interested particularly in the properties of vortex evolution in a quasi-steady gravitoturbulence, we chose a simple cooling law with a constant cooling time large enough that the disc settles into this quasi-steady state. Our analysis showed that in the self-gravitating case, anticyclonic vortices appear as transient structures undergoing recurring phases of formation, growth to sizes comparable to the local Jeans scale, and eventual shearing and destruction due to the combined action of self-gravity (gravitational instability) and background Keplerian shear. Each such phase lasts for a few orbital periods. As a result, the overall dynamical picture of PV is irregular, consisting of many transient vortices at different evolutionary stages and, therefore, with various sizes up to the local Jeans scale. By contrast, in the non-self-gravitating case, long-lived vortices form and grow in size via merging into larger ones until eventually their size reaches the disc scale height. SDWs generated by these vortices due to the shear-induced vortex-wave coupling mechanism turn into shocks afterwards, which are clearly seen in the density field together with overdense and underdense regions produced by anticyclonic vortices (negative PV regions).

It is well known that overpressure/overdensity regions, if present in a disc, can act as traps for dust particles in protoplanetary discs (e.g., Haghighipour and Boss, 2003; Rice et al., 2004, 2006; Fromang and Nelson, 2005; Lyra et al., 2009). As has been demonstrated in our simulations, in self-gravitating discs, overpressure/overdensities produced by anticyclonic vortices are not long-lived, they have a transient character and vary on the dynamical time-scale. Hence, given such an irregular and rapidly changing nature of vortex evolution in self-gravitating discs, it seems difficult for corresponding overdensities/overpressures to effectively trap dust particles in their centres. Preliminary results (Gibbons et al., in preparation) on the evolution of a coupled system – dust particles embedded in a self-gravitating gaseous disc – have shown that particles can nevertheless be caught with small concentrations by these overdensities. But because these simulations have been performed without self-gravity for particles, such particle aggregations do not appear to persist long (do not seem to be bound) and dissolve together with the gas overdensity. So, one of the goals of my future research is to investigate whether inclusion of self-gravity for particles causes them to clump and remain bound once a parent overdensity has dissolved. In this connection, we should mention that particle self-gravity has already been shown to effectively produce bound clumps (planetesimals) of dust particles, which have been trapped within gas overdensities produced by MRI-turbulence in a compressible, non-self-gravitating disc flow (Johansen et al., 2007).

Here, we have considered the simplified case of a razor-thin (2D) disc in order to gain first insight into the effects of self-gravity on vortex dynamics. Obviously, for a fuller understanding 3D treatment is necessary. The situation can be different and more interesting in the 3D self-gravitating case, which is evidently stratified in the vertical direction. As mentioned in section 2.2.4, unstratified 3D simulations by Shen et al. (2006) showed that vortices are unstable and are quickly destroyed by the elliptical instability. The presence of stratification greatly reduces the growth rate of the elliptical instability, somewhat helping the survival of vortices (Lesur and Papaloizou, 2009). Indeed, Barranco and Marcus (2005) demonstrated that in the 3D non-self-gravitating stratified shearing box, off-midplane vortices appear and survive for many orbital periods. On the other hand, as we have seen here, self-gravity does not favour long-lived vortices. But in the 3D case, the effect of self-gravity is somewhat reduced compared with that in the 2D case considered here (see also chapter 5). Thus, in a 3D generalisation of the present problem, there are two competing factors – stratification and self-gravity – and it remains to be seen in future numerical experiments which of these two prevails.

In 2D discs, it is also interesting to investigate the influence of self-gravity on the development of the baroclinic and Rossby wave instabilities. As described in section 2.2.4, these instabilities, generating PV, lead to the formation of large-scale organised vortices, while self-gravity tends to oppose this process. So, as in the above case of stratification, the aim will be to see which out of these two factors prevails and how the properties of resulting vortices differ from those in non-self-gravitating discs.

Another question of interest in the 3D case (either self-gravitating or not) is, as also pointed out by Johnson and Gammie (2005b), whether long-lived off-midplane vortices can be generated from a random PV distribution, as happens in the 2D case, rather than from specially chosen vortex solutions as in simulations of Barranco and Marcus (2005). All these are a subject of my future research.

CHAPTER 4

Excitation of spiral density waves by convection in discs

4.1 Introduction

In section 2.2.2, we outlined the latest progress in the dynamics and transport properties of convection in protoplanetary discs. We saw that recent high-resolution simulations by Lesur and Ogilvie (2010) and Käpylä et al. (2010) have demonstrated, contrary to the belief held in the past, that convective motions perpendicular to the disc midplane can actually lead to outward transport of angular momentum. This outward transport has been attributed to the non-axisymmetric structure of convection, which is established at sufficiently high Rayleigh numbers (i.e., at small molecular viscosity and/or thermal diffusivity). However, these simulations are incompressible and consequently the values of the α parameter reported, though positive, are still insufficient to match typical accretion time-scales of discs. On the other hand, in a more realistic case where compressibility of gas is taken into account, in a 3D disc with a superadiabatic vertical stratification, there obviously also exists the SDW mode due to compressibility and rotation, together with the convective mode due to the negative entropy gradient. As already mentioned in section 2.4, because these convective and SDW modes evolve in the sheared environment of the disc's Keplerian differential rotation, we should expect non-normality/shear-induced effects to come into play in the dynamics of these two perturbation modes too, despite the fact that they are characterised by different time-scales. Specifically, as we demonstrated in section 3.3.2, as a result of the shear-induced linear mode coupling phenomenon, a slowly varying vortical mode can efficiently generate rapidly oscillating SDWs if the horizontal wavelength of these modes is comparable to the disc scale height, which, in turn, implies that compressibility effects are important at such length-scales. Because usually convective motions in discs extend vertically over characteristic length-scales comparable to the disc scale height (e.g., Ruden et al., 1988; Kley et al., 1993; Cabot, 1996; Stone and Balbus, 1996), implying that compressibility should be an important ingredient in their dynamics, it is natural to suppose that convection, like vortices, can be dynamically coupled with SDWs owing to disc flow shear. In this way, SDWs could also participate in dynamical processes in convectively unstable discs and their effects must be properly understood.

Thus, our goal in this chapter is to investigate another manifestation of the shear-induced linear mode coupling phenomenon – generation of SDWs by convection. As we will demonstrate, Keplerian shear of the disc's differential rotation indeed causes these two modes to become linearly coupled (provided they are non-axisymmetric), in a similar way as SDWs are coupled to the vortical mode. So, after vortices, convection can be regarded as yet another new source of SDWs in discs. This can have implications for the angular momentum transport by convection, since SDWs are generally able to enhance angular momentum transport rates, as already demonstrated for vortices in compressible discs (see e.g., Johnson and Gammie, 2005b; Oishi and Mac Low, 2009). However, we omit the vortical mode from the present analysis, because an exponentially growing convective mode is more powerful and dominates the algebraic growth of the vortical mode (unless the disc is self-gravitating). Following an analogous work by Ryu and Goodman (1992, hereafter RG92) on the linear dynamics of non-axisymmetric convection, we adopt the local shearing box approach and, as in section 3.3, by numerically analysing the linear evolution of individual non-axisymmetric shearing waves of perturbations, describe in detail the process of generation of the SDW mode by the convective mode. We quantify the efficiency of the SDW generation as a function of the azimuthal and vertical wavelengths of these modes and show that it is maximal and powerful when both these wavelengths are comparable to the disc scale height, that is, at wavelength for which compressibility is important. Finally, we would like to note that SDW generation can also be seen in the linear analysis of RG92 and also in a related linear study of the growth of non-axisymmetric shearing waves of the convective mode by Brandenburg and Dintrans (2006), though these authors did not specifically focus on and characterise this phenomenon. In a different context, a similar linear process of acoustic wave excitation in convectively unstable non-rotating shear flows was also found and analysed by Tevzadze (2006).

This chapter is organised as follows. The physical model and basic equations are introduced in section 4.2. Shear-induced linear coupling between SDW and convective modes is investigated in detail in section 4.3. The angular momentum transport properties of these two modes is analysed in section 4.4. A summary and discussion are given in section 4.5.

4.2 Physical Model and Equations

To study the dynamics of perturbation modes existing in compressible and stratified gaseous Keplerian discs, we adopt a local shearing box approximation introduced in section 2.3. In the shearing box model, curvature effects due to cylindrical geometry are ignored and the disc dynamics is studied in a local Cartesian reference frame co-rotating with the disc at some fiducial radius r_0 from the central star (Fig. 2.5). The Coriolis force is included to take into account the effects of the coordinate frame rotation. The vertical component of the gravity force of the central star is also present. For simplicity, we neglect disc self-gravity in this chapter, as it does not qualitatively affect the linear coupling phenomenon of convection and SDWs that we wish to bring out here. In this local reference frame, the basic momentum and continuity equations are Eqs. (2.12) and (2.14), which we give here again for completeness

$$\frac{\partial \mathbf{u}}{\partial t} - q\Omega x \frac{\partial \mathbf{u}}{\partial y} + (\mathbf{u}\cdot\nabla)\mathbf{u} - q\Omega u_x \mathbf{e}_y + 2\Omega \mathbf{e}_z \times \mathbf{u} + \Omega^2 z \mathbf{e}_z = -\frac{1}{\rho}\nabla p, \tag{4.1}$$

$$\frac{\partial \rho}{\partial t} - q\Omega x \frac{\partial \rho}{\partial y} + \nabla\cdot(\rho\mathbf{u}) = 0. \tag{4.2}$$

where we have omitted the gravitational potential due to disc self-gravity. We supplement these equations with the equation of entropy conservation

$$\left(\frac{\partial}{\partial t} - q\Omega x \frac{\partial}{\partial y} + \mathbf{u}\cdot\nabla\right)\frac{p}{\rho^\gamma} = 0. \tag{4.3}$$

Here $\mathbf{u} = (u_x, u_y, u_z)$ is the perturbed velocity relative to the background Keplerian shear flow $\mathbf{u}_0 = (0, -q\Omega x, 0)$ (we recall that $q = 1.5$ in the case of Keplerian rotation), ρ, p are, respectively, the density and pressure of gas, Ω is the angular velocity of the local rotating reference frame, which is equal to the disc angular velocity at r_0, $\Omega = \Omega_0(r_0)$ and γ is the adiabatic index.

4.2.1 The equilibrium disc model

An unperturbed equilibrium state represents a stationary, axisymmetric, compressible and stratified Keplerian disc. The velocity field of Keplerian rotation, as mentioned before, appears in the local frame as the parallel azimuthal flow, $\mathbf{u_0}$, with a constant radial shear q. As noted in section 2.3, in the shearing box model, equilibrium density ρ_0 and pressure p_0 depend only on the vertical $z-$coordinate and satisfy the hydrostatic relation

$$\frac{1}{\rho_0}\frac{dp_0}{dz} + \Omega^2 z = 0, \tag{4.4}$$

according to Eq. (4.1) at zero perturbed velocity ($\mathbf{u} = 0$). Following RG92 and Brandenburg and Dintrans (2006), we assume the unperturbed disc to be vertically isothermal with a constant adiabatic sound speed

$$c_s = \sqrt{\gamma \frac{p_0}{\rho_0}}. \tag{4.5}$$

For simplicity, we use the thin disc approximation with constant gravity (see e.g., Shu, 1974; Tevzadze et al., 2003, RG92). In other words, the vertically varying gravitational acceleration $\Omega^2 z$ is replaced with its some height-averaged value $g > 0$, but because the former has opposite signs on either side of the disc midplane $z = 0$, we write

$$\Omega^2 z \rightarrow \text{sign}(z) g. \tag{4.6}$$

Although this simplification is highly idealised, it greatly eases the mathematical treatment allowing us to make Fourier analysis of perturbations in the vertical direction too in addition to horizontal decomposition into shearing waves (see below). This, in turn, makes it possible to clearly understand the key effects of shear on the dynamics of perturbation modes with various vertical and horizontal length-scales in a stratified disc and avoids the need to solve complex partial differential equations with respect to time t and the vertical coordinate z. In this case of constant g, substituting replacement (4.6) into Eq. (4.4), we find the distribution of the equilibrium density and pressure with z:

$$\frac{\rho_0(z)}{\rho_m} = \frac{p_0(z)}{p_m} = \exp\left(-\frac{|z|}{H}\right), \tag{4.7}$$

where

$$H = \frac{c_s^2}{\gamma g}$$

is the vertical stratification scale-height of the disc. $\rho_m \equiv \rho_0(0)$ and $p_m \equiv p_0(0)$ are the midplane values of the equilibrium density and pressure. As mentioned above, g is some height-averaged value of acceleration and we choose g^2 to be equal to the density-weighted average of $(\Omega^2 z)^2$, where as a weight function we use an exact expression for the equilibrium density, $\rho_m \exp(-\gamma\Omega^2 z^2/2c_s^2)$, which is obtained from Eqs. (4.4) and (4.5) with a linearly increasing gravitational acceleration $\Omega^2 z$ (RG92). This relates to the angular velocity through

$$\Omega^2 = \frac{\gamma g^2}{c_s^2}.$$

The Brunt-Väisälä frequency squared is defined as

$$N_0^2 \equiv \frac{g}{\rho_0}\left(\frac{1}{c_s^2}\frac{dp_0}{dz} - \frac{d\rho_0}{dz}\right) = \frac{\gamma-1}{\gamma}\Omega^2.$$

As is well-known, this frequency describes the buoyancy of disc vertical structure. If $\gamma > 1$, then $N_0^2 > 0$ (subadiabatic thermal stratification) and the equilibrium vertical structure of the disc is convectively stable. For $\gamma < 1$, then $N_0^2 < 0$ (superadiabatic thermal stratification) that corresponds to a convectively unstable equilibrium. For $\gamma = 1$ (adiabatic thermal stratification), $N_0^2 = 0$ and all motions/modes due to buoyancy disappear. Since our goal here is to investigate the linear coupling between vertical convection and SDWs and the possible role of this phenomenon in the disc dynamics and angular momentum transport, we focus, as in RG92, only on the superadiabatic vertical structure choosing $\gamma = 0.8$ throughout this chapter. This value of γ is close to unity reflecting the fact that the degree of superadiabaticity is not very large in protoplanetary discs (Rafikov, 2007). Strictly speaking, the case $\gamma < 1$ does not make much thermodynamic sense, because for specific heats at constant pressure and constant volume it implies $C_V > C_P$. In spite of this, there is no problem with entropy equation (4.3). In principle, the convectively unstable regime can be modelled without using the condition $\gamma < 1$, if we appropriately choose cooling and heating functions in this equation (see also RG92). So, a simple γ-prescription mimics the basic features of convectively stable/unstable discs well without introducing additional heating and cooling functions that might unnecessarily complicate the situation.

4.2.2 Perturbation equations

Consider now small perturbations to the equilibrium state (4.7) with the background velocity filed $\mathbf{u_0}$. Linearising Eqs. (4.1-4.3) about this state, we obtain the following system governing the perturbation dynamics

$$\frac{DV_x}{Dt} = -\frac{\partial p'}{\partial x} + 2\Omega V_y, \tag{4.8}$$

$$\frac{DV_y}{Dt} = -\frac{\partial p'}{\partial y} + (q-2)\Omega V_x, \tag{4.9}$$

$$\frac{DV_z}{Dt} = -\frac{\partial p'}{\partial z} - g\rho', \tag{4.10}$$

$$\frac{D\rho'}{Dt} + \nabla\cdot\mathbf{V} = 0, \tag{4.11}$$

$$\frac{Dp'}{Dt} + c_s^2\nabla\cdot\mathbf{V} + \frac{c_s^2 N_0^2}{g}V_z = 0, \tag{4.12}$$

where

$$\frac{D}{Dt} \equiv \frac{\partial}{\partial t} - q\Omega x\frac{\partial}{\partial y}$$

and $\mathbf{V} \equiv \rho_0 \mathbf{u}$ with ρ', p' being the perturbations of density and pressure, respectively.

The form of Eqs. (4.8-4.12) permits a similar decomposition of the perturbed quantities into shearing plane waves, or spatial Fourier harmonics (SFH) with time-dependent amplitudes and phases, as done in the previous chapter

$$F(\mathbf{r},t) = F(t)\mathrm{exp}\left(-\frac{|z|}{2H}\right)\mathrm{exp}(\mathrm{i}k_x(t)x + \mathrm{i}k_y y + \mathrm{i}k_z z), \tag{4.13}$$

$$k_x(t) = q\Omega k_y t,$$

where $F \equiv (\mathbf{V}, \rho', p')$ (henceforth, primes will be omitted). The azimuthal k_y and vertical k_z wavenumbers remain unchanged, whereas the radial wavenumber $k_x(t)$ varies with time at a constant rate $q\Omega k_y$ if $k_y \neq 0$ (i.e., for non-axisymmetric perturbations) due to sweeping of wave crests by the background shear flow. For convenience, in $k_x(t)$ we have shifted the origin of time towards negative values, so that for $t = 0$, $k_x(0) = 0$ as well. In this case, an initially leading SFH with $k_x(t)/k_y < 0$ at $t < 0$, eventually becomes trailing with $k_x(t)/k_y > 0$ at $t > 0$. The exponential factor involving $-|z|/2H$ in Eq. (4.13) is necessary in order to account for the fact that perturbation energy would otherwise increase exponentially with height due to stratification. With this form, the perturbation energy, proportional to $\rho_0 u^2$, is vertically uniform when averaged over the vertical wavelength (see also Lerche and Parker, 1967, RG92). The inclusion of this exponential factor also ensures that there are no imaginary terms in the dispersion relation derived below that could lead us to infer some sort of spurious instability. For further use, we combine $-1/2H$ in the exponential factor and the real vertical wavenumber k_z into a single complex total vertical wavenumber $\tilde{k}_z \equiv k_z + \mathrm{i}\cdot\mathrm{sign}(z)/2H$.

Substituting Eq. (4.13) into Eqs. (4.8-4.12) and introducing new quantities

$$\mathrm{i}p \rightarrow p, \qquad \mathrm{i}\rho \rightarrow \rho,$$

$$v = (c_s^2\tilde{k}_z^* + \mathrm{i}g)V_z,$$

$$s = (\mathrm{i}c_s^2\tilde{k}_z^* - g)\frac{p - c_s^2\rho}{(\gamma-1)g},$$

where $\tilde{k}_z^*$ is the complex conjugate of $\tilde{k}_z$ and s corresponds to the entropy perturbation, we arrive at the following system of first order ordinary differential equations that govern the linear dynamics of SFHs of perturbations

$$\frac{dV_x}{dt} = -k_x(t)p + 2\Omega V_y, \tag{4.14}$$

$$\frac{dV_y}{dt} = -k_y p + (q-2)\Omega V_x, \tag{4.15}$$

$$\frac{dv}{dt} = (N_0^2 - c_s^2\bar{k}_z^2)p - N_0^2 s, \tag{4.16}$$

$$\frac{dp}{dt} = c_s^2(k_x(t)V_x + k_yV_y) + v, \tag{4.17}$$

$$\frac{ds}{dt} = v, \tag{4.18}$$

where $\bar{k}_z^2 \equiv |\tilde{k}_z|^2 = k_z^2 + 1/4H^2$ is defined as the square of the magnitude of the total vertical wavenumber. Note that the minimum value $\bar{k}_{z,min} = 1/2H$ corresponds to vertical length-scales of perturbations comparable to the disc scale height H. By taking advantage of the new quantities v and s, we have avoided complex coefficients in Eqs. (4.14-4.18) and therefore without loss of generality, we can assume $V_x,\ V_y,\ v,\ p$ and s to be real.

It can be readily shown that the above system possesses an important time invariant – the linearised potential vorticity

$$I \equiv k_x(t)V_y - k_yV_x + (2-q)\Omega\frac{p-s}{c_s^2} = const,$$

which follows from a more general Ertel's law of potential vorticity conservation in 3D. As in 2D case (see Eq. 3.15), this time invariant I, in turn, indicates the existence of the vortical/aperiodic mode in the perturbation spectrum, which is characterised by non-zero potential vorticity ($I \neq 0$) and represents a stationary (time-independent) solution of Eqs. (4.14-4.18) in the absence of shear. In the present 3D case, this vortical mode originates from the combined action of the vertical gravity force (stratification) and the Coriolis force and represents a 3D generalisation of the 2D vortical mode considered in chapter 3. As shown in chapters 2 and 3, because of the shear/non-normality of disc flow, the vortical mode can undergo large transient amplification and effectively (linearly) couple with and excite other modes (SDWs, inertia-gravity waves, baroclinic modes, etc.) existing in the disc (Tevzadze et al., 2003; Bodo et al., 2005; Mamatsashvili and Chagelishvili, 2007; Tevzadze et al., 2008, 2010; Heinemann and Papaloizou, 2009a). However, in the present problem where disc self-gravity is absent, the transient (algebraic) growth of the vortical mode is overwhelmed by the exponential growth of the convective mode. For this reason, the vortical mode is not of much importance in convectively unstable discs. So, we exclude the latter from our analysis by setting the linearised potential vorticity to zero, $I = 0$.

Taking into account that by our choice $I = 0$, we eliminate $V_x,\ V_y,\ v$ in Eqs. (4.14-4.18) and after switching to different quantities

$$\hat{s} = s, \qquad \hat{p} = \frac{p-s}{k_\perp(t)},$$

arrive at the second order system

$$\frac{d^2\hat{s}}{dt^2} = (N_0^2 - c_s^2\bar{k}_z^2)k_\perp(t)\hat{p} - c_s^2\bar{k}_z^2\hat{s}, \tag{4.19}$$

$$\frac{d^2\hat{p}}{dt^2} = -(c_s^2 k_\perp^2(t) + \kappa^2(t))\hat{p} - c_s^2 k_\perp(t)\hat{s}, \tag{4.20}$$

where

$$\kappa^2(t) \equiv 2(2-q)\Omega^2 - \frac{4q\Omega^2 k_y^2}{k_\perp^2(t)} + \frac{3q^2\Omega^2 k_y^4}{k_\perp^4(t)},$$

and $k_\perp^2(t) = k_x^2(t) + k_y^2 = k_y^2(1 + q^2\Omega^2 t^2)$, $k^2(t) = k_\perp^2(t) + \bar{k}_z^2$. These equations are the basis for our subsequent analysis. They contain all the information on various perturbation modes present in a stratified disc apart from the vortical mode. In the next section, we remove shear from these equations for the moment in order to classify modes in the disc and then put shear back again to see how it alters the mode dynamics and what new effects it introduces.

4.2.3 Classification of modes in the absence of shear

Consider first a simple case without shear (i.e., a rigidly rotating disc) by setting $q = 0$ in Eqs. (4.19) and (4.20). After that, all the coefficients in these equations become time-independent and, therefore, we can look for solutions in the form $\hat{p}, \hat{s} \propto \exp(\mathrm{i}\omega t)$. Substituting this into Eqs. (4.19) and (4.20), we obtain the following dispersion relation in the absence of shear

$$\omega^4 - (c_s^2 k^2 + \kappa^2)\omega^2 + c_s^2(N_0^2 k_\perp^2 + \kappa^2 \bar{k}_z^2) = 0,$$

where $\kappa^2 = 4\Omega^2$ is constant when $q = 0$. This dispersion relation has two different solutions corresponding to two different types of perturbation modes:[1]

1. A high-frequency acoustic mode with

$$\omega_s^2 = \frac{c_s^2 k^2 + \kappa^2}{2}\left(1 + \sqrt{1 - \frac{4c_s^2(N_0^2 k_\perp^2 + \kappa^2 \bar{k}_z^2)}{(c_s^2 k^2 + \kappa^2)^2}}\right) \tag{4.21}$$

   the restoring force for which is mainly provided by compressibility/pressure forces, but is modified by stratification/bouyancy and the Coriolis force. So, in this mode, pressure perturbations dominate over entropy perturbations. For large wavenumbers $kH \gg 1$ (i.e., for wavelengths much smaller than the disc scale height), the effects of stratification and rotation are negligible and the frequency of the acoustic mode reduces to $\omega_s^2 \simeq c_s^2 k^2$. In an unstratified disc, for z-independent perturbations (i.e., in the razor-thin disc approximation), from Eq. (4.21) we obtain $\omega_s^2 = c_s^2(k_x^2 + k_y^2) + \kappa^2$, which is a classical dispersion relation of two-dimensional SDWs in non-self-gravitating discs (cf. Eq. 2.5 without self-gravity, see also Balbus, 2003). From this it follows that

---

[1] If we included the vortical mode, it would correspond, as noted above, to the stationary solution $\omega = 0$, since $I \neq 0$ for this mode.

in non-self-gravitating discs, the SDW mode is actually a 2D counterpart of the 3D acoustic (compressible) mode, also referred to as p-mode. So, for convenience, in this chapter we call the latter mode SDW too from now on, regarding expression (4.21) as a generalised form of the SDW mode frequency in the presence of vertical stratification. However, as will be demonstrated in chapter 5, situation is drastically different in stratified self-gravitating discs, where we can no longer unambiguously identify SDWs with the acoustic p-mode.

2. A low-frequency mode, which we call inertia-buoyancy mode, with

$$\omega_g^2 = \frac{c_s^2 k^2 + \kappa^2}{2}\left(1 - \sqrt{1 - \frac{4c_s^2(N_0^2 k_\perp^2 + \kappa^2 \bar{k}_z^2)}{(c_s^2 k^2 + \kappa^2)^2}}\right) \tag{4.22}$$

   the restoring force for which is mainly provided by vertical buoyancy and the Coriolis force, but modified by compressibility. So, in this mode, entropy perturbations dominate over pressure perturbations. The inertia-buoyancy mode represents an inertia-gravity wave in the case of subadiabatic and adiabatic stratifications ($N_0^2 \geq 0$) and the convective mode in the case of superadiabatic stratification ($N_0^2 < 0$), which is considered here.[2] The dynamics of inertia-gravity waves in Keplerian discs was studied elsewhere (see e.g., Lubow and Pringle, 1993b; Ogilvie, 1998; Balbus, 2003; Tevzadze et al., 2003, 2008; Latter and Balbus, 2009) and will not be dealt with here. Again, in the limit of large wavenumbers $kH \gg 1$, the effect of compressibility on the inertia-buoyancy mode is small and we get $\omega_g^2 \simeq N_0^2 \frac{k_\perp^2}{k^2} + \kappa^2 \frac{\bar{k}_z^2}{k^2}$, which can also be derived using the incompressible (Boussinesq) approximation (Tevzadze et al., 2008).

Thus, in a stratified and compressible disc, perturbation modes can be classified into two basic types – the acoustic (SDW) and inertia-buoyancy modes (apart from the vortical mode) and any general perturbation (with zero potential vorticity) can be decomposed into the sum of these two modes.[3] To analyse the behaviour of these two modes, we define new characteristic quantities, or eigenfunctions, ψ_s and ψ_g, for them as

$$\psi_s = \frac{(c_s^2 \bar{k}_z^2 - N_0^2) k_\perp \hat{p} - (\omega_g^2 - c_s^2 \bar{k}_z^2)\hat{s}}{\omega_s^2 - \omega_g^2}, \tag{4.23}$$

---

[2] Although we call this a convective mode, it actually exhibits convective instability, i.e. it has $\omega_g^2 < 0$, only when the wavenumbers satisfy $\kappa^2 \bar{k}_z^2 < -N_0^2 k_\perp^2$ in order to overcome the stabilising effect of rotation.

[3] Strictly speaking, the mode classification done in this section is applicable in the case of the constant-g approximation (Eq. 4.6). A more general classification of vertical modes in stratified discs without invoking this approximation will be done in section 5.3 of the next chapter. Nevertheless, the classification adopted here allows us to grasp the key effects of shear on the dynamics of the inertia-buoyancy (convective) and acoustic (SDW) modes.

$$\psi_g = \frac{(c_s^2\bar{k}_z^2 - N_0^2)k_\perp \hat{p} - (\omega_s^2 - c_s^2\bar{k}_z^2)\hat{s}}{\omega_g^2 - \omega_s^2} \tag{4.24}$$

describing, respectively, the SDW and inertia-buoyancy modes. These eigenfunctions are convenient and physically revealing, because substituting them into Eqs. (4.19) and (4.20) without shear, we obtain two separate equations for each mode eigenfunction

$$\frac{d^2\psi_s}{dt^2} + \omega_s^2\psi_s = 0, \qquad \frac{d^2\psi_g}{dt^2} + \omega_g^2\psi_g = 0 \tag{4.25}$$

that allow us to study these modes individually. All the other quantities (V_x, V_y, v, ρ, p) can be expressed through the modal eigenfunctions ψ_s, ψ_g and their respective first order time derivatives. Hence, we can fully determine the perturbation field corresponding to a specific mode by setting the eigenfunction of the other mode and its time derivative to zero. In other words, if we want to have either only SDWs or only the inertia-buoyancy mode in the disc flow, we should initially set simply $\psi_g(-\infty) = d\psi_g/dt(-\infty) = 0$ or $\psi_s(-\infty) = d\psi_s/dt(-\infty) = 0$, respectively. Modal equations (4.25) governing the dynamics of the SDW and inertia-buoyancy modes are not coupled, implying that in the absence of shear the perturbation modes evolve independently, that is, initially exciting one either mode with a specific characteristic time-scale (frequency) does not lead to the excitation of other modes with different time-scales. In the following sections, we investigate how shear modifies modal equations (4.25) and its influence on the dynamics of the SDW and inertia-buoyancy modes. Specifically, we will show that in the presence of Keplerian shear, these equations are no longer independent from each other, i.e. become coupled that, in turn, results in the linear coupling between these two modes and, in particular, between SDWs and convection.

4.2.4 Effects of shear on mode dynamics

Here we rederive the modal equations from Eqs. (4.19) and (4.20) including shear q. Let us first introduce the non-dimensional variables and quantities being used throughout this chapter

$$t \to \Omega t, \quad \omega_{s,g} \to \frac{\omega_{s,g}}{\Omega}, \quad \kappa^2 \to \frac{\kappa^2}{\Omega^2} = 1, \quad N_0^2 \to \frac{N_0^2}{\Omega^2} = \frac{\gamma - 1}{\gamma},$$

$$K_y \equiv \frac{c_s k_y}{\Omega}, \quad K_z \equiv \frac{c_s k_z}{\Omega}, \quad \bar{K}_z \equiv \sqrt{K_z^2 + \frac{\gamma}{4}}, \quad H \to \frac{\Omega H}{c_s} = \frac{1}{\sqrt{\gamma}},$$

$$s \to \frac{s}{\rho_m c_s^2}, \qquad p \to \frac{p}{\rho_m c_s^2}, \qquad \Psi_{s,g} \equiv \frac{\psi_{s,g}}{\rho_m c_s^2}.$$

A similarly normalised radial wavenumber, K_x, varies with (normalised) time as $K_x(t) = qK_yt$. Also, as defined above, $K_\perp^2(t) = K_x^2(t) + K_y^2 = K_y^2(1 + q^2t^2)$, $K^2(t) = K_\perp^2(t) + \bar{K}_z^2$.

Since $K_x(t)$ and $\kappa^2(t)$ in Eqs. (4.21) and (4.22) are now time-dependent as a result of shear, the non-dimensional frequencies ω_s and ω_g also become functions of time:

$$\omega_s^2(t) = \frac{K^2(t)+\kappa^2(t)}{2}\left(1+\sqrt{1-\frac{4(N_0^2 K_\perp^2(t)+\kappa^2(t)\bar{K}_z^2)}{(K^2(t)+\kappa^2(t))^2}}\right)$$

$$\omega_g^2(t) = \frac{K^2(t)+\kappa^2(t)}{2}\left(1-\sqrt{1-\frac{4(N_0^2 K_\perp^2(t)+\kappa^2(t)\bar{K}_z^2)}{(K^2(t)+\kappa^2(t))^2}}\right).$$

At large times, $K(t) \gg 1$ and from these expressions it follows that $\omega_s^2(t) \simeq K^2(t)$, $\omega_g^2(t) \simeq N_0^2$. As noted above, all the other quantities can readily be expressed through Ψ_s, Ψ_g and their first order time derivatives, so we carry out the further analysis working mostly with the eigenfunctions.

Expressing $\hat{s}, \hat{p}$ through Ψ_s, Ψ_g from Eqs. (4.23) and (4.24), taking into account the above normalisation, and substituting them into Eqs. (4.19) and (4.20), we arrive at the following system of coupled second order differential equations for the eigenfunctions

$$\frac{d^2\Psi_s}{dt^2} + f_s\frac{d\Psi_s}{dt} + (\omega_s^2 + \Delta\omega_s^2)\Psi_s = f_g\frac{d\Psi_g}{dt} + \Delta\omega_g^2\Psi_g, \tag{4.26}$$

$$\frac{d^2\Psi_g}{dt^2} + f_g\frac{d\Psi_g}{dt} + (\omega_g^2 + \Delta\omega_g^2)\Psi_g = f_s\frac{d\Psi_s}{dt} + \Delta\omega_s^2\Psi_s, \tag{4.27}$$

where the new coefficients $f_s, \Delta\omega_s^2$ and $f_g, \Delta\omega_g^2$, compared with Eqs. (4.25), describe modifications to the dynamics of the SDW and inertia-buoyancy modes and their coupling brought about by the shear of disc flow. They depend on time as well as on the azimuthal K_y and vertical $\bar{K}_z$ wavenumbers:

$$f_s(K_y, \bar{K}_z, t) \equiv \frac{2K_\perp(t)}{\omega_s^2(t)-\omega_g^2(t)}\left(\frac{\omega_s^2(t)-\bar{K}_z^2}{K_\perp(t)}\right)',$$

$$f_g(K_y, \bar{K}_z, t) \equiv \frac{2K_\perp(t)}{\omega_g^2(t)-\omega_s^2(t)}\left(\frac{\omega_g^2(t)-\bar{K}_z^2}{K_\perp(t)}\right)',$$

$$\Delta\omega_s^2(K_y, \bar{K}_z, t) \equiv \frac{K_\perp(t)}{\omega_s^2(t)-\omega_g^2(t)}\left(\frac{\omega_s^2(t)-\bar{K}_z^2}{K_\perp(t)}\right)'',$$

$$\Delta\omega_g^2(K_y, \bar{K}_z, t) \equiv \frac{K_\perp(t)}{\omega_g^2(t)-\omega_s^2(t)}\left(\frac{\omega_g^2(t)-\bar{K}_z^2}{K_\perp(t)}\right)''$$

(here and below prime denotes the time derivative). Note that all the time derivatives in these coefficients are proportional to the shear parameter q as well as to the azimuthal wavenumber K_y (time enters through qK_yt) and thus vanish in the shearless limit $q = 0$ or in the case of axisymmetric perturbations with $K_y = 0$, reducing Eqs. (4.26) and (4.27) to decoupled Eqs. (4.25). In vertically stratified Keplerian discs, the dynamics of axisymmetric perturbations (modes) was extensively studied in the past (e.g., Ruden et al., 1988; Kley et al., 1993; Lubow and Pringle, 1993b; Korycansky and Pringle, 1995; Ogilvie, 1998; Mamatsashvili and Rice, 2010), so we omit them from the present analysis and concentrate only on non-axisymmetric perturbations having non-zero azimuthal wavenumber, $K_y \neq 0$, which is assumed to be positive ($K_y > 0$) throughout without loss of generality.

Equations (4.26) and (4.27) describe the linear dynamics of the SDW and inertia-buoyancy modes and their coupling in a vertically stratified, compressible disc with Keplerian differential rotation. As mentioned above, novel features in the mode dynamics, in comparison with the case of no shear, arise because of the coefficients $f_s, \Delta\omega_s^2$ and $f_g, \Delta\omega_g^2$ that originate from shear. The homogeneous (left hand side) parts of Eqs. (4.26) and (4.27) describe the individual dynamical behaviour of the SDW and inertia-buoyancy modes, respectively, in the presence of shear. The most interesting aspect here is the source terms on the right hand side of these equations that ensure dynamical coupling between these two modes in the linear regime. In other words, initially imposed only one mode acts as a source for another mode and can excite it in the course of evolution. Thus, the shear of disc flow introduces a phenomenon of mode conversion, as in chapter 3. Below, we will demonstrate by means of numerical analysis how this mode coupling occurs in practice.

For the purpose of numerical integration presented below, we change the eigenfunctions to $\Psi_s \to \Psi_s/h_s,\ \Psi_g \to \Psi_g/h_g$, where

$$h_s = \exp\left(-\frac{1}{2}\int_0^t f_s dt'\right), \quad h_g = \exp\left(-\frac{1}{2}\int_0^t f_g dt'\right).$$

These new eigenfunctions are more convenient, because substituting them into Eqs. (4.26) and (4.27), we arrive at a simpler system without first derivatives on the left hand side

$$\frac{d^2\Psi_s}{dt^2} + \hat{\omega}_s^2\Psi_s = \chi_{sg1}\frac{d\Psi_g}{dt} + \chi_{sg2}\Psi_g \tag{4.28}$$

$$\frac{d^2\Psi_g}{dt^2} + \hat{\omega}_g^2\Psi_g = \chi_{gs1}\frac{d\Psi_s}{dt} + \chi_{gs2}\Psi_s. \tag{4.29}$$

Equations (4.28) and (4.29) with time-dependent frequencies

$$\hat{\omega}_s^2 = \omega_s^2 + \Delta\omega_s^2 - \frac{f_s'}{2} - \frac{f_s^2}{4},$$

$$\hat{\omega}_g^2 = \omega_g^2 + \Delta\omega_g^2 - \frac{f_g'}{2} - \frac{f_g^2}{4},$$

respectively, for the SDW and inertia-buoyancy modes, resemble equations of two coupled oscillators. These frequencies differ from corresponding ω_s^2 and ω_g^2 due to the presence of the

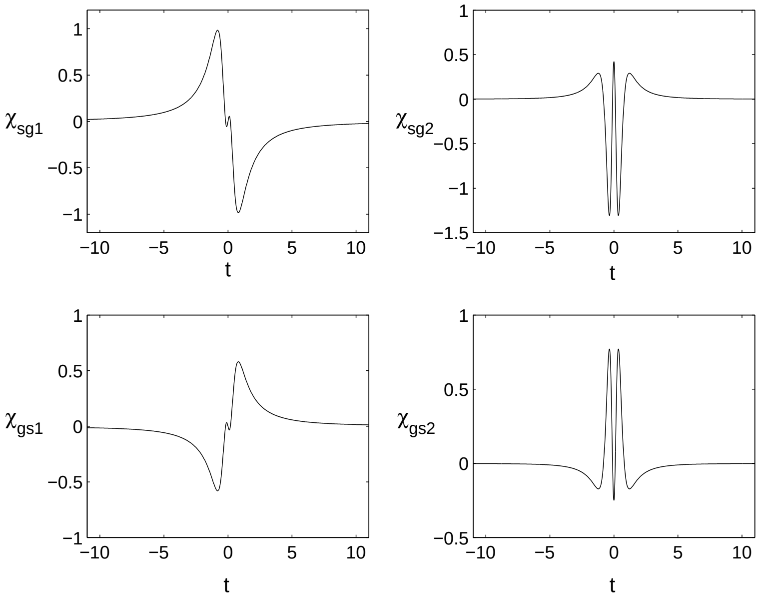

Figure 4.1: Coupling parameters χ as a function of time for $K_y = 2,\ K_z = 3$. They reach the highest values during $|t| \lesssim 1$ and fall off at large times $|t| \gg 1$.

shear-induced terms. For large times as well as for $K_y \gg 1$ and/or $\bar{K}_z \gg 1$, these terms are small and $\hat{\omega}_s^2$ and $\hat{\omega}_g^2$ go to corresponding values in the shearless limit:

$$\hat{\omega}_s^2(t) \simeq K^2(t), \qquad \hat{\omega}_g^2(t) \simeq N_0^2 \frac{K_\perp^2(t)}{K^2(t)} + \kappa^2(t) \frac{\bar{K}_z^2}{K^2(t)}.$$

(Since we use the total vertical wavenumber $\bar{K}_z$ everywhere below, the bar over it will be omitted hereafter, so it should not be confused with the real vertical wavenumber K_z). The χ parameters, describing the coupling between the modes, are given by

$$\chi_{sg1} = \frac{f_g h_g}{h_s}, \qquad \chi_{sg2} = \frac{h_g}{h_s}\left(\Delta\omega_g^2 - \frac{f_g^2}{2}\right),$$

$$\chi_{gs1} = \frac{f_s h_s}{h_g}, \qquad \chi_{gs2} = \frac{h_s}{h_g}\left(\Delta\omega_s^2 - \frac{f_s^2}{2}\right).$$

The coupling parameters χ_{sg1}, χ_{sg2} describe the excitation of the SDW mode by the inertia-buoyancy mode, while χ_{gs1}, χ_{gs2} describe the excitation of the inertia-buoyancy mode by the SDW mode. Figure 4.1 shows the temporal variation of the coupling parameters during the swing of a perturbation SFH from leading to trailing in the case of superadiabatic stratification with our chosen $\gamma = 0.8$. Accordingly, from now on we speak only of the convective mode instead of the more general inertia-buoyancy mode. In this case, $\hat{\omega}_g^2(t)$ has a negative sign for most of the time and determines the shear-modified instantaneous growth rate of convective instability. It is seen from this figure that for fixed K_y and K_z, these parameters reach their maximum values in the interval $|t| \lesssim 1$, when the time-dependent radial wavenumber is not

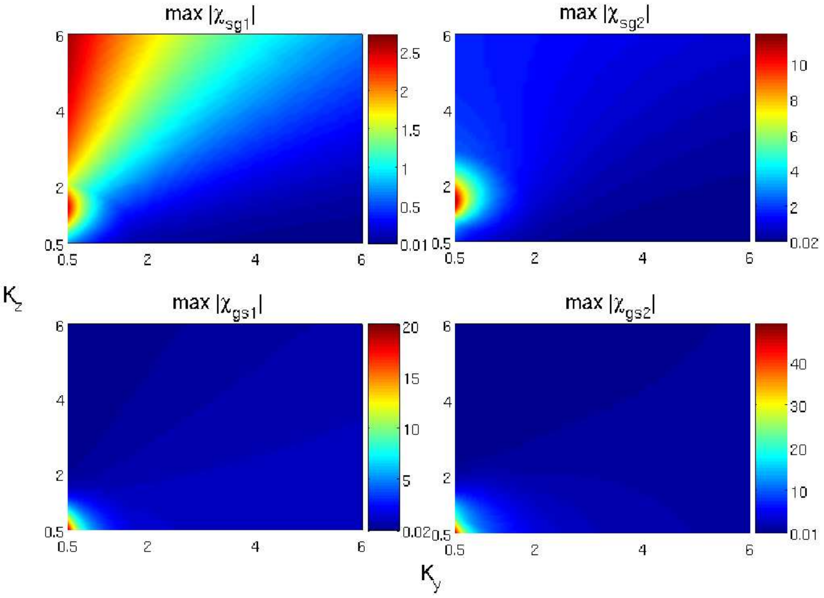

Figure 4.2: Maximum values (with respect to time) of modulus of coupling parameters, $|\chi|$, as a function of K_y and K_z. $\max|\chi_{sg1}|$ and $\max|\chi_{sg2}|$ are appreciable over a broader range of wavenumbers than $\max|\chi_{gs1}|$ and $\max|\chi_{gs2}|$.

large, $|K_x(t)/K_y| = q|t| \lesssim 1$, and rapidly decay at $|t| \gg 1$, when $|K_x(t)/K_y| = q|t| \gg 1$ as well. Thus, because the coupling parameters χ evidently also originate from shear, we can conclude that at given K_y and K_z, the influence of shear on the mode dynamics can be significant only for moderate radial wavenumbers and an efficient energy exchange between the modes and the mean shear flow should be expected to occur just during this interval, which we will call a coupling interval. For $|K_x(t)/K_y| \gg 1$, the coupling parameters are small and hence the modes become dynamically decoupled from each other and evolve independently (see section 4.3). In Fig. 4.2, the maximum values (over the interval $|K_x(t)/K_y| = q|t| \lesssim 1$) of the modulus of the coupling parameters are plotted in the (K_y, K_z)-plane. These maximum values behave differently. The maximum of $|\chi_{sg1}|$ is appreciable for a broader range of wavenumbers, but it has larger values at smaller K_y. The maximum of $|\chi_{sg2}|$ is appreciable for K_z of the order of unity and smaller K_y, and rapidly decays with the increase of these wavenumbers. On the other hand, the maximum values of both $|\chi_{gs1}|$ and $|\chi_{gs2}|$ are only appreciable at much smaller K_y and K_z. Because the coupling parameters are generally one of the factors determining the generation of one mode by another, we should expect the efficiency of coupling of the SDW and convective modes as a function of K_y and K_z to follow, to a certain extent, the same trend as that of the coupling parameters shown in Fig. 4.2. In particular, this implies that in a superadiabatically stratified disc, the generation (triggering) of convective motions by SDWs, which is determined by the coupling parameters χ_{gs1}, χ_{gs2}, ought to be a minor process occurring in a narrower range of wavenumbers in comparison with the opposite one – excitation of SDWs by convection determined by χ_{sg1}, χ_{sg2}. Besides, convection can be easily triggered by other mechanisms in a superadiabatically stratified disc. So, below we focus only on the latter process.

4.3 SDW and convective modes – shear-induced coupling

In this section, we study the shear-induced dynamics of the non-axisymmetric SDW and convective modes by numerically solving Eqs. (4.28) and (4.29). At large times/radial wavenumbers ($|K_x(t)/K_y| = q|t| \gg 1$), the adiabatic (WKBJ) condition with respect to time is satisfied for the frequency (growth rate) of the modes

$$|\hat{\omega}'_{s,g}(t)| \ll |\hat{\omega}^2_{s,g}(t)|.$$

A physical interpretation of this inequality is that the time-scale of shear-induced variation of the frequency/growth rate is much larger than the characteristic time-scale (i.e., inverse frequency/growth rate) of the mode itself. In this adiabatic regime, $\hat{\omega}^2_g(t) \simeq N_0^2 < 0$ is nearly independent of time and is much less than the frequency of SDWs, which increases linearly with time, $\hat{\omega}_s(t) \simeq K(t) \approx qK_y|t| \gg 1$, so that the mode time-scales are well separated. In addition, all four coupling parameters χ are negligible at these times (Fig. 4.1). All these imply that in the adiabatic regime, shear plays only a minor role in the dynamics of SDWs and convective instability. Below we will refer to this asymptotic stage as adiabatic/non-coupling region in the **K**-space. Consequently, if convective mode is in the adiabatic region, it will evolve with time as in the shearless limit, without any exchange of energy and exciting other modes (i.e., SDWs) in the disc. However, if $K_x(t)$ during its drift crosses the coupling interval $|K_x(t)/K_y| \lesssim 1$, where, as we have seen, the influence of shear on the mode dynamics becomes significant, the excitation of SDWs by non-axisymmetric convective motions can take place. So, the goal of this section is to examine this phenomenon in detail.

For further use we also introduce the mode energies as

$$E_s \equiv \Psi'^2_s + \hat{\omega}^2_s(t)\Psi^2_s, \qquad E_g \equiv \Psi'^2_g + \hat{\omega}^2_g(t)\Psi^2_g$$

for the SDW and convective modes, respectively. In the absence of shear they are conserved quantities, as it follows from Eqs. (4.25). The mode energies are useful in characterising mode dynamics for asymptotically large times, in the adiabatic regime. In this regime, the energies vary with time due solely to the mean shear flow. In particular, the SDW mode energy grows linearly with time,

$$E_s \propto \hat{\omega}_s(t) \simeq K(t) \propto |t|,$$

which follows from Eq. (4.28) in the WKBJ regime neglecting the coupling parameters, which are small in this regime anyway. Such an asymptotic behaviour of the SDW mode energy will later prove to be a useful diagnostic in analysing the generation of SDWs.

4.3.1 Generation of SDW mode by convective mode

We wish to investigate the convective instability in the presence of Keplerian shear and, specifically, how convective motions can excite SDWs. To this end, initially at $t = -t_0$, where $t_0 \gg 1$ is some large positive parameter, we impose a tightly leading (i.e., with $K_x(-t_0)/K_y \ll -1$) SFH of the convective mode on the flow without any mixture of a SDW mode SFH and follow the subsequent evolution of the eigenfunctions and perturbed quantities until $t = t_0$, at which stage their SFHs become tightly trailing (with $K_x(t_0)/K_y \gg 1$). For numerical integration we use a standard Runge-Kutta scheme (MATLAB ode45 RK implementation).

To prepare such initial conditions, we make use of the fact that at $t \ll -1$ the dynamics is adiabatic and the modes do not interact with each other that, in turn, permits us to pick out only the convective mode at the beginning. In this adiabatic/non-coupling regime, the convective mode is given by the WKBJ asymptotic solution of the homogeneous part of Eq. (4.29), which we take to have the form

$$\Psi_g = \frac{C_0}{\sqrt{|\hat{\omega}_g(t)|}} \exp\left(\int_{-t_0}^{t} |\hat{\omega}_g(t')| dt'\right) \quad at \quad t \ll -1, \tag{4.30}$$

where C_0 is some arbitrary constant setting the initial value of Ψ_g at the start of integration at $t = -t_0$. This solution means that the convective mode is evanescent in the distant past (at $t \to -\infty$) and grows with time afterwards. As for the initial value of the SDW mode eigenfunction, a high-frequency SDW component is absent at the outset, so that Ψ_s should be given only by non-oscillatory particular solution $\Psi_s^{(\mathrm{g})}$ of Eq. (4.28) due to the source term on the right hand side that is associated with the convective mode. Asymptotically at $t \ll -1$, we can ignore the second order time derivative in Eq. (4.28) and represent the initial value of Ψ_s at these times, to a good approximation, as

$$\Psi_s = \Psi_s^{(\mathrm{g})} \approx \frac{\chi_{sg1}\Psi'_g + \chi_{sg2}\Psi_g}{\hat{\omega}_s^2(t)}. \tag{4.31}$$

Indeed, as mentioned above, in the adiabatic regime, (i) the period of the SDW oscillations is much smaller than the shear/dynamical time, or equivalently $|\hat{\omega}_s'| \ll \hat{\omega}_s^2$, (ii) the coupling parameters vary a little during the oscillation period and at the same time (iii) the frequency of SDWs is much larger than the characteristic growth rate of the convective mode, $|\hat{\omega}_g| \ll \hat{\omega}_s$. As a result, the second time derivative of the particular solution (4.31) turns out to be much smaller than the right hand side term of Eq. (4.28), thereby validating our approximation. This slowly varying solution ensures that there is no contribution from the oscillatory SDW mode in the initial conditions. Note also that as the coupling coefficients are vanishing at $t \to -\infty$, we also have $\Psi_s^{(\mathrm{g})}(-\infty) \to 0$, $\Psi_s^{(\mathrm{g})\prime}(-\infty) \to 0$. This in fact coincides with the condition of the absence of SDWs in the shearless limit. Thus, asymptotic solutions (4.30) and (4.31) can be viewed as a generalisation of the recipe for imposing only the convective mode, without a mixture of SDWs, on the flow in the absence of shear (i.e., $\Psi_s(-\infty) = \Psi_s'(-\infty) = 0$) to the case of non-zero shear provided, however, that the adiabatic approximation is still met, which is always the case at large times.

Figure 4.3 shows the subsequent temporal evolution corresponding to initial conditions/solutions (4.30) and (4.31). Together with the eigenfunctions, which more clearly illustrate mode coupling, we also plot the time-development of the perturbed entropy s, pressure p and mode energies E_s and E_g. At the beginning, while the convective mode SFH is still in the adiabatic region $K_x(t)/K_y = qt \ll -1$ and far from the coupling region, the eigenfunction Ψ_g, following expression (4.30), grows exponentially with the instantaneous growth rate $|\hat{\omega}_g|$ and induces a similar behaviour in the entropy, whereas Ψ_s, evolving according to (4.31), still remains very small and non-oscillatory. The evolution of the corresponding energy E_s is similar to that of Ψ_s. It is also seen in Fig. 4.3 that at this initial asymptotic/adiabatic stage of evolution, the entropy perturbation is dominant over the pressure perturbation, which, like Ψ_s, is small and aperiodic initially. It can also be shown that the form of our initial conditions for the convective mode implies $E_g = 0$ at $t \ll -1$, because of the negative $\hat{\omega}_g^2 < 0$ (of course, this does not mean that the convective mode is zero).

As time passes, the time-dependent radial wavenumber $K_x(t)$ of the convective mode SFH gradually approaches the coupling region $|K_x(t)/K_y| \lesssim 1$ and begins to cross it. In this region, the effects of shear come into play: the coupling parameters become appreciable (Fig. 4.1) and at the same time the characteristic time-scales of the SDW mode, convective mode and the dynamical/shear time $1/q$ (in the non-dimensional form) become comparable,[4] or

---

[4] This is not the case at large K_y and/or K_z, because the mode time-scales still remain separated even for $|K_x(t)/K_y| \lesssim 1$ and therefore the mode coupling is negligible at large wavenumbers; see section 4.3.2.

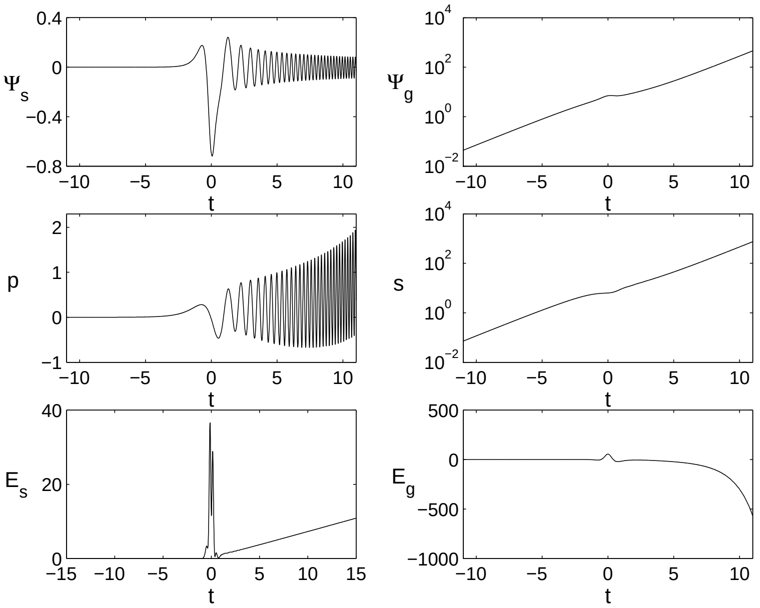

Figure 4.3: Evolution of the eigenfunctions Ψ_s, Ψ_g, perturbed pressure p, entropy s and corresponding energies E_s, E_g pertaining to an initially imposed purely convective mode SFH with $K_y = 2, K_z = 3$. At the beginning, Ψ_g and s evolve adiabatically, exponentially growing with the characteristic growth rate $|\hat{\omega}_g|$, while Ψ_s and p are still nearly zero. At around $t = 0$, rapid oscillations abruptly emerge in the evolution of Ψ_s and p that indicate a trailing SFH of the SDW mode being excited. In addition, the convective mode causes p to grow as well while oscillating. Accordingly, the energy of the SDW mode, E_s, being negligible at $t < 0$, after the transient amplification event in the vicinity of $t = 0$, increases linearly with time in the subsequent adiabatic regime at $t \gg 1$.

equivalently $q \sim |\hat{\omega}_g| \sim \hat{\omega}_s$, also implying that in the coupling region, the adiabatic condition for the SDW frequency breaks down, $|\hat{\omega}_s'| \sim |\hat{\omega}_s^2|$, that is, the coupling region is also non-adiabatic. This results in an efficient energy exchange between these two modes and the background shear flow in this coupling region. First, at $-1 \lesssim t < 0$, the eigenfunction Ψ_s, pressure p and the energies E_s, E_g undergo transient amplification by extracting energy mainly from the background flow, but still remain aperiodic. The evolution of Ψ_s has now deviated from (4.31), because the second derivative of Ψ_s has become important in Eq. (4.28), though it is still following the non-oscillatory solution $\Psi_s^{(g)}$. Then, during a short period of time around $t = 0$, when $K_x(t)$ crosses the point $K_x = 0$, swinging from negative (leading) to positive (trailing), rapid oscillations abruptly appear in the evolution of Ψ_s and the pressure, signaling the generation of a trailing SFH of the SDW mode. So, now at $t > 0$, there are the newly excited trailing SDW mode SFH and the former convective mode SFH. It can be said that the convective mode, in some sense, acts as a mediator between SDWs and the disc shear flow. The energy needed for the SDW excitation is mainly extracted from the shear with the help of the convective mode. After that, on leaving the coupling region, $K_x(t)$ moves into the next adiabatic region $K_x(t)/K_y \gg 1$, where the linear dynamics of SDWs and convection are decoupled with no further energy exchange. The time-scales of these two modes get separated again: the pressure p rapidly oscillates as a result of the new SDW component in Ψ_s with the frequency $\hat{\omega}_s(t)$, which increases linearly with time and soon becomes much larger than the growth rate, $|\hat{\omega}_g|$, of convective instability. The eigenfunction Ψ_g continues to grow exponentially with this growth rate and causes a similar growth in the entropy. As typical of SDWs in the adiabatic regime, the corresponding energy E_s increases linearly with time solely due to the background flow.

Finally, we would like to stress again that the studied here linear coupling of the SDW and convective modes caused by shear is essentially of the same nature as other linear mode coupling phenomena in disc flows with Keplerian shear: coupling of vortical and wave modes, that is, the generation of SDWs by vortices studied in chapter 3 (see also Bodo et al., 2005; Heinemann and Papaloizou, 2009a), coupling of vortices and inertia-gravity waves (Tevzadze et al., 2003, 2008) and coupling of baroclinic and SDW modes (Tevzadze et al., 2010). Thus, the linear mode coupling is a generic phenomenon inevitably taking place whenever the mean flow is inhomogeneous (i.e., the velocity profile has a non-zero shear) and all these cases are its special manifestations for particular physical settings. In this connection, we should mention that it also occurs in MHD shear flows, where in general there are larger number of mode branches with different time-scales and coupling among them becomes quite complex (see e.g., Chagelishvili et al., 1996). For example, Heinemann and Papaloizou (2009b) described the generation of SDWs by the vortical perturbations in a MRI-turbulent disc shear flow. The coupling among different MHD modes can also be seen in the linear analysis of non-axisymmetric MRI by Balbus and Hawley (1992), however, the authors did not identify it as such.

Thus, the primary effect of shear of the disc's differential rotation on the perturbation dynamics is that it couples, or introduces a new channel of energy exchange among different types of perturbation modes existing in the disc as well as with the disc flow. In the present case, the linear dynamics of the convective mode is accompanied by the generation of high-frequency SDWs during a finite-time interval as SFHs swing from leading to trailing. As noted before, the Kelvin mode, or shearing wave approach because not involving spectral expansion in time, allows us to trace an entire temporal evolution of perturbations and thus to reveal new aspects of non-axisymmetric convection in disc shear flows. This, in turn, can provide further insight into the role of convection and its coupling with SDWs in angular momentum transport, especially now when the capability of convection to transport angular momentum outwards in discs has recently been established via numerical simulations (we address the transport properties of the modes in section 4.4). Previous linear studies of non-axisymmetric convection were either in the framework of modal approach (e.g., Lin et al., 1993) or using the shearing wave (non-modal) approach (RG92; Korycansky, 1992; Brandenburg and Dintrans, 2006) as here, but without identifying the mode coupling process. As pointed out in section 2.4, the results of the modal approach are actually applicable for asymptotically large times a and therefore tend to overlook dynamical effects arising as a result of the flow non-normality/shear. Since the excitation of SDWs by convection is just one such examples, occurring during a limited time interval, in fact it cannot be captured in the framework of the modal approach.

4.3.2 Generation of SDWs for various K_y and K_z

Here we examine how the efficiency of the above-described SDW generation process depends on the azimuthal, K_y, and vertical, K_z, wavenumbers. Figure 4.4 shows the evolution of Ψ_s under the same initial conditions (4.30) and (4.31) as above, that is, consisting of an initially imposed only tightly leading convective mode SFH at various K_y and K_z, without a mixture of the SDW mode SFH. For large K_y and/or K_z (i.e., when the azimuthal and/or vertical wavelengths of the modes are much smaller than the disc scale height), starting out with a very small value, Ψ_s mainly undergoes only small transient variation in the coupling region (at $|t| \lesssim 1$) followed by very weak, almost no generation of high-frequency oscillations, i.e., SDW component. The reason for such a behaviour is the following. At large K_y and/or K_z, the time-scales of the SDW and convective modes remain well separated during an entire course of evolution, not only at large times, and the adiabatic condition for the SDW frequency holds. As a result, non-oscillatory particular solution (4.31), valid for the initial adiabatic stage of evolution at $t \ll -1$, in fact continues to hold at all times to a good approximation (because its second time derivative always remains negligible in Eq. 4.28) and, hence, an efficient coupling between the SDW and convective modes is not feasible at large K_y and/or K_z. So, in Fig. 4.4, the time-development of Ψ_s at these wavenumbers is given by this particular solution that does not involve any oscillations. (To be more precise, the SDW mode is still being generated, but with a very small amplitude, as evident from the panels with $K_y = 5, K_z = 5$; $K_y = 10, K_z = 5$; $K_y = 10, K_z = 1$; $K_y = 10, K_z = 0.6$ of Fig. 4.4 and also from Fig. 4.5.) Thus, the regime of large azimuthal and/or vertical wavenumbers is a weak coupling regime, where shear plays only a minor role in the mode dynamics. By

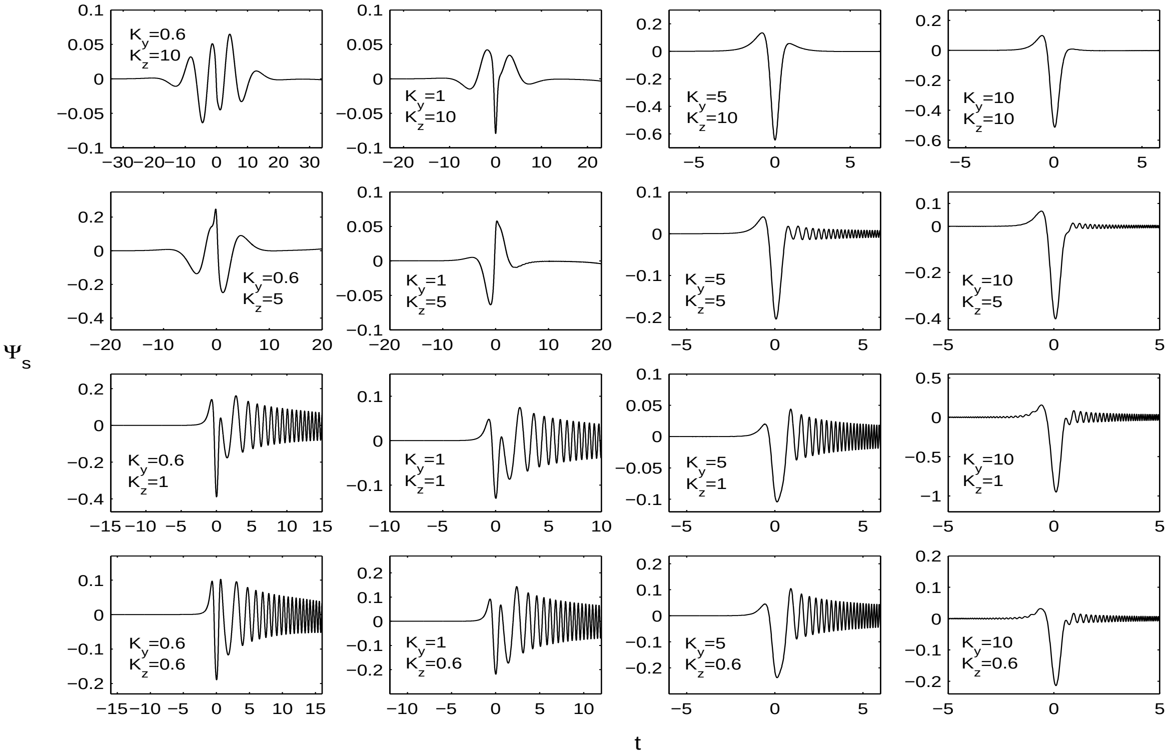

Figure 4.4: Evolution of Ψ_s corresponding to an initially imposed convective mode SFH with various K_y and K_z. Appreciable oscillations in the time-development of Ψ_s appear and, therefore an efficient SDW mode excitation occurs, at $K_y, K_z \lesssim 1$. By contrast, at large K_y and/or K_z, Ψ_s undergoes only small transient variation (growth) in the interval $|t| \lesssim 1$ with very weak, almost no SDW generation.

contrast, for smaller $K_y, K_z \lesssim 1$ (i.e., for wavelengths comparable to the disc scale height) the favourable conditions for mode coupling – violation of the adiabatic condition and comparable time-scales of the SDW and convective modes and the shear time – can occur in the vicinity of $t = 0$, leading to an efficient generation of the SDW mode by the convective one, as described in section 4.3.1.

We can quantitatively characterise the mode coupling as follows. At $t > 0$, the total solution for Ψ_s is a sum of the two components

$$\Psi_s = \Psi_s^{(\mathrm{g})} + \Psi_s^{(\mathrm{w})},$$

where $\Psi_s^{(\mathrm{g})}$, as before, is the non-oscillatory, slowly varying particular solution of Eq. (4.28) associated with the convective mode and $\Psi_s^{(\mathrm{w})}$ is an oscillatory component associated with the SDW mode generated by this particular solution. So, the relative intensity of the wave generation process can be quantified by comparing the value of $\Psi_s^{(\mathrm{w})}$ to that of $\Psi_s^{(\mathrm{g})}$. To do this, we introduce the parameter

$$\epsilon \equiv \frac{\max_{t>0} |\Psi_s^{(\mathrm{w})}(t)|}{\max_{t>0} |\Psi_s^{(\mathrm{g})}(t)|},$$

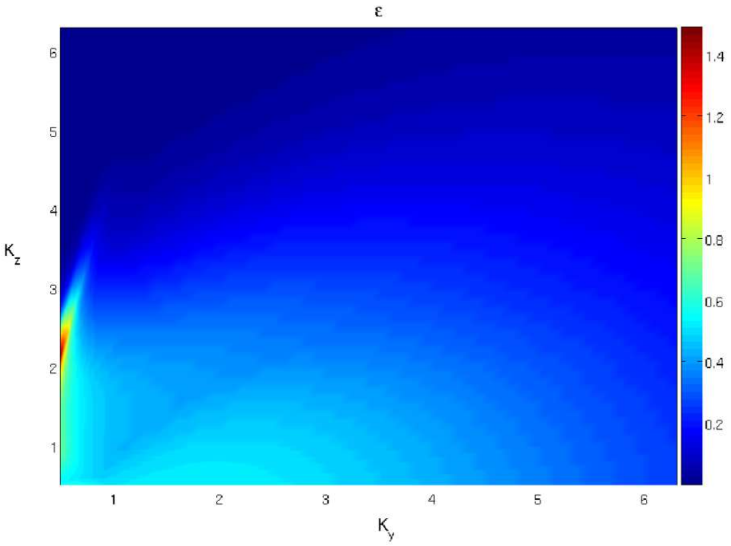

Figure 4.5: Ratio, ϵ, of the maximum values of the SDW mode solution to its generator non-oscillatory particular solution of Eq. (4.28) originating from the convective mode as a function of K_y and K_z.

which is the ratio of the maximal values of $|\Psi_s^{(\mathrm{w})}|$ and $|\Psi_s^{(\mathrm{g})}|$ over time (typically, both these functions achieve their maximal values in the vicinity of $t = 0$, as seen in Fig. 4.4). Obviously, with our initial conditions (4.30) and (4.31), this ratio does not depend on the arbitrary parameter C_0 and is a function of K_y and K_z only, which is plotted in Fig. 4.5. It is seen from this figure that at large K_y and/or K_z, as expected, $\epsilon \lesssim 0.1$ is small, implying that the non-oscillatory particular solution dominates over the oscillatory SDW mode solution, i.e., the wave generation is weak. For $K_y, K_z \lesssim 1$, $\epsilon \sim 1$ implying that the wave component is comparable to the non-oscillatory solution and therefore the SDW generation is appreciable. This also confirms the situation in Fig. 4.4.

4.3.3 Amplitudes of generated SDWs

We can go further and calculate the amplitudes of SDWs generated by the convective mode, which also represent a measure of the coupling efficiency between these two modes. In this section, we choose the constant C_0 in the WKBJ solution (4.30), describing an initially tightly leading convective mode SFH at $t \ll -1$, such that to have

$$\Psi_g = \frac{C_0}{\sqrt{|\hat{\omega}_g(t)|}} \exp\left(\int_{-t_0}^{t} |\hat{\omega}_g(t')| dt'\right) = \frac{1}{\sqrt{|N_0|}} \mathrm{e}^{|N_0|t}.$$

An advantage of this form is that the calculated below SDW amplitudes depend only on K_y and K_z and are independent of the choice of the starting point $-t_0$ (as long as $t_0 \gg 1$).

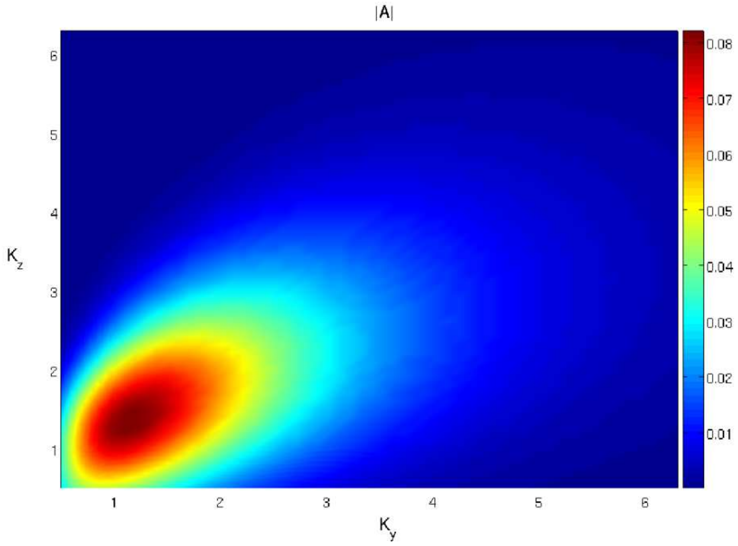

Figure 4.6: Absolute value of the amplitude $|A|$ of SDWs generated by the convective mode as a function of K_y and K_z. It achieves the largest value $|A|_{max} = 0.082$ at $K_{y,m} = 1.2, K_{z,m} = 1.4$ and decreases at smaller and larger K_y, K_z, implying also the decrease in the mode coupling efficiency at these wavenumbers.

In this initial adiabatic regime, as before, $\Psi_s = \Psi_s^{(\mathrm{g})}$, because SDWs are not present at the outset. Then, in the non-adiabatic region at $|t| \lesssim 1$, the mode coupling is at work and subsequently, at around $t = 0$, the SDW mode SFH abruptly emerges. After crossing the non-adiabatic/coupling region, in the next adiabatic region at $t \gg 1$, there are the former convective mode SFH and the excited by it SDW mode SFH, both with tightly trailing orientation. So, now, as noted above, the full solution for Ψ_s is the sum of the following parts:

$$\Psi_s = \Psi_s^{(\mathrm{g})} + \Psi_s^{(\mathrm{w})} = \frac{\chi_{sg1}\Psi_g' + \chi_{sg2}\Psi_g}{\hat{\omega}_s^2(t)} + \\ + \frac{A}{\sqrt{\hat{\omega}_s(t)}}\exp\left(-\mathrm{i}\int_0^t \hat{\omega}_s(t')dt'\right) + \frac{A^*}{\sqrt{\hat{\omega}_s(t)}}\exp\left(\mathrm{i}\int_0^t \hat{\omega}_s(t')dt'\right) \qquad at \qquad t \gg 1,$$

where the first term is the familiar non-oscillatory slowly varying particular solution due to the convective mode and the other two terms are oscillatory WKBJ solutions of the homogeneous part of Eq. (4.28) that correspond to the newly excited SDWs. Since the initial conditions are real, these solutions come in complex conjugate pairs with different signs of frequency and with amplitudes A and its complex conjugate A^*. Thus, the SDW mode SFH generated by the convective mode SFH is in fact a superposition of two SFHs corresponding to SDWs propagating in the opposite directions. In other words, convection always pairwise excites SDWs that propagate oppositely, similar to the situation with SDWs generated by the vortical mode in 2D discs considered in section 3.3.2.

Figure 4.6 shows the dependence of the absolute value of the amplitude, $|A|$, on K_y and K_z. When either of these wavenumbers is large, $|A|$ is small, because, as shown above, the SDW generation/mode coupling inefficient. With decreasing K_y and K_z, as expected, the amplitude increases and attains the largest values at $K_y, K_z \lesssim 1$, because an appreciable SDW generation takes place at such wavenumbers, as also seen in Figs. 4.4 and 4.5. Note that $K_y, K_z \lesssim 1$ is the region where the coupling parameters χ_{sg1}, χ_{sg2} are appreciable too (see Fig. 4.2). Comparing Figs. 4.5 and 4.6 we also see that the amplitude $|A|$ and ϵ, which is a measure of the mode coupling efficiency, behave with K_y and K_z more or less similarly.

4.4 Angular momentum transport by the convective and SDW modes

In this section, we analyse the angular momentum transport properties of the SDW and convective modes. Specifically, we are interested in the radial component of angular momentum flux carried by these two modes, because it determines the radial transport of mass and hence accretion onto the central star. The angular momentum conservation law for linear perturbations in the shearing box model follows from the invariance of the system under translations in the azimuthal y-direction. It can be derived by application of Noether's theorem to the second order Lagrangian density (RG2; Narayan et al., 1987; Heinemann and Papaloizou, 2009a)

$$L = \frac{\rho_0}{2}\left\{\left(\frac{D\boldsymbol{\xi}}{Dt}\right)^2 - \frac{p_0}{\rho_0}\left[(\gamma-1)(\nabla\cdot\boldsymbol{\xi})^2 + \frac{\partial\xi_i}{\partial x_j}\frac{\partial\xi_j}{\partial x_i}\right] + 2(\boldsymbol{\Omega}\times\boldsymbol{\xi})\cdot\frac{D\boldsymbol{\xi}}{Dt} + 2q\Omega^2\xi_x^2\right\},$$

in this direction and yields

$$\frac{D}{Dt}\left(\frac{\partial L}{\partial(D\xi_i/Dt)}\frac{\partial\xi_i}{\partial y}\right) + \frac{\partial}{\partial x_j}\left(\frac{\partial L}{\partial(\partial\xi_i/\partial x_j)}\frac{\partial\xi_i}{\partial y} - L\delta_{j2}\right) = 0,$$

where $\boldsymbol{\xi}$ is the displacement vector related to the perturbed velocity by

$$u_x = \frac{D\xi_x}{Dt}, \quad u_y = \frac{D\xi_y}{Dt} + q\Omega\xi_x, \quad u_z = \frac{D\xi_z}{Dt},$$

and δ_{ij} is the Kronecker delta and the summation is assumed over the repeated indices $i, j = 1, 2, 3$ corresponding to $(x_1, x_2, x_3) \equiv (x, y, z)$. (It can also be readily shown that this Lagrangian density through variation gives the original set of linear perturbation Eqs. 4.8-4.12). Since we are mainly concerned with the radial x-component of the angular momentum flux, we can average this equation over both y- and z-coordinates. After averaging, the canonical angular momentum density in brackets in the first term, as demonstrated by Narayan et al. (1987), actually coincides with the (radial) density of the true physical angular momentum, which in the local rotating frame has the from

$$B \equiv r_0\rho_0\left(\frac{d\xi_y}{dt} + 2\Omega\xi_x\right),$$

but taken with the minus sign and divided by a fiducial radius r_0, which does not play any role in the local analysis, so that we have

$$B = -\left\langle \frac{\partial L}{\partial(D\xi_i/Dt)}\frac{\partial\xi_i}{\partial y}\right\rangle_{yz},$$

where the angle brackets denote averaging over y and z. On similarly averaging the second term and taking it with the minus sign too, we obtain the desired radial component of the angular momentum flux

$$F_x = -\left\langle \frac{\partial L}{\partial(\partial\xi_i/\partial x)}\frac{\partial\xi_i}{\partial y}\right\rangle_{yz} = -\left\langle p\frac{\partial\xi_x}{\partial y}\right\rangle_{yz},$$

where p is the pressure perturbation as used in original Eqs. (4.8-4.12) before normalisation, which is related to the displacement vector by $p = -\rho_0 c_s^2\nabla\cdot\boldsymbol{\xi} + \rho_0 g\xi_z$. So, the angular momentum conservation takes more compact form

$$\frac{\partial B}{\partial t} + \frac{\partial F_x}{\partial x} = 0.$$

Because we work in terms of $\mathbf{V} = \rho_0\mathbf{u}$, it is convenient to make a change $\boldsymbol{\xi} \rightarrow \rho_0\boldsymbol{\xi}$ and normalise the new displacement and flux as $\xi_x \rightarrow \Omega\xi_x/\rho_m c_s,\ F_x \rightarrow F_x/\rho_m c_s^2$. After that, for the angular momentum flux carried by an individual SFH we get

$$F_x = \frac{1}{2}K_y p(t)\xi_x(t),$$

where now $p(t)$ and $\xi_x(t)$ are the amplitudes of SFHs depending only on time, as used in the previous sections, and $V_x(t) = d\xi_x(t)/dt$. Since, we have two modes – SDWs and convection, we decompose p and ξ_x into two parts corresponding to these modes

$$p = p_s + p_g, \quad \xi_x = \xi_{x,s} + \xi_{x,g},$$

where p_s and $\xi_{x,s}$ are related to the SDW component and, therefore, are rapidly oscillating in time with the frequency of the latter, while p_g and $\xi_{x,g}$ are related to the convective mode and vary on its corresponding time-scale, relatively slowly compared with the SDW mode. So, the total flux

$$F_x = \frac{1}{2}K_y(p_s + p_g)(\xi_{x,s} + \xi_{x,g}),$$

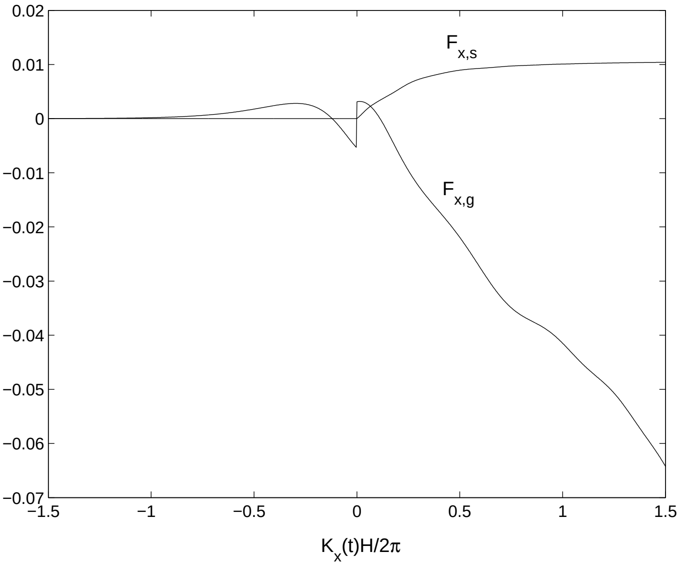

Figure 4.7: Radial angular momentum fluxes associated with the convective mode, $F_{x,g}$, and with the generated by it the SDW mode, $F_{x,s}$, for $K_{y,m} = 1.2, K_{z,m} = 1.4$. The jump of $F_{x,g}$ at $t = 0$ is due to the emergence of the SDW mode at around this time.

will also be of oscillatory type, because of the presence of the SDW mode, but we can conveniently smooth it by averaging over the oscillation period much smaller than the growth time of the convective mode (we denote the time averaging with angle brackets without subscript). As a result, we get

$$\langle F_x \rangle = \frac{1}{2} K_y \langle p_s \xi_{x,s} \rangle + \frac{1}{2} K_y p_g \xi_{x,g}. \tag{4.32}$$

All the cross-products vanish after averaging and the total angular momentum flux appears as a sum of angular momentum fluxes

$$F_{x,s} = \frac{1}{2} K_y \langle p_s \xi_{x,s} \rangle , \quad F_{x,g} = \frac{1}{2} K_y p_g \xi_{x,g}$$

related, respectively, to the SDW and convective modes (we have omitted the angle brackets in $F_{x,g}$, as it does not change much during the oscillation period). In an analogous calculation of the angular momentum transport by non-axisymmetric shearing waves, RG92 obtained rapidly oscillating, on average negative, angular momentum flux and attributed it only to the convective mode without separating/analysing the contribution from SDWs. In fact, this contribution is present, because, as demonstrated here, SDWs are inevitably generated by the convective mode due to background shear. So, decomposition (4.32) allows us to individually analyse the angular momentum fluxes carried by each mode and contrast them.

In Fig. 4.7, we show the time-development of the fluxes $F_{x,s}$ and $F_{x,g}$ when a tightly leading purely convective mode is imposed initially. At $t < 0$, before SDW mode generation, the angular momentum transport is solely due to the convective mode SFH and is still small. Then, in the vicinity of $t = 0$, trailing SDW mode SFH emerges, giving rise to its corresponding flux $F_{x,s}$ at $t > 0$, which initially increases and after a short time settles to

$$F_{x,s} \simeq \frac{K_y K_x(t)(\omega_s^2(t) - K_z^2)(\omega_s^2(t) - N_0^2)}{K_\perp^2(t)(K_z^2 - N_0^2)^2}\frac{h_s^2(t)}{\hat{\omega}_s(t)}|A|^2,$$

where A is the amplitude of generated SDWs as defined in section 4.3.3. With time, this expression tends to a positive and constant value, as clearly seen in Fig. 4.7. Thus, trailing SDWs always transport angular momentum outwards. A similar result was also obtained by Heinemann and Papaloizou (2009a) for SDWs generated by the vortical mode. As for $F_{x,g}$, during a short period of time after $t = 0$, it is still positive, but quickly switches to increasingly negative due to the exponential growth of Ψ_g, and dominates over the positive flux of the SDW mode. This is also evident from the asymptotic expression of the flux of the convective mode

$$F_{x,g} \simeq \frac{K_y K_z^2(\omega_g^2(t) - K_z^2)(\kappa^2(t) - N_0^2)}{2K_\perp(t)K^2(t)(K_z^2 - N_0^2)^2}h_g^2(t)\Psi_g^2(t)$$

at $t \gg 1$, which is always negative, as $\omega_g^2(t) \simeq N_0^2 < 0, \kappa^2(t) \simeq 2(2-q) = 1$ at these times. However, as demonstrated by Lesur and Ogilvie (2010) and Käpylä et al. (2010), this flux of angular momentum associated with the convective mode, being negative in the linear regime, can actually reverse sign in the non-linear regime. We also see from Fig. 4.7 that shortly after the SDW mode generation, at $K_x(t)H/2\pi \lesssim 1$ (i.e., when the radial wavelength is still larger than or equal to the scale height), the corresponding wave flux, $F_{x,s}$, achieves a maximum value comparable to, though still less by an absolute value, than the convective mode flux, $F_{x,g}$. As Heinemann and Papaloizou (2009b) showed, just these values of the angular momentum flux due to SDWs during the time when their time-dependent $K_x(t)$ are still in the range $K_x(t)H/2\pi \lesssim 1$, are important and determine transport in the non-linear regime, because SDWs with radial wavenumbers larger than this are subsequently damped via shock formation in the trailing phase (see also Fig. 3.10).[5] Thus, SDWs generated by convection are, in principle, capable of fully contributing to the transport before undergoing significant damping. Based on this, in the non-linear regime, we may expect that the positive flux due to SDWs will aid and enhance the outward angular momentum transport due to the convective mode only, although a further non-linear study is required to ascertain this.

---

[5] But overall, SDW perturbations and associated with them transport are not damped, as successive leading SFHs are regenerated due to non-linearity, maintaining a steady flux

4.5 Summary and discussion

In this chapter, we have investigated a new linear mechanism of spiral density wave excitation by vertical thermal convection in compressible Keplerian discs with superadiabatic vertical stratification using the local shearing box approximation. The wave excitation results from the shear of the disc's Keplerian differential rotation. As is usually done in the shearing box, perturbations were decomposed into spatial Fourier harmonics, or shearing plane waves. The temporal evolution of the corresponding amplitudes was followed by numerical integration of the linearised equations. Only non-axisymmetric perturbations were considered for which the effects of shear are important. Three basic types of perturbation modes can be distinguished in this system: SDWs due mainly to compressibility, the convective mode due mainly to the negative vertical entropy gradient, and the vortical mode due to the combined action of stratification and Coriolis force. However, in the present case of non-self-gravitating discs with a superadiabatic vertical structure, the main energy-carrying mode is convection. In such a setup, the algebraic growth of the vortical mode can be neglected compared with the exponential amplification of the convective mode and hence we concentrated primarily on the dynamics of SDWs and the convective mode, which both have zero potential vorticity. We first characterised the properties of the SDW and convective modes by finding the dispersion relations in the presence of disc rotation and stratification, but neglecting shear of disc flow. In this limit of rigid rotation, these two modes evolve independently, i.e., once either of these modes is initially excited, it does not lead to the excitation of another. We then demonstrated that in the presence of differential rotation/shear, an initially tightly leading SFH of the convective mode evolving in the disc flow, excites a corresponding SFH of SDWs. The main cause of wave generation is the following. Because of shear, the radial wavenumber of SFH varies with time, making the characteristic time-scales of these modes time-dependent as well. As a result, during a short period of time, when SFHs swing from leading to trailing, the mode time-scales can become comparable to each other and to the dynamical/shear time, leading to the possibility of an efficient energy exchange between the SDW mode, convective mode and the mean disc flow. Because of this, SDWs are generated in this process, mainly at the expense of shear flow energy, with the help of the convective mode. We also quantified the efficiency of wave generation for different azimuthal and vertical wavenumbers and found that it is maximal when these length-scales are comparable to the disc scale height. This, in turn, implies that SDWs excited by convection have weak vertical dependence, similar to SDWs generated by vortices (see also Heinemann and Papaloizou, 2009a). We calculated the angular momentum flux associated with non-axisymmetric SDWs generated by the convective mode and found that it is positive, that is, transport due to SDWs is outward, as opposed to that of the convective mode.

Analysed here linear coupling of SDWs and convection is similar in nature to the linear coupling of SDWs and vortices that has already been extensively studied and represents a special manifestation of a more general phenomenon of shear-induced linear coupling of perturbation modes inevitably taking place in any flow with an inhomogeneous velocity profile, or shear. Thus, although convection is not generally shear-driven – in that it does not directly tap energy from shear but from the unstable superadiabatic thermal distribution – its dynamics is still affected by shear and its shear-induced coupling with SDWs, in some sense, makes convection a participant in kinematic processes as well.

Shearing box simulation of compressible MRI turbulence in a Keplerian disc indicates that the angular momentum transport due to non-axisymmetric SDWs generated by vertical perturbations constitutes a significant fraction of that associated with the total turbulent Reynolds stress (Heinemann and Papaloizou, 2009b). In other words, a purely hydrodynamic part of transport is mostly due to these SDWs and is quite appreciable. As mentioned in the Introduction, in their non-linear simulations, Lesur and Ogilvie (2010) demonstrated that, contrary to previous results, in fact vertical convection has a non-axisymmetric structure and thus can transport angular momentum outwards. But the magnitude of the corresponding α reported in their simulations still turned out to be small, $\lesssim 10^{-4}$. However, these simulations were performed in the Boussinesq/incompressible limit, where high-frequency SDWs are filtered out. In a related study, Käpylä et al. (2010) solved full hydrodynamical equations in the shearing box with superadiabatic stratification without explicitly making the assumption of incompressibility, but the parameter regime considered was such as to give small Mach numbers, so that the effects of compressibility and, hence of SDWs, were probably small in their simulations and were not particularly investigated. As mentioned before, SDWs are generally able to enhance angular momentum transport rates. For example, Johnson and Gammie (2005b) demonstrated that compressible simulations of vortices yield at least an order of magnitude larger transport rates, which are primarily due to shocks of SDWs emitted by vortices, compared with those in the incompressible case, which are due to vortices only. Based on this property, we might anticipate a similar situation in compressible simulations of convectively unstable discs, where SDWs generated by the convective mode owing to disc flow shear also boost outward angular momentum transport due to the convective mode only. If so, then convection will become yet another promising candidate capable of mediating anomalous angular momentum transport in neutral discs, although this needs to be further investigated in greater detail and represents one of subjects of my future research. The existence of coupling between SDWs and convection revealed in this chapter is at least an essential factor and a first important step towards enhancing convective transport and makes the latter worthy of a subsequent more general (non-linear) examination. We emphasise once more that the studied here shear-induced coupling between the SDW and convective modes occurs only when these modes are non-axisymmetric. So, in the earlier simulations of Cabot (1996) and Stone and Balbus (1996), where vertical convection had an axisymmetric structure, SDW generation could not occur in spite of moderate Mach numbers associated with convective motions.

Another point I plan to investigate in the future is where in the disc SDWs generated by convection primarily cause dissipation through shock formation (see also Stone and Balbus, 1996), because in all the above-mentioned simulations, convection was not generated self-consistently, but maintained by artificially imposed heat flux. This becomes even more interesting, in the light of the result obtained here that in the linear regime SDWs do not vary much with height. If the shocks of SDWs dissipate near the surface, this may stifle convection, while if – near the midplane, this can then sustain convection.

Chapter 5

Axisymmetric modes in vertically stratified self-gravitating discs

5.1 Introduction

We have seen in chapters 2 and 3 that self-gravity plays an important role in the dynamics of protostellar and protoplanetary discs. It provides one of the main sources of outward angular momentum transport and is able to cause fragmentation of a disc into bound clumps, or planets, under certain conditions. Starting with the seminal paper by Toomre (1964), there have been a large number of studies of the stability of self-gravitating gaseous discs, both in the linear (e.g., Goldreich and Lynden-Bell, 1965a,b; Goldreich and Tremaine, 1978; Adams et al., 1989; Bertin et al., 1989; Papaloizou and Lin, 1989; Laughlin et al., 1997) and non-linear regimes, including other relevant physical factors (e.g., heating, cooling, radiation transport) with up-to-date numerical techniques (e.g., Papaloizou and Savonije, 1991; Laughlin and Bodenheimer, 1994; Boss, 1998; Pickett et al., 2000, 2003; Gammie, 2001; Boss, 2003; Johnson and Gammie, 2003; Rice et al., 2003, 2005; Boss, 2004; Mejía et al., 2005; Boley et al., 2006; Mayer et al., 2007; Stamatellos and Whitworth, 2008).

Linear stability analysis in a vast majority of cases is restricted, for simplicity, to razor-thin, or 2D discs that are obtained by vertically averaging all quantities. In other words, perturbations are assumed to have large horizontal length-scales compared with the disc thickness. In this limit, the well-known Toomre's parameter

$$Q_{2D} = \frac{c_s\Omega}{\pi G\Sigma}$$

controls the stability of self-gravitating discs (Toomre, 1964).[1] In this case, density waves and vortices, which are the only possible modes in a 2D disc, are influenced by self-gravity and thus can become unstable, as the analysis presented in chapter 3 clearly demonstrates.

Stability analysis in a more realistic case of self-gravitating three-dimensional discs is more complicated. The disc is vertically stratified due to both its own self-gravity and the vertical component of the gravity of a central object. Depending on the nature of the stratification, there exists a whole new set of various vertical modes in the system (see below), some of which can become unstable due to self-gravity on horizontal length-scales comparable to the disc thickness. In this situation, the vertical variation of perturbations is important and for a correct characterisation of the gravitational instability it is necessary to introduce another parameter not involving height-dependent variables, such as the sound speed in Toomre's parameter. Furthermore, not all types of stratification permit two-dimensional modes, that is, modes with no vertical motions commonly occurring in the 2D treatment. For example, in non-self-gravitating discs with polytropic vertical structure, there are no 2D modes (Lin et al., 1990; Lubow and Pringle, 1993b; Korycansky and Pringle, 1995), implying that the dynamics does not always reduce to that of the 2D case. Therefore, a more accurate stability analysis of self-gravitating discs should necessarily be three-dimensional.

Obviously, before studying the gravitational instability of stratified discs, one must first classify and characterise vertical normal modes of perturbations in the simplified case of no self-gravity (partly we have done such a classification in chapter 4 in the constant-g approximation). Analysis of the modal structure of stratified, polytropic, compressible, non-self-gravitating gaseous discs has been done in several papers: Ruden et al. (1988, hereafter RPL), Korycansky and Pringle (1995, hereafter KP), Ogilvie (1998). In convectively stable discs, i.e., with a subadiabatic vertical stratification, four principal types of vertical modes can be distinguished. These modes are: acoustic p-modes, surface gravity f-modes, buoyancy g-modes and inertial r-modes. The modes are named after their corresponding restoring forces, which can be well identified for each mode at horizontal wavelengths smaller than the disc height and are provided by one of the following: compressibility/pressure, displacements of free surfaces of a disc, buoyancy due to vertical stratification and inertial forces due to disc

---

[1] In this chapter, we distinguish between 2D and introduced below 3D Toomre's parameters with corresponding subscripts.

rotation, respectively, for the p-, f-, g- and r-modes. In the case of a superadiabatic stratification, the r- and g-modes merge and appear as a single mode, which becomes convectively unstable for horizontal wavenumbers larger than a certain value (RPL); the p- and f-modes remain qualitatively unchanged. For a neutral/adiabatic stratification, buoyancy is absent and the g-mode disappears. Our main goal in this chapter is to investigate how self-gravity modifies the frequencies and the structure of the eigenfunctions of these modes, which mode acquires the largest positive growth rate due to self-gravity and, therefore, determines the onset criterion and nature of the gravitational instability of a stratified disc. So, the mode dynamics in the 3D case can appear more complex than that in the 2D one, where only the density wave and vortical modes can be subject to the influence of self-gravity. Previously, Goldreich and Lynden-Bell (1965a, hereafter GLB) considered the gravitational instability of a uniformly rotating gaseous slab with an adiabatic vertical stratification, thereby leaving out all modes associated with buoyancy. Other studies also considered the gravitational instability in 3D in the context of galactic discs, however, the analysis was essentially 2D, finite-thickness effects were only taken into account by means of various reduction factors in the 2D dispersion relations (Shu, 1968; Romeo, 1992, 1994). In all these studies, as in GLB, the main focus was on finding the criterion for the onset of gravitational instability, so a full analysis of various types of vertical normal modes existing in stratified self-gravitating discs had not been carried out. Actually, we generalise the study of GLB to subadiabatic and superadiabatic stratifications having different modal structure.

Another motivation for our study is that the f-mode is thought to play an important dynamical role in self-gravitating discs. The non-linear behaviour of 3D perturbations involving large surface distortions, as seen in numerical simulations, has been attributed to the surface gravity f-mode (Pickett et al., 2000). However, this was done without analysing the behaviour of other vertical modes under self-gravity. It was shown that the f-mode leads to a large energy dissipation in the vicinity of the disc surface, which may facilitate disc cooling, because the energy is deposited at smaller optical depth where it can be radiated away more quickly (see e.g., Johnson and Gammie, 2003; Boley et al., 2006). Later it was realised that in fact the non-linear vertical motions in self-gravitating discs can be much more complex than just the f-mode and can have a shock character (shock bores, Boley and Durisen, 2006). Thus, in the 3D case, the dynamics of self-gravitating discs is much richer and diverse than that of idealised 2D ones and requires further study. To fully understand the origin of such three-dimensional effects and what type of instability they are associated with, one should start with a rigorous linear study of the characteristic properties of all the types of vertical normal modes mentioned above, not only the f-mode, in the presence of self-gravity. The present work is just a first step in this direction.

Numerical simulations of self-gravitating discs are often in the context of global discs (e.g., Pickett et al., 2000, 2003; Rice et al., 2003, 2005; Lodato and Rice, 2004, 2005; Boley et al., 2006; Boley, 2009) and, therefore, are not always able to well resolve vertical motions, which, as shown in the present study, inevitably arise during the development of the gravitational instability associated with intrinsically three-dimensional modes. So, these simulations may not quite accurately capture all the subtleties of the gravitational instability in 3D discs. In this connection, we should mention the work by Nelson (2006) that extensively discusses the issue of vertical resolution and its importance in the outcome of the gravitational instability in numerical simulations of self-gravitating discs. Resolving and analysing vertical motions are also crucial for properly understanding cooling processes in discs and, particularly, whether convection is able to provide sufficiently effective cooling for disc fragmentation to occur, which is still a matter for debate in the literature (Boss, 2004; Mayer et al., 2007; Boley et al., 2006, 2007; Rafikov, 2007). In addition, these studies, for simplicity, use the criterion for gravitational instability based on the 2D Toomre's parameter, which, as we will demonstrate, cannot always be uniquely mapped into an analogous three-dimensional stability parameter and give a precise criterion for the onset of gravitational instability.

In this chapter, following other works in a similar vein: Lubow and Pringle (1993b, hereafter LP), KP, Lubow and Ogilvie (1998, hereafter LO98), we again adopt the shearing box approximation and consider the linear dynamics of vertical normal modes of perturbations in a compressible, stratified, self-gravitating gaseous disc with Keplerian rotation (without making the simplifying, constant-g assumption). In the unperturbed disc, pressure and density are related by a polytropic law, which is a reasonably good approximation for optically thick discs (see e.g., LO98). This allows us to consider the specific features of the dynamics for subadiabatic, adiabatic and superadiabatic vertical stratifications by simply varying the polytropic index. As a first step towards understanding the effects of self-gravity on the perturbation modes, we restrict ourselves to axisymmetric perturbations only. Although with this we cannot study the effects of disc flow shear on the dynamics of 3D modes, the linear results obtained here will form the basis for studying the non-linear development of gravitational instability in the local shearing box approximation that allows much higher numerical resolution than global disc models can afford.

The chapter is organised as follows.[2] The physical model and basic equations are intro-

---

[2] This chapter is based on my paper with the collaborator (supervisor): Mamatsashvili and Rice (2010).

duced in section 5.2. The classification of vertical modes in the absence of self-gravity is performed in section 5.3. Effects of self-gravity on all normal modes are analysed in section 5.4. In section 5.5, we focus on the properties of gravitational instability in 3D. Comparison between the criteria of gravitational instability in 3D and 2D is made in section 5.6. A summary and discussions are given in section 5.7.

5.2 Physical Model and Equations

We study the dynamics of three-dimensional modes in self-gravitating gaseous discs in the local shearing box approximation, as in chapter 4. Now we include both the vertical component of the gravity force of the central object and disc self-gravity. The basic equations in this case are Eqs. (2.12-2.14), which we give here again

$$\frac{\partial \mathbf{u}}{\partial t} - q\Omega x\frac{\partial \mathbf{u}}{\partial y} + (\mathbf{u}\cdot\nabla)\mathbf{u} - q\Omega u_x\mathbf{e}_y + 2\Omega\mathbf{e}_z\times\mathbf{u} + \Omega^2 z\mathbf{e}_z = -\frac{1}{\rho}\nabla p - \nabla\psi, \tag{5.1}$$

$$\frac{\partial \rho}{\partial t} - q\Omega x\frac{\partial \rho}{\partial y} + \nabla\cdot(\rho\mathbf{u}) = 0. \tag{5.2}$$

and the equation of conservation of entropy

$$\left(\frac{\partial}{\partial t} - q\Omega x\frac{\partial}{\partial y} + \mathbf{u}\cdot\nabla\right)\frac{p}{\rho^\gamma} = 0. \tag{5.3}$$

This set of equations is supplemented with Poisson's equation to take care of disc self-gravity

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}\right)\psi = 4\pi G\rho. \tag{5.4}$$

Here, as before, $\mathbf{u} = (u_x, u_y, u_z)$ is the velocity in the local Cartesian frame relative to the background Keplerian shear flow $\mathbf{u}_0 = (0, -q\Omega x, 0)$, p, ρ, ψ are, respectively, the pressure, density and the gravitational potential of the disc gas. The shear parameter is $q = 1.5$. The adiabatic index, or the ratio of specific heats, $\gamma = 1.4$ as typical of a disc composed of molecular hydrogen; we adopt this value throughout this chapter.

5.2.1 The equilibrium disc model

Equations (5.1-5.4) have an equilibrium solution that is stationary and axisymmetric. In this unperturbed state, the velocity is zero and there is only a parallel azimuthal shear flow, $\mathbf{u_0}$. The equilibrium density ρ_0, pressure p_0 and gravitational potential ψ_0, depending only on the vertical $z-$coordinate, satisfy the hydrostatic relation with self-gravity (cf. Eq. 4.4)

$$g_0 \equiv -\frac{1}{\rho_0}\frac{dp_0}{dz} = \Omega^2 z + \frac{d\psi_0}{dz}, \tag{5.5}$$

$$\frac{d^2\psi_0}{dz^2} = 4\pi G\rho_0. \tag{5.6}$$

As in LP, KP and RPL, pressure and density in the unperturbed disc are related by a polytropic relationship of the form

$$p_0 = K\rho_0^{1+1/s}, \tag{5.7}$$

where K is the polytropic constant and $s > 0$ is the polytropic index. The Brunt-Väisälä frequency squared is

$$N_0^2 \equiv \frac{g_0}{\rho_0}\left(\frac{1}{c_s^2}\frac{dp_0}{dz} - \frac{d\rho_0}{dz}\right) = \left(\frac{\gamma s}{s+1} - 1\right)\frac{g_0^2}{c_s^2}, \tag{5.8}$$

where $c_s^2 = \gamma p_0/\rho_0$ is the adiabatic sound speed squared. If $1+1/s < \gamma$ (subadiabatic thermal stratification), then $N_0^2 > 0$ all along the height and the equilibrium vertical structure of the disc is convectively stable. In the opposite case $1+1/s > \gamma$ (superadiabatic thermal stratification), $N_0^2 < 0$ everywhere and this corresponds to a convectively unstable equilibrium. For $1+1/s = \gamma$ (adiabatic thermal stratification), $N_0^2 = 0$ and all the motions/modes due to buoyancy disappear. Later we will consider these three regimes separately.

To determine the vertical structure, we need to solve Eqs. (5.5-5.6) subject to the boundary condition that the pressure vanish at the free surface of the disc. Because we have a polytropic model, it is convenient to work with the pseudo-enthalpy

$$w_0 = (s+1)K\rho_0^{1/s}.$$

The disc structure is also symmetric with respect to the midplane, $z = 0$, and, as a consequence, it follows from Eqs. (5.5-5.7) that the derivative of w_0 at the midplane vanishes, $dw_0/dz(0) = 0$. Because of this reflection symmetry, we consider only the upper half of the disc, $z \geq 0$. At the surface of the disc $w_0 = 0$, similar to the pressure. This allows us to determine the equilibrium vertical structure of the disc. The non-dimensional variables being used throughout the chapter are:

$$x \to \frac{x\Omega}{c_{sm}}, \quad y \to \frac{y\Omega}{c_{sm}}, \quad z \to \frac{z\Omega}{c_{sm}},$$

$$\rho_0 \to \frac{\rho_0}{\rho_m}, \quad w_0 \to \frac{w_0}{w_m}, \quad c_s \to \frac{c_s}{c_{sm}},$$

where $\rho_m \equiv \rho_0(0)$, $w_m \equiv w_0(0)$, $c_{sm} \equiv c_s(0)$ are the midplane values of the equilibrium density, pseudo-enthalpy and sound speed. We define the three-dimensional analogue of Toomre's parameter as

$$Q_{3D} = \frac{\Omega^2}{4\pi G\rho_m},$$

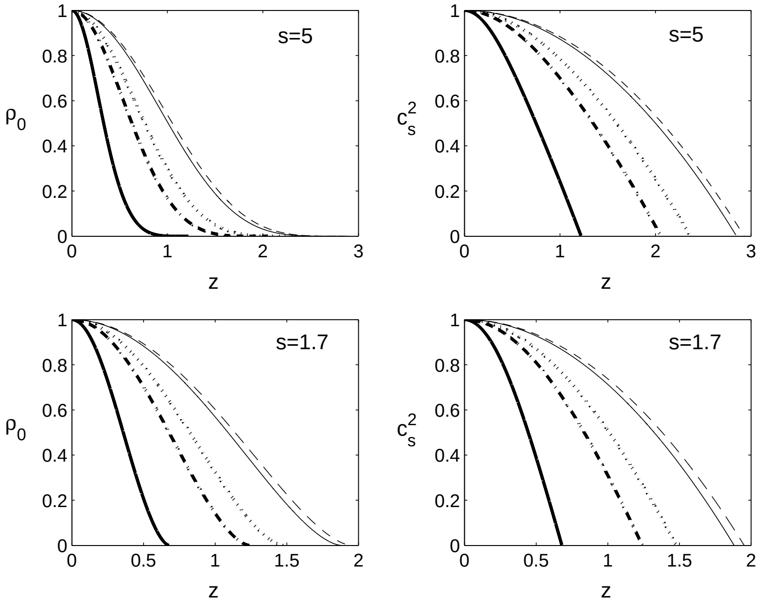

Figure 5.1: Vertical variations of the normalised density and squared sound speed for subadiabatic ($s = 5$) and superadiabatic ($s = 1.7$) equilibrium states with $Q = 0.1$ (thick solid lines), $Q = 0.5$ (dashed-dotted lines), $Q = 1$ (dotted lines), $Q = 10$ (thin solid lines) and the non-self-gravitating case ($Q \to \infty$, dashed lines). The lines corresponding to $Q = 10$ are close to those in the non-self-gravitating limit, implying that the role of self-gravity is negligible for $Q \geq 10$. The disc heights, h, are different depending on Q and s.

which is a measure of disc self-gravity (from now on until section 5.6, we will use Q_{3D} without subscript everywhere, so it should not be confused with the 2D Toomre's parameter). Using that at the midplane $dw_0/dz(0) = 0$, we can derive from Eqs. (5.5-5.7) a single equation for the normalised pseudo-enthalpy

$$\frac{s+1}{2\gamma}\left(\frac{dw_0}{dz}\right)^2 = (1 - w_0) + \frac{1}{Q(s+1)}(1 - w_0^{s+1}). \tag{5.9}$$

The normalised density and sound speed are found from $\rho_0 = w_0^s$, $c_s^2 = w_0$. Equation (5.9) shows that even though the disc is in Keplerian rotation, that is, self-gravity in the radial direction can be neglected, it must be included in determining the vertical structure. We integrate Eq. (5.9), starting at $z = 0$ with $w_0(0) = 1$, until w_0, monotonically decreasing with z, reaches zero at some finite height/edge $z = h$, which is therefore determined as a result of the integration process. At this edge, the density and sound speed also vanish, $\rho_0(h) = 0, c_s(h) = 0$ (see also GLB and RPL). The height and entire vertical structure of a polytropic disc are therefore uniquely specified solely by Q and s, which are free parameters in Eq. (5.9).

Figure 5.1 illustrates the sub- and superadiabatic equilibrium vertical structures obtained from Eq. (5.9) for various Q. For a fixed s, the disc height h decreases with decreasing Q. The sound speed is a decreasing function of z that does not permit the existence of two-dimensional modes in three-dimensional models of polytropic discs as opposed to isothermal ones (Lin et al., 1990, LP, KP). We also see that the equilibrium structures for sub- and superadiabatic cases look similar except that they have, respectively, everywhere positive and negative N_0^2 that diverges at the surface because the sound speed vanishes there.

5.2.2 Perturbation equations

Let us now consider small axisymmetric ($\partial/\partial y = 0$) perturbations of the form $\mathbf{u}, \rho', p', \psi' \propto \exp(-i\omega t + ikx)$ about the equilibrium states found above, where ω and k are the frequency and the radial wavenumber, respectively. Without a loss of generality, we assume throughout that k is non-negative, $k \geq 0$. After switching to non-dimensional variables:

$$t \to \Omega t, \quad \omega \to \frac{\omega}{\Omega},$$

$$k \to \frac{kc_{sm}}{\Omega}, \quad N_0 \to \frac{N_0}{\Omega}, \quad g_0 \to \frac{g_0}{\Omega c_{sm}},$$

$$u_x \to \frac{u_x}{c_{sm}}, \quad u_y \to \frac{u_y}{c_{sm}}, \quad u_z \to \frac{u_z}{c_{sm}},$$

$$\rho' \to \frac{\rho'}{\rho_m}, \quad p' \to \frac{p'}{\rho_m c_{sm}^2}, \quad \psi' \to \frac{\psi'}{c_{sm}^2},$$

and linearising Eqs. (5.1-5.4), we obtain the following set of equations governing the linear dynamics of axisymmetric perturbations

$$-i\omega u_x = -\frac{ikp'}{\rho_0} + 2u_y - ik\psi' \tag{5.10}$$

$$-i\omega u_y = (q-2)u_x \tag{5.11}$$

$$-i\omega u_z = -\frac{1}{\rho_0}\frac{dp'}{dz} - g_0\frac{\rho'}{\rho_0} - \frac{d\psi'}{dz} \tag{5.12}$$

$$-i\omega\rho' + ik\rho_0 u_x + \frac{d}{dz}(\rho_0 u_z) = 0 \tag{5.13}$$

$$-i\omega(p' - c_s^2\rho') + \frac{\rho_0 c_s^2 N_0^2}{g_0}u_z = 0 \tag{5.14}$$

$$\left(-k^2 + \frac{d^2}{dz^2}\right)\psi' = \frac{\rho'}{Q}. \tag{5.15}$$

If we now make the changes: $u_z \to iu_z$, $p' \to \omega p'/\rho_0$, $\psi' \to \omega\psi'$ and eliminate u_x, u_y, ρ' from Eqs. (5.10), (5.11) and (5.13), we arrive at the following set of equations for the three basic quantities u_z, p', ψ' (henceforth primes will be omitted)

$$\frac{du_z}{dz} = \frac{g_0}{c_s^2}u_z - \left(\frac{1}{c_s^2} - \frac{k^2}{\omega^2 - \kappa^2}\right)p + \frac{k^2\psi}{\omega^2 - \kappa^2} \tag{5.16}$$

$$\frac{dp}{dz} = \frac{N_0^2}{g_0}p + (\omega^2 - N_0^2)u_z - \frac{d\psi}{dz} \tag{5.17}$$

$$\frac{d^2\psi}{dz^2} - k^2\psi = \frac{\rho_0}{Q}\left(\frac{p}{c_s^2} + \frac{N_0^2}{g_0}u_z\right), \tag{5.18}$$

where the non-dimensional epicyclic frequency κ is given by $\kappa^2 = 2(2-q) = 1$. Notice that the density perturbation on the right hand side of the linearised Poisson's equation (5.18) consist of two physically different parts – the density perturbation due to compressibility, which are inversely proportional to the squared sound speed and proportional to the pressure perturbation, and the density perturbation due to the stratified background, which are proportional to the vertical velocity perturbation and N_0^2. This latter term due to stratification is obviously absent in the two-dimensional analysis of gravitational instability as well as when the disc is adiabatic (see GLB). Equations (5.16-5.18) form the basis for determining the axisymmetric normal modes of perturbations existing in compressible, stratified, polytropic, self-gravitating gaseous discs. In the non-self-gravitating limit, these equations reduce to the main equations of KP. The reflection symmetry of the unperturbed disc allows us to take the eigenfunctions/modes to be either even or odd with respect to z. So, to determine the eigenfrequencies (dispersion diagrams, that is, ω as a function of k for various branches) and eigenfunctions, we can numerically integrate the main Eqs. (5.16-5.18) only in the upper half, $0 \leq z \leq h$, of the full vertical extent of the disc, provided suitable boundary conditions are imposed at $z = 0$ and $z = h$. In other words, we need to solve a boundary value problem. Here the disc height h, as noted above, is determined by the parameters Q and s. By inspecting Eqs. (5.16-5.18), it is clear that a normal mode whose pressure p and potential ψ perturbations are even functions under reflection in z, has odd vertical velocity u_z and gravitational potential derivative $d\psi/dz$ perturbations under reflection and vice versa. We define a normal mode as being 'even' ('odd') if the corresponding vertical velocity and gravitational potential derivative are odd (even) functions of z (this convention agrees with that of Ogilvie, 1998, but differs from LP and KP). Thus, the boundary conditions at the midplane, $z = 0$, for the even modes are:

$$u_z(0) = 0,\ d\psi/dz(0) = 0 \tag{5.19}$$

and for the odd modes are:

$$p(0) = 0,\ \psi(0) = 0. \tag{5.20}$$

At $z = h$ we impose the usual free-surface boundary condition for which the Lagrangian pressure perturbation vanishes. In our new non-dimensional variables this translates as

$$p = g_0 u_z \quad at \quad z = h. \tag{5.21}$$

The boundary condition for the gravitational potential can be derived by treating the surface displacement as a surface distribution of gravitating matter at $z = h$. Using the continuity of potential across the boundary, Gauss's flux theorem and the fact that outside the disc, at $z \to \pm\infty$, gravitational potential should vanish, we arrive at the following condition at the disc surface (see e.g., GLB, Lubow and Pringle, 1993a)

$$\frac{d\psi}{dz} + k\psi = -\frac{u_z \rho_0}{Q} \qquad at \qquad z = h. \tag{5.22}$$

Equations (5.16-5.18) supplemented with boundary conditions (5.21), (5.22) at the surface $z = h$ and (5.19), (5.20) at the midplane fully determine a boundary value problem. Using these boundary conditions and variational principle, it can be easily shown that ω^2 is real (GLB, RPL) that much alleviates the search of eigenfrequencies. We integrate these equations from $z = h$ downwards to $z = 0$ separately for the even modes under conditions (5.19) and for the odd modes under conditions (5.20). We will see below that in the presence of self-gravity, the dispersion diagrams for these two mode parities are quite different. In other words, self-gravity influences even and odd modes, or changes their dispersion characteristics, in different ways. As is typical, for a given equilibrium structure, that is, for given Q and s, and for a given radial wavenumber k, midplane boundary conditions can be satisfied only for certain values of ω, yielding dispersion relations $\omega(k)$ for different branches of modes classified below. We use a Runge-Kutta integrator and root-finding algorithms of MATLAB package to first numerically solve differential Eqs. (5.16-5.18) and then find eigenfrequencies.

5.3 The classification of vertical modes

In this section, we for the moment remove self-gravity (put $\psi \to 0, Q \to \infty$) from the perturbation Eqs. (5.16-5.18) and from the equilibrium in order to first classify and characterise all the vertical axisymmetric normal modes existing in a three-dimensional disc. An analogous classification of modes in non-self-gravitating gaseous discs has been done before in special cases (RPL, LP, RG92, KP, Ogilvie, 1998, and also in chapter 4 under the simplifying assumption of constant vertical gravity). The aim of this section is to briefly review and synthesise these results. In the next section, we will again switch on self-gravity in these equations and investigate how it affects the characteristics of these modes. Accordingly, in the non-self-gravitating case, we need the midplane and surface boundary conditions discussed above, but only for pressure and vertical velocity perturbations.

Figure 5.2 shows the typical dispersion diagrams for three types of equilibria: subadiabatic ($1 + 1/s < \gamma$), adiabatic ($1 + 1/s = \gamma$) and superadiabatic ($1 + 1/s > \gamma$) stratifications. In the subadiabatic case, one can identify four distinct classes of vertical normal modes (see also KP, Ogilvie, 1998). These modes are: the acoustic p-mode, the surface gravity f-mode, the buoyancy g-mode and the inertial r-mode the restoring forces for which at large radial wavenumbers $kh \gg 1$ are mainly provided, respectively, by compressibility/pressure, by the displacements of free surface of the disc, by buoyancy due to the vertical entropy gradient and by inertial forces due to disc rotation.[3] The existence of the g-mode in polytropic discs is attributable to the variation of the sound speed and N_0 with height; the latter diverges near the disc surface giving rise to the trapped g-mode there. For $1 + 1/s < \gamma$, all these modes have $\omega^2 > 0$ and therefore are stable. The p-, f- and g-modes are high-frequency modes with frequencies always larger than the epicyclic frequency, $\omega^2 \geq \kappa^2$, while the r-mode is of low-frequency with $\omega^2 \leq \kappa^2$ (we remind that hereafter $\kappa = 1$).

For an adiabatic stratification, there is no buoyancy ($N_0^2 = 0$) and, therefore the g-mode disappears, while the other modes remain qualitatively unchanged. Similarly, the r-mode does not exist in a disc that has zero epicyclic frequency. For a superadiabatic stratification, $1+1/s > \gamma$, the g-mode becomes convectively unstable ($\omega^2 < 0$) because of negative buoyancy for radial wavenumbers larger than a certain value. Its characteristic time-scale (growth rate) can be of the order of the epicyclic frequency or less and due to this, it interferes with motions corresponding to the r-mode. Consequently, in the superadiabatic case, the r- and g-modes merge at $\omega^2 \leq \kappa^2$ and appear as a single mode, which we still call the g-mode and not the r-mode, because the behaviour of corresponding eigenfunctions with height in this case is similar to that of the convectively stable g-mode (see also RPL); the p- and f-modes are not much affected.

All these modes come in even and odd pairs. In Fig. 5.2, even (odd) modes are numbered according to the number of nodes of the vertical velocity (pressure) perturbation in the interval $0 < z \leq h$ (the node at $z = 0$ is not counted). We also see the coalescence of the dispersion curves of the even and odd p- and f-modes as well as the convectively unstable even and odd g-modes as k increases past a certain point, which depends on the mode number (see RPL, KP). This is associated with a transition from an oscillatory to evanescent nature of eigenfunctions. We do not plot eigenfunctions in this non-self-gravitating case here, as an extensive discussion of their properties can be found in RPL, KP, LO98 and Ogilvie (1998). We only mention that the eigenfunctions of the p- and g-modes are trapped near the surfaces and decay towards the midplane, while the eigenfunctions of the r-mode are

---

[3] Performing the above classification of vertical modes, we have omitted the vortical mode, because, as pointed out before, in the axisymmetric case it is a stationary solution of Eqs. (5.10-5.15) with $\omega = 0$.

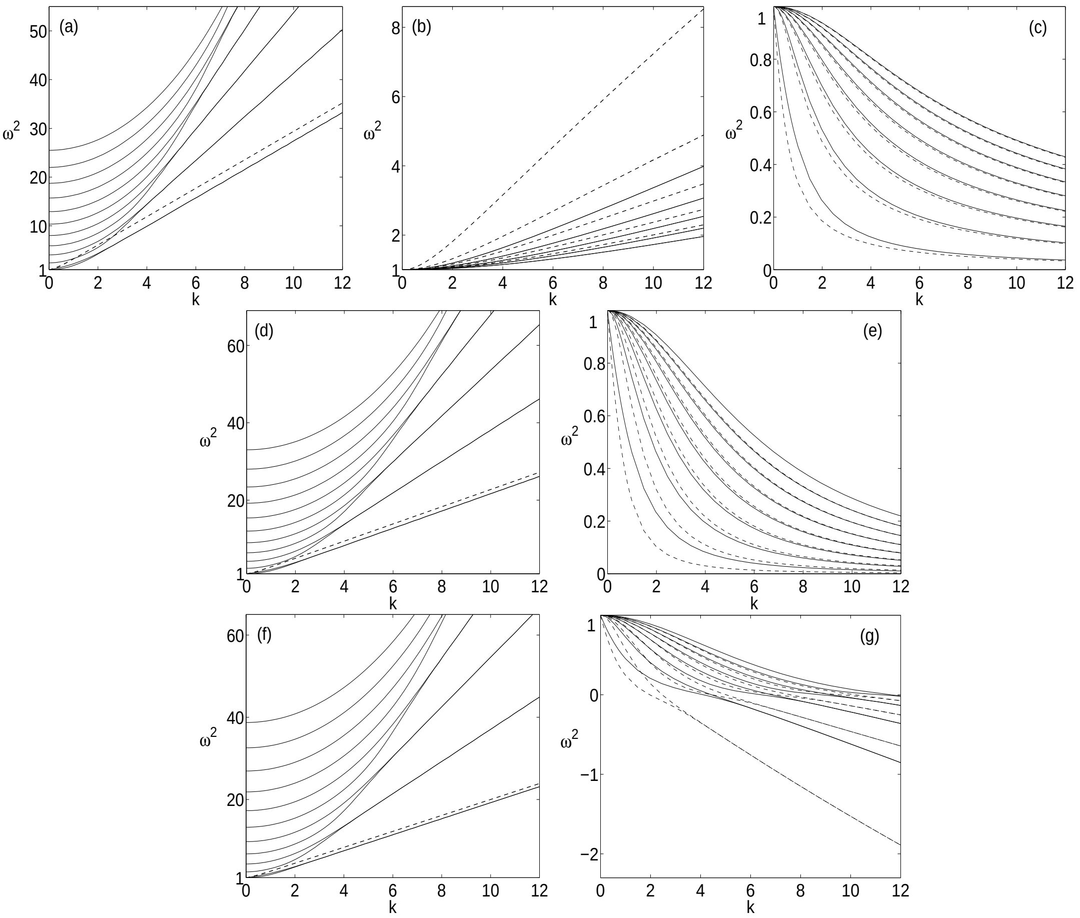

Figure 5.2: Dispersion diagrams in the non-self-gravitating case for a subadiabatic disc with $s = 5$ (panels a, b, c), for an adiabatic disc with $s = 2.5$ (panels d, e) and for a superadiabatic disc with $s = 1.7$ (panels f, g). The solid lines in panels (a), (d), (f) show the frequencies for different branches of the surface gravity f- and acoustic p-modes vs. radial wavenumber k. In order of increasing frequency these branches/curves are $\mathrm{f^e, f^o, p_1^e, p_1^o, p_2^e, p_2^o, p_3^e, p_3^o, p_4^e, p_4^o, p_5^e, p_5^o}$ (superscripts denote the even and odd modes, which are numbered, respectively, according to the number of nodes of vertical velocity and pressure perturbations in $0 < z \leq h$). For large k, the frequencies of even and odd modes coalesce. The dashed lines in these panels correspond to frequencies of the f-modes computed in the incompressible limit. The solid lines in panel (b) show the convectively stable g-mode branches, which in order of decreasing frequency are $\mathrm{g_1^{o,e}, g_2^{o,e}, g_3^{o,e}, g_4^{o,e}, g_5^{o,e}}$. Here the frequencies of even and odd modes coincide. The solid lines in panels (c),(e) show the r-mode branches, which in order of increasing frequency are $\mathrm{r_0^o, r_0^e, r_1^o, r_1^e, r_2^o, r_2^e, r_3^o, r_3^e}$ ($\mathrm{r_0^{o,e}}$ are the basic odd/even r-modes and subscript '0' means that, respectively, pressure/vertical velocity for them has no nodes in $0 < z \leq h$. In this respect, our numbering of modes differs from that of Ogilvie, 1998). The solid lines in panel (g) show the convectively unstable g-modes, which in order of increasing ω^2 are $\mathrm{g_0^o, g_0^e, g_1^o, g_1^e, g_2^o, g_2^e, g_3^o, g_3^e}$ ($\mathrm{g_0^{o,e}}$ are the basic odd/even g-modes with no nodes, respectively, of pressure/vertical velocity in $0 < z \leq h$). Similar to the p-modes, the convectively unstable even and odd g-modes coalesce at large k. In each panel, the dashed lines show the corresponding mode branches computed for the incompressible case with the same ordering of eigenfrequencies. From panels (b), (g), we see that compressibility most strongly affects the convectively stable and unstable g-modes.

concentrated near the midplane and decay towards the surfaces. As for the eigenfunctions of the fundamental f-mode, they have no nodes and monotonically decay from the surfaces to the midplane. Below we show that these properties of eigenfunctions are altered by self-gravity. Specifically, the number of nodes along each branch of the dispersion diagrams, which is preserved in the non-self-gravitating limit, is no longer constant in the presence of self-gravity. The frequencies of the p- and r-modes increase with mode number. The frequencies and growth rates, respectively, of the convectively stable and unstable g-modes, decrease with the mode number. These are the well-known general properties of vertical modes in polytropic discs.

In chapter 4, we considered discs with an isothermal vertical structure, where the equilibrium density and pressure fall off exponentially with height, instead of forming a sharp boundary as in polytropic discs. As a result, in isothermal discs, as noted above, only p- and r-modes can exist for $N_0^2 \geq 0$ (the convectively stable g-mode is not possible, see LP) and the p- and convectively unstable g-modes for $N_0^2 < 0$. Accordingly, the classified in section 4.2.3 acoustic mode is the same as the p-mode and the inertia-buoyancy mode is the same as the r- or g-mode, depending on the sign of N_0. The constant-g approximation adopted in chapter 4 is in fact appropriate at large vertical wavenumbers, so the dispersion relations (4.21) and (4.22) derived there, as a matter of fact, describe, respectively, the p- and r-,g-modes at high mode numbers, when the characteristic vertical length-scale of these modes is smaller than the disc scale height, though they still remain reasonably valid even at smaller mode numbers, that is, at vertical length-scales comparable to the scale height.

Although each mode type has its dominant restoring force, one out of the above mentioned four types (compressibility, surface gravity, buoyancy, inertial forces), the other three forces also contribute to a certain degree at smaller radial wavenumbers $kh \lesssim 1$. For example, the p- and f-modes are modified both by rotation/inertial forces and buoyancy, the f-mode is also modified by compressibility, the g-mode is modified by rotation and compressibility and the r-mode is modified by buoyancy and compressibility. For the sake of comparison with the relevant results from previous studies, we are particularly interested in to what degree compressibility modifies generally non-compressive f-, g- and r-modes. So, we decided to juxtapose the dispersion diagrams for these modes computed separately in the compressible and incompressible cases. We take the incompressible limit by formally letting the adiabatic index go to infinity, $\gamma \to \infty$ (Ogilvie, 1998). After that Eqs. (5.16) and (5.17) without self-gravity take the form

$$\frac{du_z}{dz} = \frac{k^2}{\omega^2 - \kappa^2} p$$

$$\frac{dp}{dz} = \frac{N_0^2}{g_0} p + (\omega^2 - N_0^2) u_z.$$

We solved these equations with the same boundary conditions and found the dispersion diagrams of the f-, g- and r-modes that survive in this limit (obviously, the acoustic p-mode disappears). The results are plotted in Fig. 5.2 with dashed lines. Notice that by taking the incompressible limit in this manner, we have been able to retain the f-mode, as expected. Instead, KP set the density perturbations to zero in the linearised continuity equation (anelastic approximation) that resulted in the f-mode disappearing in their incompressible limit. We see that compressibility most strongly affects the convectively stable and unstable g-mode branches with small mode numbers (Figs. 5.2b and 5.2g), or equivalently with vertical extent comparable to the disc height, even at large radial wavenumbers (Ogilvie, 1998). The frequencies of the f- and r-modes do not change much, indicating that these modes are nearly incompressible.

5.4 Vertical modes in the presence of self-gravity

Having characterised all types of axisymmetric normal modes in a stratified disc, let us now compute the dispersion diagrams taking into account self-gravity in the perturbation equations and in the equilibrium. This will allow us to understand how the frequencies and the structure of the eigenfunctions of the above-described mode types are altered by self-gravity, which mode acquires the largest positive growth rate in the presence of self-gravity and, thus determines the onset criterion and nature of gravitational instability of a disc. In other words, we return to the boundary value problem formulated in section 5.2, which is represented by Eqs. (5.16-5.18) supplemented with boundary conditions (5.21), (5.22) at the surface and conditions (5.19), (5.20) of the even and odd symmetry of a solution at the midplane. For each Q and s, we first determine the corresponding vertical distribution of the equilibrium quantities in Eqs. (5.16-5.18), $\rho_0, c_s^2, g_0, N_0^2$, with height as described in section 5.2 and then based on this compute the normal modes.

Figures 5.3, 5.4, 5.5 and 5.6 show the typical dispersion diagrams for the p-, f-, g- and r-modes in a self-gravitating disc for subadiabatic, adiabatic and superadiabatic vertical stratifications. We separately plot the dispersion diagrams of the even and odd parities for each mode type to clearly bring out the influence of self-gravity on them, which, as evident from these figures, depends on the mode parity. Unlike in the non-self-gravitating limit, the number of nodes of the vertical velocity and pressure perturbations in the presence of self-gravity are not preserved along each mode branch. However, as we will see below, at large k, the

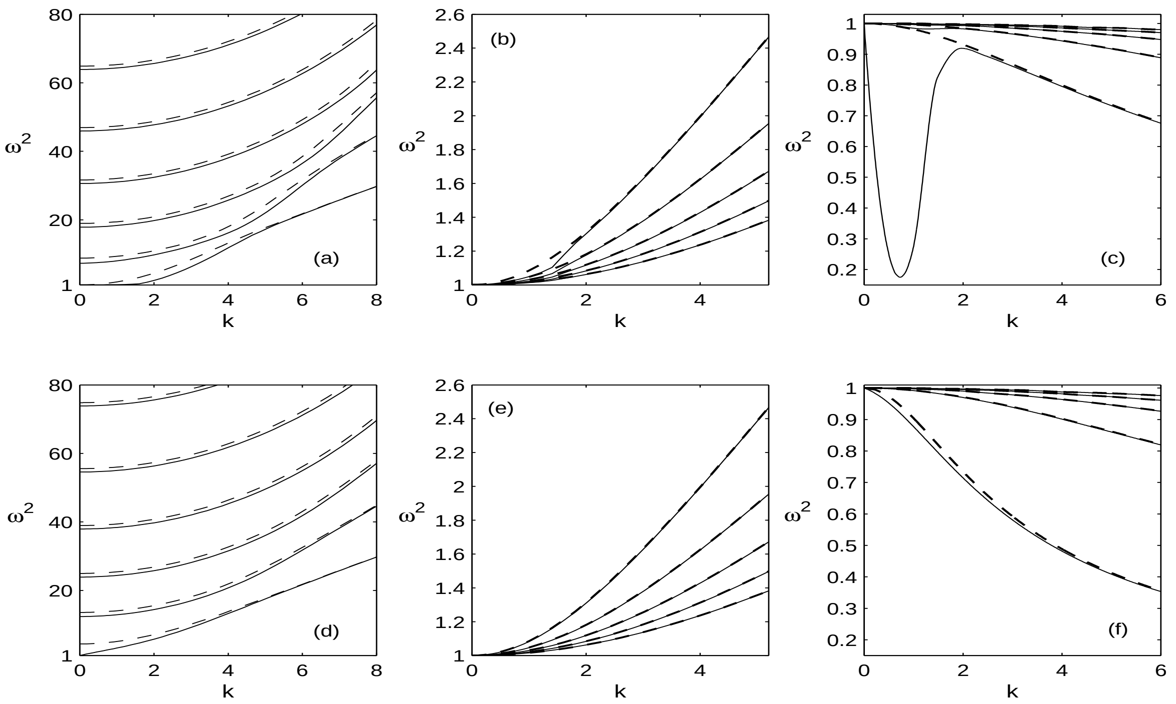

Figure 5.3: Dispersion diagrams of vertical modes in a subadiabatic self-gravitating disc with $s = 5$ and $Q = 0.2$. Shown are: (a) the even p- and f-modes, (b) the even g-mode, (c) the even r-mode, (d) the odd p- and f-modes, (e) the odd g-mode, (f) the odd r-mode. Dashed lines are the corresponding branches computed without self-gravity in perturbation Eqs. (5.16-5.18), but with the same self-gravitating equilibrium. The frequency ordering and mode naming are the same as in Fig. 5.2. Self-gravity reduces (shifts) the frequencies of the even and odd p-, f-modes and the even g- and r-modes mainly in the range $0 \leq k \leq 2$; the frequencies of the odd g- and r-modes are almost unaffected by self-gravity. The frequency of the basic even r-mode (denoted above as $\mathrm{r}_0^{\mathrm{e}}$) is modified most strongly by self-gravity, as evidenced by the largest dip on the corresponding dispersion curve in panel (c). For large k and/or large mode numbers, the effect of self-gravity becomes weak and the dispersion curves merge with the dashed ones for the non-self-gravitating case.

influence of self-gravity on the mode dynamics is small and the dispersion diagrams merge with their non-self-gravitating counterparts (shown with dashed curves in Figs. 5.3 and 5.5). So, the naming of the modes for smaller k, where the effect of self-gravity is important, is done by continuity with the large-k limit for each mode branch.

5.4.1 Subadiabatic stratification

Consider first the subadiabatic case (Fig. 5.3), where we again have the p-, f-, g- and r-modes, but their dispersion diagrams are modified (shifted) from their non-self-gravitating counterparts towards lower values due to self-gravity. As illustrated in Figs. 5.3a and 5.3d, self-gravity reduces the frequencies of all branches of the p- and f-modes, for both even and odd parities, but it more affects the f-modes. The situation for the g- and r-modes is different (Figs. 5.3b, 5.3c, 5.3e and 5.3f): only the frequencies of the even g- and r-modes are reduced by self-gravity mostly for radial wavenumbers in the range $0 \leq k \leq 2$ (dips on the corresponding dispersion curves in Figs. 5.3b and 5.3c indicating deviations from the non-self-gravitating dashed ones), whereas the frequencies of the odd g- and r-modes are almost unaffected by self-gravity. The frequencies of the fundamental f-modes, the first few branches of the p-modes and also the first few branches of the even g- and r-modes are reduced noticeably. With increasing mode number and/or radial wavenumber, the effect of self-gravity on the eigenfrequencies gradually falls off, because the corresponding eigenfunctions become of smaller and smaller horizontal and/or vertical scale (the number of nodes increases) and take the form similar to that in the non-self-gravitating limit. As a result, the dispersion diagrams become more and more identical to their non-self-gravitating counterparts. The ordering of frequencies for the even and odd p-, f-, g- and r-modes, as it is in the non-self-gravitating case (see Fig. 5.2): $\omega^2(\mathrm{r}) \leq \kappa^2 \leq \omega^2(\mathrm{g}) < \omega^2(\mathrm{f}) < \omega^2(\mathrm{p})$, is preserved in the self-gravitating case as well, however small the value of Q. Figure 5.4 shows that the frequencies of the even and odd f-modes and also of the basic even g-mode, for which the influence of self-gravity is stronger, never fall below the epicyclic frequency with decreasing Q and, therefore, these modes always remain gravitationally stable (we also checked that the same is true for the dispersion curves of the even and odd f-modes in the adiabatic and superadiabatic cases described below). This is a consequence of the three-dimensionality of the disc. Thus, as evident from Figs. 5.3c and 5.4d, self-gravity most significantly affects the basic branch of the even r-mode and only this branch can become unstable due to self-gravity. From Fig. 5.4d, it is seen that the dip on this branch deepens and broadens with decreasing Q. (We denote the wavenumber at which the minimum of this dip is located by k_m.) The gravitational instability sets in when the dip's minimum first touches the $\omega^2 = 0-$axis at some Q_{cr} and then extends farther into the unstable $\omega^2 < 0$ region for $Q < Q_{cr}$. We will demonstrate in section 5.5 that the basic even r-mode becomes strongly compressible at $k \sim k_m$ (Fig. 5.8), as opposed to the r-mode in a non-self-gravitating disc, and the density perturbations due to compressibility are responsible for its gravitational instability. For $s = 5$, we find $Q_{cr} = 0.168$, which gives $H = 2h = 2.96$ for the disc total thickness, and $k_m = 0.8$ (see Fig. 5.12). Note that the characteristic radial wavelength of instability $\lambda_m = 2\pi/k_m = 7.85$ is not much greater than the disc thickness H. In self-gravitating discs, compressibility and inertial forces play an important role in the dynamics of the basic even r-mode at wavelengths $\lambda \sim \lambda_m \gtrsim H$ and

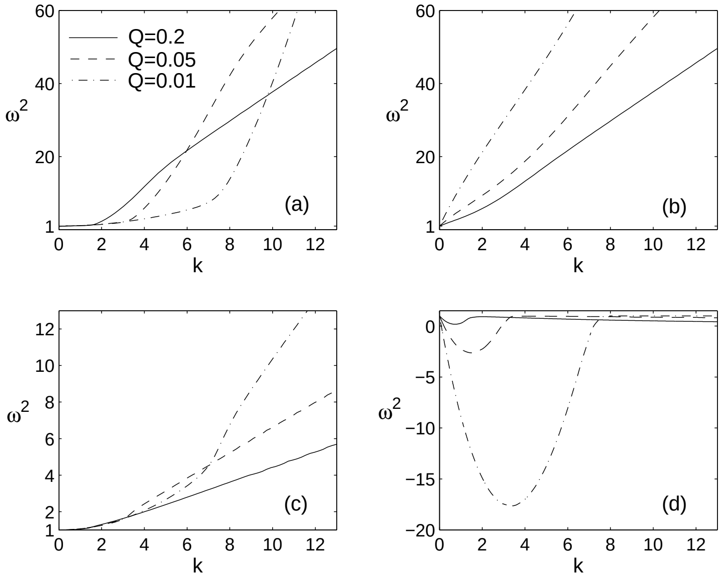

Figure 5.4: Dispersion curves for $s = 5$ and $Q = 0.01$ (dashed-dotted lines), $Q = 0.05$ (dashed lines), $Q = 0.2$ (solid lines). Panels (a),(b) show the even and odd f-modes, respectively. Panel (c) shows the basic even g-mode and panel (d) shows the basic even r-mode. The frequencies of the f- and g-modes never fall below the epicyclic frequency with decreasing Q and therefore these modes are always stable. The basic even r-mode, which is most significantly affected by self-gravity, becomes unstable when the dip on the corresponding dispersion curve crosses the $\omega^2 = 0-$axis with decreasing Q. This dip is deeper and broader at smaller Q.

so it resembles a 2D density wave at such wavelengths (see section 5.4.4). The fact that the characteristic length-scale, λ_m, of gravitational instability turns out to be comparable to the disc thickness may offer a clue why angular momentum transport due to self-gravity tends to be a local phenomenon (e.g., Gammie, 2001; Lodato and Rice, 2004; Boley et al., 2006). In this respect, it should also be mentioned that the analogous gravitational instability of low-frequency modes in a strongly compressed gaseous slab was also found by Lubow and Pringle (1993a). In their analysis, these modes, called neutral modes, are basically degenerate r-modes, since the compressed gaseous slab was not rotating.

5.4.2 Adiabatic stratification

In the adiabatic case (Figs. 5.5a-5.5d), there is no g-mode and the behaviour of the p-, f- and r-modes under self-gravity has qualitatively the same character as in the subadiabatic case above. Again, the basic branch of the even r-mode appears to be most significantly affected by self-gravity (the corresponding dispersion curve in Fig. 5.5c has the same form as that in Fig. 5.3c) and becomes gravitationally unstable. The ω^2 of the p- and f-modes, although

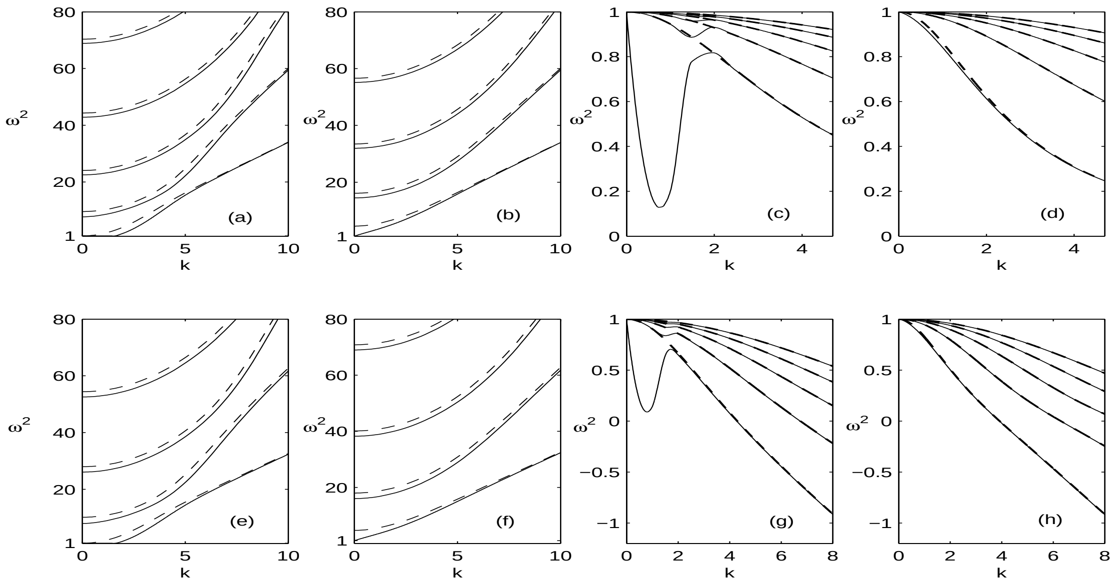

Figure 5.5: Dispersion diagrams for vertical modes in a self-gravitating disc with $Q = 0.2$ for an adiabatic stratification with $s = 2.5$ (panels a-d) and for a superadiabatic stratification with $s = 1.7$ (panels e-h). Shown are (a, e) the even p- and f-modes, (b, f) the odd p- and f-modes, (c) the even r-mode, (d) the odd r-mode, (g) the convectively unstable even g-mode and (h) the odd g-mode. The frequency ordering and mode naming are the same as in Fig. 5.2. As before, dashed curves correspond to the dispersion diagrams calculated without self-gravity for the perturbations. The behaviour of the p-, f- and r-modes in the adiabatic case is similar to that in Fig. 5.3. In the superadiabatic case, the dispersion curves of the even g-mode have dips caused by self-gravity in the interval $0 \leq k \leq 2$, while the odd g-mode is almost unaffected by self-gravity. The largest dip occurs for the basic branch of the even g-mode (denoted above as g_0^e) in panel (g), which appears to be most significantly influenced by self-gravity. For large k and/or large mode numbers, the effect of self-gravity becomes weak and the dispersion curves merge with dashed ones for the non-self-gravitating case. Due to the superadiabatic stratification, the even and odd g-modes are convectively unstable (i.e., have $\omega^2 < 0$) in the range $k > 3.91$, where the influence of self-gravity is small.

lowered by self-gravity, always remain larger or equal to κ^2 irrespective of the value of Q. In this case, $s = 2.5$ and the corresponding $Q_{cr} = 0.177$ and $k_m = 0.81$. Again the radial wavelength of instability $\lambda_m = 2\pi/k_m = 7.75$ is not far from the disc thickness $H = 2h = 2.08$ for these parameters.

5.4.3 Superadiabatic stratification

The superadiabatic case (Figs. 5.5e-5.5h) is interesting, because at $\omega^2 \leq \kappa^2$, as classified above, instead of the r-mode there is the g-mode, which in addition to convective instability can also exhibit gravitational instability. As in the non-self-gravitating case, for radial

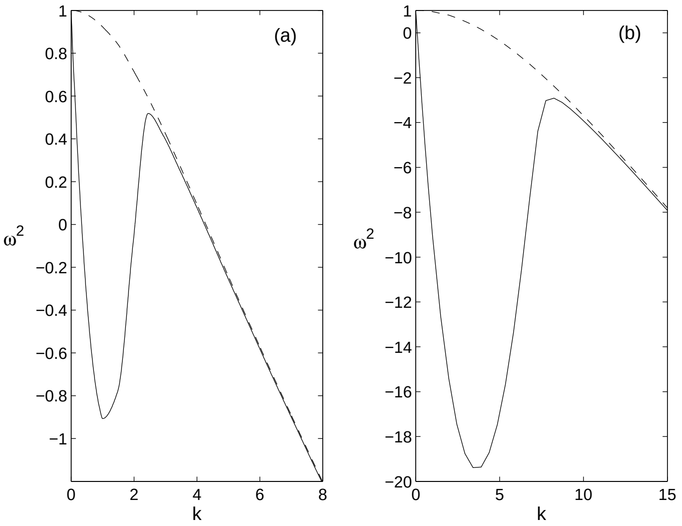

Figure 5.6: Dispersion curves of the basic even g-mode for a superadiabatic stratification with $s = 1.7$ for (a) $Q = 0.1$ and (b) $Q = 0.01$. Dashed lines show the same branches computed without self-gravity for the perturbations. For $Q = 0.1$, the basic even g-mode exhibits gravitational instability in the vicinity of $k_m = 1.15$ and convective instability in the range $k > 4.27$, where the influence of self-gravity is weak. For $Q = 0.01$, the effect of self-gravity is much stronger and the radial length-scales of gravitational and convective instabilities overlap in the range $4.25 \leq k \leq 8.5$.

wavenumbers larger than a certain value, all the branches of the g-mode become unstable because of the negative vertical entropy gradient. From Fig. 5.5g we see that for $0 \leq k \leq 2$, self-gravity produces dips on the g-mode dispersion curves, again preferably for the even parity ones. When Q drops sufficiently, the dip on the basic branch of the even g-mode, which appears to be most affected by self-gravity, starts to cross the $\omega^2 = 0$-axis in a way similar to that of the basic even r-mode branch above and this signals the onset of gravitational instability. For $s = 1.7$, adopted in Fig. 5.5, we find $Q_{cr} = 0.184$, which gives a thickness $H = 2h = 1.76$, and $k_m = 0.83$, so that $\lambda_m = 2\pi/k_m = 7.57 \gtrsim H$. As for the p- and f-modes, they behave under self-gravity in exactly the same manner as in the above cases. In particular, although their frequencies are reduced, they can never become gravitationally unstable even for very small values of Q. Thus, for a superadiabatic stratification, the basic even g-mode can exhibit two types of instabilities: gravitational and convective (see Fig. 5.6). As we see from Figs. 5.5g and 5.6a, at moderate $Q \sim Q_{cr}$, the radial length-scales for the activity of self-gravity and convective instability are well separated, so that these two effects do not interfere with each other in the linear regime. For $Q = 0.1$ in Fig. 5.6a, self-gravity is dominant at $0 \leq k \leq 2.55$, while convective instability occurs at $k > 4.27$, where the effect of self-gravity is weak (a similar situation is for $Q = 0.2$ in Fig. 5.5g). However, for very small

$Q \ll Q_{cr}$, the radial scales of gravitational and convective instabilities overlap (Fig. 5.6b), but the gas motion for such radial scales in the case of very strong gravitational instability would hardly resemble convective motions (it will look more like that in Fig. 5.11 below); convection is simply disrupted by gravitational instability. Therefore, we can conclude that unless a disc is strongly self-gravitating, self-gravity has little influence on the properties of convective motions and on the (Schwarzschild) criterion for the onset of convective instability that is equivalent to $N_0^2 < 0$.

5.4.4 Analogy with the 2D density wave theory

From the three cases considered above, we can conclude that the basic branch of the low-frequency ($\omega^2 \le \kappa^2$) even mode, which is the r-mode, in the case of subadiabatic and adiabatic stratifications, and the g-mode, in the case of a superadiabatic stratification, is most subject to the influence of self-gravity. These modes determine the gravitational instability in vertically stratified discs. As we have clearly seen in all the above cases, the radial wavelength of the instability is comparable, though a bit larger than the disc thickness, $\lambda_m \gtrsim H$, which in turn implies that a 2D analysis of the gravitational instability is at most marginally valid and the rigorous stability (linear) analysis of self-gravitating discs must be three-dimensional. However, despite the 2D treatment only being marginal, we can still find similarities between our results and the 2D density wave theory in thin self-gravitating discs, which actually appears doing a decent qualitative job.

Consider an equivalent zero-thickness disc with the sound speed c_{sm}[4] and the surface density $\Sigma_0 = \int_{-h}^{h} \rho_0 dz$. Then the well-known dispersion relation of axisymmetric 2D density waves in this thin self-gravitating disc is given by Eq. (2.5) (see also Goldreich and Tremaine, 1978; Binney and Tremaine, 1987):

$$\omega^2 = k^2 - \frac{2}{Q_{2D}} k + 1,$$

where we have used the same normalisation as before and because of that $Q_{2D} = c_{sm}\Omega/\pi G\Sigma_0$. This dispersion relation is a parabola with a minimum at the Jeans wavenumber $k_J = 1/Q_{2D}$. This is the wavenumber, at which the effect of self-gravity is most prominent, and if $Q_{2D} < 1$ it gives the characteristic length-scale of gravitational instability. At small $k \ll k_J$, 2D density waves are dominated by self-gravity and inertial forces, while at large $k \gg k_J$, pressure/compressibility dominates over self-gravity and density waves appear as an acoustic

---

[4] In fact, there are other options for choosing the sound speed as a some kind of height-averaged value. We address this question in section 5.6, but for the present purpose this does not make much difference.

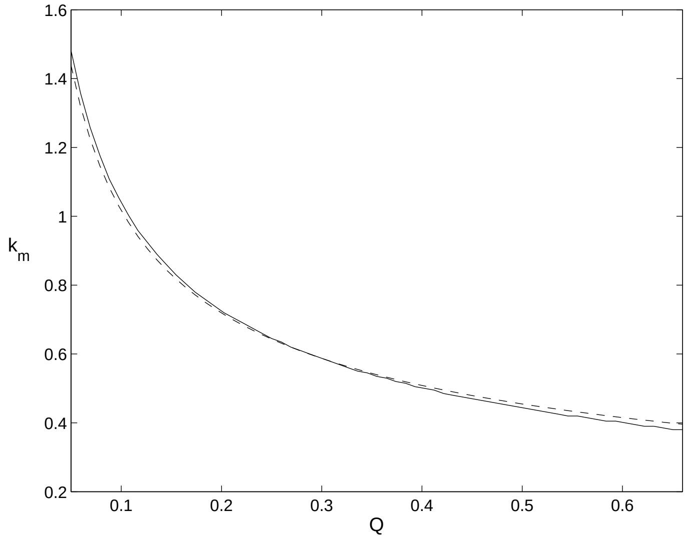

Figure 5.7: Dependence of k_m on Q for $s = 5$, which very closely follows the power law $Q^{-1/2}$ (shown with dashed line and scaled appropriately). The gravitational instability sets in at $Q = Q_{cr} = 0.168$ that corresponds to $k_m = 0.8$.

mode. At $k \sim k_J$ all three factors can be important. Let us now look at the dispersion curves of the basic branches of the even r- and convectively unstable even g-modes in Figs. 5.3c, 5.5c and 5.5g. They have a similar parabolic shape in the self-gravity and compressibility dominated regime at $k \sim k_m$ with the linear phase at smaller $k \ll k_m$, where only self-gravity and inertial forces play a role. This linear phase at long wavelengths is well reproduced by the 2D dispersion relation. Therefore, we can identify the wavenumber k_m, at which the effect of self-gravity on the 3D modes is largest, with the Jeans wavenumber k_J. In some sense, in the case of instability when $Q < Q_{cr}$, k_m gives a more accurate value for the radial wavelength, λ_m, of the gravitationally most unstable mode than that given by the Jeans wavenumber k_J in the 2D model. For example, from Fig. 5.12 we find that the vertical structure (ρ_0) calculated at $s = 5$ and $Q_{cr} = 0.168$ gives the corresponding $Q_{2D} = 0.76$ and $k_J = 1.32$, which differs from $k_m = 0.8$ found above for these parameters. Moreover, we will see in section 5.6 that the criterion for gravitational instability based on the 3D calculations, i.e. $Q < Q_{cr}$, is more exact than the 2D criterion $Q_{2D} < 1$.

In the 2D case, as shown above, the Jeans wavenumber $k_J = 1/Q_{2D}$, implying that it is determined solely by competition between self-gravity and compressibility/pressure. It is now interesting to see how an analogous characteristic wavenumber k_m depends on Q in the 3D case. From Fig. 5.7 we see that this dependence has a power law character $Q^{-1/2}$, which means that the value of k_m is again set by self-gravity and compressibility, as in the 2D case, so that the disc rotation plays a role only in determining ω^2 but not k_m. Indeed, the only length-scale that can be constructed from the sound speed c_{sm}, density ρ_m and gravitational constant G without the angular velocity, Ω, of disc rotation is $c_{sm}/\sqrt{G\rho_m}$ (analogous to the Jeans length of a collapsing 3D cloud). If we normalise it by c_{sm}/Ω, we get $2\sqrt{\pi}Q^{1/2}$, giving the above power law dependence for the corresponding non-dimensional wavenumber $\sqrt{\pi}Q^{-1/2}$ to which k_m is proportional.

Thus, the basic even r-mode or, in the case of a superadiabatic stratification, the basic even g-mode at $k \lesssim k_m$ exhibit many of the characteristics of 2D density waves and could be regarded as their 3D analogues at such radial wavenumbers (see also section 5.5). If continued to $k \gg k_m$, the density wave mode would thus connect up to the large-k parts of these two modes dominated, respectively, by inertial forces and by negative buoyancy, instead of merging with the compressible p-mode, as might seem at first sight from the 2D dispersion relation. (Although, in non-self-gravitating discs, as noted in chapter 4, the 2D density wave mode can be related to the basic p-mode.) However, the limit $k \gg k_m$, equivalent to $\lambda \ll H$, means the breakdown of the 2D approximation and, therefore, the concept of 2D density waves at $k \gg k_m$ would not be well-defined.

5.5 Gravitational instability in 3D: properties of the basic branch of the low-frequency even r-mode

In this section we concentrate only on the properties of the gravitationally unstable basic even r-mode. As we have seen, the behaviour of the convectively unstable basic even g-mode under self-gravity has a qualitatively similar character.

5.5.1 Effect of compressibility

As noted above, in the Poisson's equation, the density perturbation giving rise to the gravitational potential perturbation can be due to compressibility and due to the stratification of the equilibrium vertical structure. As is evident from Fig. 5.2c, in non-self-gravitating discs, the r-mode is nearly incompressible. So, it is interesting to see if it still remains incompressible in self-gravitating discs and which out of these two sources of density perturbations is ultimately responsible for its gravitational instability. In order to explore this, we again computed the dispersion curve of the basic even r-mode in the incompressible limit of Eqs. (5.16-5.18):

$$\frac{du_z}{dz} = \frac{k^2}{\omega^2 - \kappa^2}(p + \psi) \qquad (5.23)$$

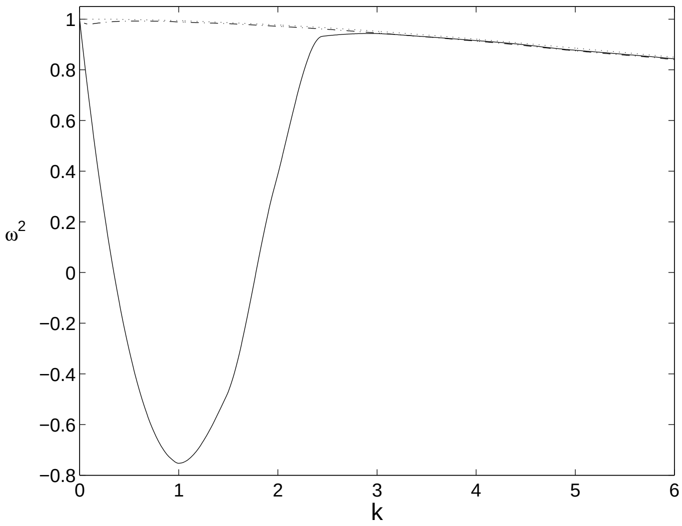

Figure 5.8: Dispersion curves of the basic even r-mode at $s = 5, Q = 0.1$ with corresponding $h = 1.22$. Dotted line corresponds to the non-self-gravitating case, dashed-dotted – to the incompressible case with self-gravity and solid line – to the self-gravitating case with compressibility. The influence of self-gravity on the basic even r-mode is very weak in the incompressible limit and the dashed-dotted dispersion curve almost coincides with the dotted one in the non-self-gravitating limit; a slight deviation (dip) at small k is due to the density perturbation arising from the stratified background.

$$\frac{dp}{dz} = \frac{N_0^2}{g_0}p + (\omega^2 - N_0^2)u_z - \frac{d\psi}{dz} \tag{5.24}$$

$$\frac{d^2\psi}{dz^2} - k^2\psi = \frac{\rho_0}{Q}\frac{N_0^2}{g_0}u_z. \tag{5.25}$$

supplemented with the same boundary conditions. Notice that on the right hand side of Poisson's equation (5.25), only the density perturbation due to stratification is left in the absence of compressibility. In Fig. 5.8, we compare the dispersion curves of the basic even r-mode obtained in the incompressible limit to those computed for the compressible self-gravitating and non-self-gravitating cases. It is clear that in self-gravitating discs, the basic even r-mode becomes compressible at $k \sim k_m$ thereby providing density perturbation for the gravitational potential and, therefore, accounting for the large dip on the dispersion curve. The dispersion curve in the incompressible limit deviates only very slightly from that in the non-self-gravitating case, because of the second source of density perturbation, i.e. stratification, which turns out to be much smaller than that associated with compressibility. Such a comparison also shows that the primary cause of gravitational instability in incompressible discs, as described in GLB and Lubow and Pringle (1993a), is the displacements of free-surfaces. In our case, the equilibrium density vanishes at the surface, so that the effect of surface displacements on the growth rate of instability is null (see boundary condition 5.22)

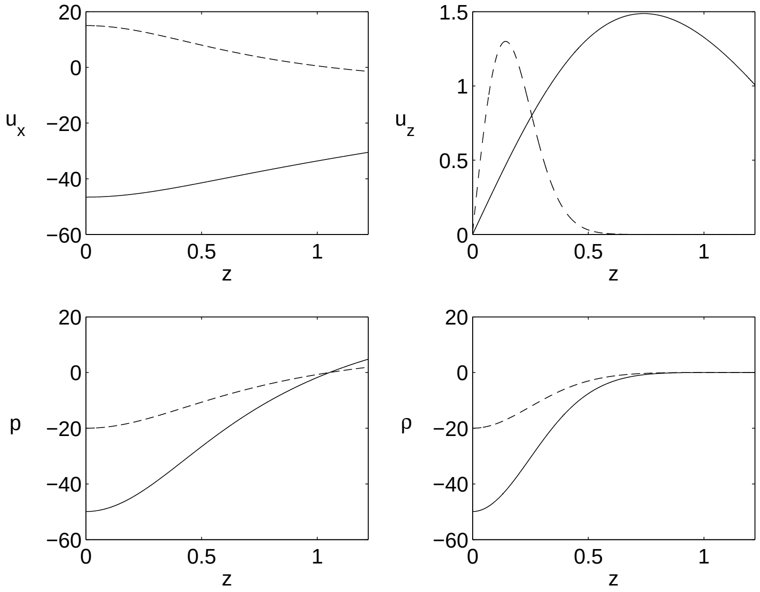

Figure 5.9: Vertical structures of the radial velocity u_x, the vertical velocity u_z, the pressure p and the density ρ perturbations constituting the eigenfunctions of the gravitationally unstable basic even r-mode for $s = 5, Q = 0.1, h = 1.22, k_m = 1$ with the corresponding largest growth rate $\omega^2_{min} = -0.75$. Dashed lines show the eigenfunctions (scaled arbitrarily for plotting purposes) in the non-self-gravitating case for the same parameters except $\omega^2 = 0.99$. Notice that in the non-self-gravitating case, the eigenfunctions are more concentrated near the midplane, whereas in the self-gravitating case they vary over the whole vertical extent.

and in the incompressible case only density perturbation due to stratification, as shown, is hardly sufficient for gravitational instability.

5.5.2 Eigenfunctions and spatial structure

Here we compute the vertical structure of the eigenfunctions of the gravitationally unstable basic even r-mode and also find what type of motions it induces. We take $Q = 0.1$ and $s = 5$, which gives $h = 1.22$ for the disc height. For these parameters, the dispersion curve of this mode in Fig. 5.8 has a minimum $\omega^2_{min} = -0.75$, i.e. gives the largest growth rate of gravitational instability, at $k_m = 1$. In Fig. 5.9, we plot the corresponding eigenfunctions only in the upper half of the disc's full vertical extent. Because the mode has even parity, the vertical velocity and the derivative of the gravitational potential are odd functions and, correspondingly, the pressure and potential are even functions of z. Only the pressure perturbation has one node in the interval $0 < z \leq h$, other quantities have no nodes and vary over the whole vertical extent. This demonstrates that the spatial structure of the gravitationally most unstable mode is three-dimensional with non-zero vertical velocity. To see how

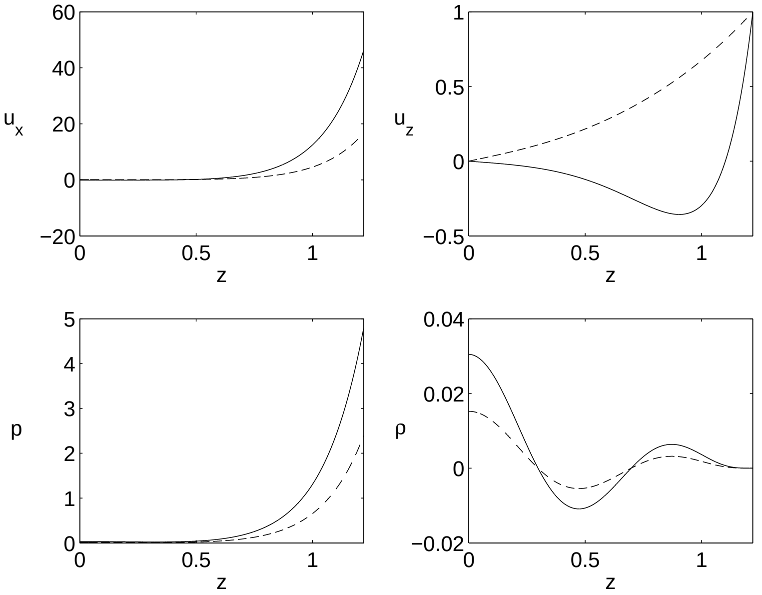

Figure 5.10: Eigenfunctions of the even f-mode with the same values of parameters as in Fig. 5.9. Unlike the basic even r-mode, the even f-mode is gravitationally stable. The corresponding eigenfrequencies at $k_m = 1$ are $\omega^2 = 1.1$ for the self-gravitating case (solid lines) and $\omega^2 = 1.6$ for the non-self-gravitating case (dashed lines). Again, non-self-gravitating eigenfunctions have been scaled arbitrarily for plotting purposes.

self-gravity changes the form of the eigenfunctions, we also show the eigenfunctions of the same branch in the non-self-gravitating limit, which appear to be more concentrated towards the midplane than the self-gravitating ones and have no nodes in the same vertical range.

The f-mode was shown to play an important role in the angular momentum and energy transport in stratified discs (see LO98). In Fig. 5.10 we also plot its eigenfunctions in the presence of self-gravity for the same parameters. Comparing Figs. 5.9 and 5.10, we clearly see that self-gravity modifies the form of the eigenfunctions of both the f- and r-modes. The perturbed quantities for both mode types vary with height somewhat similarly, especially the vertical velocities, which are, however, still at least an order of magnitude less than the horizontal ones, so that the motions can be viewed, to leading order, as 2D (see also Fig. 5.11). Thus, the gas motion associated with the gravitationally unstable r-mode is of similar type to that of the f-mode, which, in turn, implies that the r-mode might be as important as the f-mode, i.e., might be able to do similar 'dynamical jobs' as the f-mode in self-gravitating discs. As mentioned earlier, the non-linear behaviour of 3D perturbations in self-gravitating discs has been attributed solely to the f-mode dynamics without analysing other modes in the disc (Pickett et al., 2000; Boley and Durisen, 2006).

The perturbed quantities have the form $f(z)\exp(ik_m x), f \equiv (u_x, u_z, p, \rho)$, from which, by

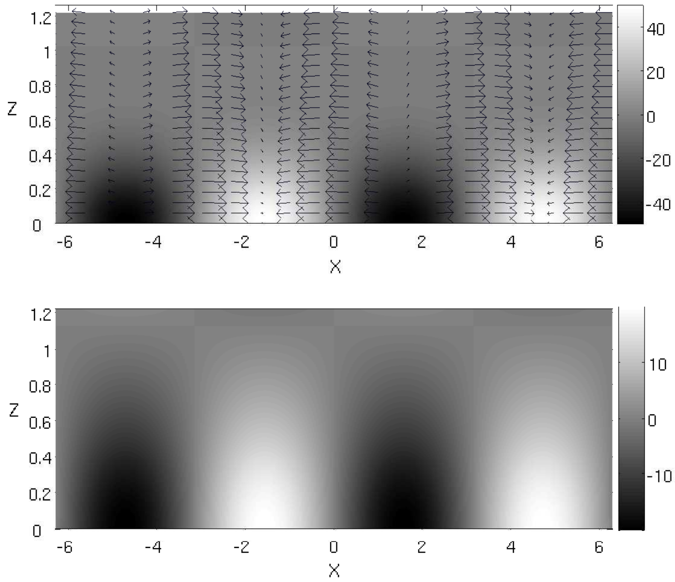

Figure 5.11: Distribution of the density (upper plot) and temperature (lower plot) perturbations in the (x,z)-plane constructed from the eigenfunctions of the gravitationally unstable basic even r-mode in Fig. 5.9. Superimposed on the density field are the velocity vectors of the induced gas flow.

taking the real parts, we can construct the spatial picture of the velocity, density and temperature perturbations for the gravitationally unstable basic even r-mode. The temperature perturbation is found from the ideal gas equation of state and is equal to $T = -c_s^2\rho/\rho_0 + \gamma p/\rho_0$ (here the pressure perturbation p is as used in the original Eqs. 5.10-5.15, i.e., before switching to new variables). Figure 5.11 shows the density and temperature fields constructed in this way with velocity vectors showing the direction of gas motion superimposed on the density field, which in fact resembles the classical density profile of a 2D density wave. Near the midplane, matter flows (converges), almost parallel to the $x-$axis, towards high density regions. With increasing z, the flow becomes more arc-like, because of the variation of the vertical velocity with height. An analogous velocity field in the (x,z) plane was also observed in some non-linear simulations of self-gravitating discs (Boley and Durisen, 2006; Boley et al., 2006), which suggests that the latter may be a result of the non-linear development of the type of motion we see here in the linear regime. We might also speculate that in relatively high-mass discs, the non-axisymmetric gravitationally unstable basic r-mode will produce similar high-density regions, which can be precursors of spiral shock fronts (if the disc does not fragment), and arc-like motions in the upper layers can then give rise to non-linear shock bores producing large distortions of the disc surface, as described by Boley and Durisen (2006). The temperature perturbations are fairly non-uniform as well, varying both vertically and radially on comparable length-scales. Obviously, larger temperature perturbations correspond to overdense regions that are contracting because matter flows into them, and lower – to underdense regions.

5.6 Stability criteria in 2D and 3D

In a 3D disc, the degree and effect of self-gravity is characterised by $Q_{3D} = \Omega^2/4\pi G\rho_m$, where ρ_m is the value of equilibrium density at the midplane (see also GLB), which plays here a role similar to that of the standard 2D Toomre's parameter $Q_{2D} = c_s\Omega/\pi G\Sigma$ (Toomre, 1964). However, in three-dimensional simulations, researchers often still tend to use the 2D Toomre's parameter to characterise the onset of gravitational instability in discs (Rice et al., 2003, 2005; Lodato and Rice, 2004, 2005; Boley et al., 2006). In this section, we investigate how a more general and exact 3D criterion of gravitational instability can be related to the 2D one. In general, these two parameters are different, because the 2D parameter contains the sound speed and surface density, which by definition do not vary vertically for razor-thin discs, whereas the 3D parameter does not contain the sound speed and surface density, but the midplane value of the density. So, if we still wish to characterise the 3D instability in terms of the 2D Toomre's parameter, we need to use some height-averaged, or effective sound speed (surface density can be uniquely calculated from the vertical density distribution). In some simulations the midplane values (Mejía et al., 2005; Boley et al., 2006), while in others vertically averaged values (Rice et al., 2003, 2005) of the sound speed are used. Here we calculate the corresponding critical $Q_{cr,2D}$ for these cases.

As is well-known, in razor-thin 2D discs, axisymmetric perturbations (density waves) are gravitationally unstable if $Q_{2D} < 1$ (Toomre, 1964; Goldreich and Tremaine, 1978). Let us now find the critical $Q_{cr,3D}$ that determines the gravitational instability in 3D, which, as found in this study, is associated with the basic branches of the even parity low-frequency modes. In Fig. 5.12, we show the critical $Q_{cr,3D}$ as a function of the polytropic index s. Given $Q_{cr,3D}$, we can now calculate $Q_{1\ cr,2D}$ defined in terms of the midplane sound speed and $Q_{2\ cr,2D}$ defined in terms of the vertically averaged sound speed:

$$Q_{1\ cr,2D} = \frac{2Q_{cr,3D}}{\int_0^h \rho_0 dz}, \qquad Q_{2\ cr,2D} = \frac{2Q_{cr,3D}\int_0^h c_s dz}{h\int_0^h \rho_0 dz},$$

where ρ_0 and c_s, as before, are the normalised equilibrium density and sound speed the z-dependence of which, as well as the value of h, are determined by $Q_{cr,3D}$ and the polytropic index s. So, the critical $Q_{1\ cr,2D}$ and $Q_{2\ cr,2D}$ calculated with the two different methods are

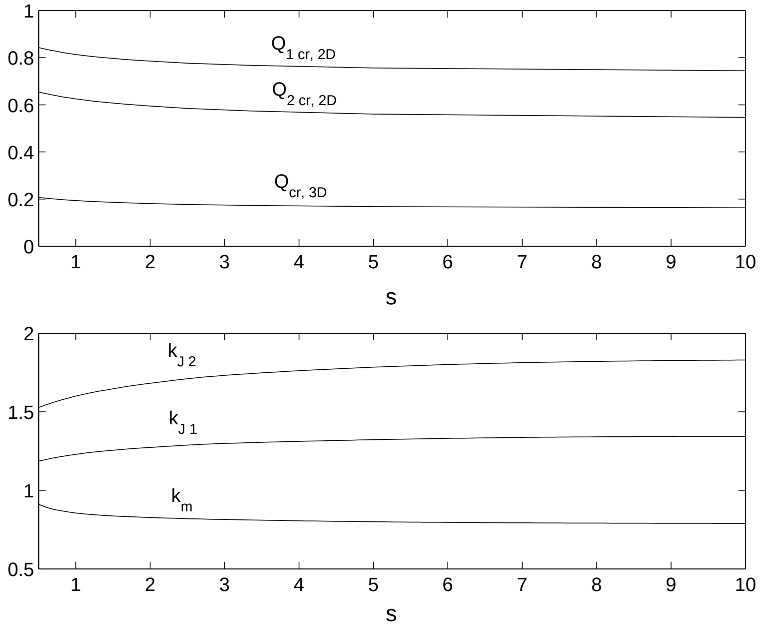

Figure 5.12: Dependence of the critical $Q_{cr,3D}$ and the corresponding critical $Q_{1\ cr,2D}$ and $Q_{2\ cr,2D}$ on the polytropic index s. Shown also are the k_m at $Q_{cr,3D}$ and the 2D Jeans wavenumbers $k_{J1} = 1/Q_{1\ cr,2D}$, $k_{J2} = 1/Q_{2\ cr,2D}$ as a function of s.

different, though not far from each other, and less than unity which is their critical value in the 2D case. This again confirms the well-known result that 3D discs are more stable and to make them unstable smaller value of Q_{2D} is necessary compared with that required for instability of razor-thin 2D discs (see also GLB). But in such cases, using Q_{3D} seems more appropriate as it does not involve uncertainties in how to average the sound speed over height.

In Fig. 5.12 we also examine the dependence of the critical wavenumbers of the gravitational instability on s. We see that the 2D Jeans wavenumbers are always larger than the actual more exact wavenumber k_m at the onset of gravitational instability when $Q = Q_{cr,3D}$. Thus, in the 3D case, the critical wavenumbers depend differently on the thermodynamics than in the 2D case, where the critical $k_J = 1/Q_{2D} = 1$ at the onset of instability.

5.7 Summary and discussions

In this chapter, we have analysed the axisymmetric normal modes of perturbations in stratified, compressible, self-gravitating gaseous discs with subadiabatic, adiabatic and superadiabatic vertical stratifications. First, we performed a classification of perturbation modes in stratified discs in the absence of self-gravity to compare with previous studies. Four well-known main types of modes can be distinguished: the acoustic p-mode, the surface gravity f-mode, the buoyancy g-mode and the inertial r-mode. The restoring forces for these modes for large radial wavenumbers are mainly provided by one the following: pressure/compressibility, displacements of the disc surface, buoyancy and inertial forces due to disc rotation for the p-,f-,g- and r-modes, respectively. For smaller wavenumbers, the restoring force for each mode is a combination of these forces. In the case of adiabatic stratification, buoyancy is zero and, therefore, the g-mode disappears, while other modes remain qualitatively unchanged. For superadiabatic stratification, the g-mode becomes convectively unstable and merges with the r-mode, so that only a single convectively unstable mode appears in the dispersion diagram at $\omega^2 \leq \kappa^2$, which we still call the g-mode. Due to the reflection symmetry of the equilibrium vertical structure with respect to the midplane, each mode comes in even and odd pairs. By our terminology, for even (odd) modes, pressure and gravitational potential perturbations are even (odd), while the perturbations of vertical velocity and derivative of potential are odd (even) functions of the vertical coordinate. After classifying and characterising modes without self-gravity, we included self-gravity in the perturbation equations and investigated how it modifies the properties of these modes. We found that self-gravity, to some extent, reduces the frequencies of all normal modes at radial wavelengths comparable to the disc height, but its influence on the basic even r-mode, in the case of subadiabatic and adiabatic stratifications, and on the basic even g-mode, in the case of superadiabatic stratification, appears to be strongest. With decreasing Q_{3D}, these modes become unstable due to self-gravity and thus determine the gravitational instability of a vertically stratified disc. The basic even g-mode also exhibits convective instability due to a negative entropy gradient but, unless the disc is strongly self-gravitating, these two instabilities grow concurrently in the linear regime, because their corresponding radial length-scales are separated. We also obtained the corresponding criterion for the onset of gravitational instability in 3D, which is more exact than the standard instability criterion in terms of the 2D Toomre's parameter, $Q_{2D} < 1$, for axisymmetric density waves in razor-thin discs. By contrast, the p-, f- and convectively stable g-modes have their ω^2 reduced by self-gravity, but never become unstable for however small value of Q_{3D}. This is a consequence of the three-dimensionality of the disc. The eigenfunctions associated with the gravitationally unstable modes are intrinsically three-dimensional, that is, have non-zero vertical velocity and all perturbed quantities vary over the whole vertical extent of the disc. In this regard, resolving the gravitationally most unstable length-scale in numerical simulations thus reduces to properly resolving the disc height (together with resolving the corresponding radial scale, which, as shown here, appears somewhat larger than the disc height). So, the criterion of Nelson (2006) that at least ~ 4 particle smoothing lengths should fit into per scale height may apply in three-dimensional SPH simulations. This implies a substantially larger number of SPH particles per vertical column because the disc itself may extend over many scale heights. He also shows that a similar criterion applies to grid-based simulations.

Here for simplicity, and also to gain the first insight into the effects of self-gravity on the vertical modes in stratified discs, we considered only axisymmetric perturbations. Non-axisymmetric perturbations are dynamically richer, though more complicated, because the phenomena induced by Keplerian shear – strong transient amplification and linear coupling of modes (not to be confused with non-linear mode-mode interactions) – come into play for this type of perturbations. Transient (swing) amplification of perturbations (density waves) has been studied previously in razor-thin (2D) approximation, ignoring disc's vertical structure (Goldreich and Lynden-Bell, 1965b; Goldreich and Tremaine, 1978; Toomre, 1981; Kim and Ostriker, 2001; Mamatsashvili and Chagelishvili, 2007). From the 3D analysis presented here, we may expect that in the linear regime, the non-axisymmetric basic even r-mode can undergo larger transient amplification due to self-gravity than other modes in the disc. This transient amplification of perturbations may be relevant for explaining the large burst phases seen in numerical simulations at the initial stages of the development of gravitational instability in discs (Rice et al., 2003, 2005; Lodato and Rice, 2004, 2005; Boley et al., 2006). So, in this respect one should analyse and quantify the transient amplification of non-axisymmetric perturbations in stratified 3D self-gravitating discs starting with linear theory, which represents one of my future research plans. As for the linear coupling of modes, so far it has been mainly studied in razor-thin 2D discs; in 3D discs it has only been treated using simplifying assumptions (e.g., constant-g approximation, see Tevzadze 2006; Tevzadze et al. 2008 and also chapter 4) that do not fully bring out all the diversity of modes existing in 3D discs. There are a larger number of modes in 3D than that in 2D discs and it is quite possible that some of them may appear linearly coupled due to shear and therefore be able to generate each other during evolution, especially the f-mode and the basic branch of the r-mode (because they vary on comparable vertical scales in the presence of self-gravity, see Figs. 5.9 and 5.10). Another related problem also of interest in this context is the interaction between self-gravity and the MRI in magnetised discs (see e.g., Fromang, 2005). In particular, how self-gravity can modify the growth of magnetic normal modes responsible for the MRI. Actually, this will be the generalisation of the extensive analysis of normal modes in magnetised discs by Ogilvie (1998).

LO98 showed that in non-self-gravitating discs with polytropic vertical stratification, an external forcing preferentially excites the f-mode, because it exhibits the largest responsiveness to an external driving compared to other modes. This mode, propagating through a disc, results in energy dissipation near the disc surface. Based on these results and partly on the properties of f-modes in stellar dynamics, Pickett et al. (2000) identified the behaviour of 3D perturbations in self-gravitating discs involving large surface distortions and the resulting energy dissipation in the upper layers, with f-mode dynamics. The self-stimulated potential was thought to play the role of an external/tidal force. However, it is not obvious that the effect of a self-stimulated potential is the same as that of the external potential. In fact, our analysis has revealed that in self-gravitating discs, in addition to the f-mode, the r-mode can also be dynamically important, because this mode appears to be subject to gravitational instability, while the f-mode is not. The eigenfunctions of the gravitationally unstable basic even r-mode differ from those of the r-mode in non-self-gravitating discs in that they are no longer concentrated near the midplane and behave somewhat similarly to the eigenfunctions of the f-mode: they vary over the whole vertical extent of the disc and also involve noticeable perturbations of the disc surface. Consequently, like the f-mode, the gravitationally unstable r-mode can, in principle, also induce gas motion causing large surface distortions and resultant energy dissipation in the upper layers of the disc, which is thought to play a role in enhancing disc cooling (because the energy is deposited in the upper layers with small optical depth, it can be radiated away more quickly and effectively cool the disc, but this is a subject of further study, see e.g., Johnson and Gammie, 2003; Boley et al., 2006). Furthermore, in the case of non-axisymmetric perturbations, as mentioned above, because of shear, the gravitationally unstable r-mode can couple with and generate the strong f-mode. So, the surface distortions may be caused by a combination of the f- and r-modes. In order to explore where dissipation can predominantly occur in a self-gravitating disc, one needs to generalise the analysis of Lin et al. (1990), LP, LO98, Bate et al. (2002) on the propagation of waves in stratified non-self-gravitating discs and consider the propagation of non-axisymmetric modes in stratified self-gravitating discs, which I also plan to look into in the future as the next step of the present study. The dispersion properties of modes in the presence of self-gravity as found here are one of the necessary things for studying mode propagation.

Another point we want to raise concerns the spatial distribution of temperature. In order to realistically model the cooling of protoplanetary discs, Boley et al. (2007) employed the radiative transfer. In the vertical $z-$direction, the ray equation was solved exactly, approximating the disc vertically as a plane-parallel atmosphere, but in the radial direction only the radiation diffusion approximation was used. However, as our linear results (Fig. 5.11) and other non-linear simulations (e.g., Mejía et al., 2005; Boley et al., 2006) show, temperature and, therefore, opacity may vary on comparable scales in both the radial and vertical directions and have very complex structure in the non-linear regime. This implies that a more general radiative transfer treatment based on solving the ray equation in all directions, rather than using the diffusion approximation in either direction, would be more appropriate for a better understanding of cooling processes.

Chapter 6

Summary and future work

In this thesis, we have studied new aspects of the dynamics of perturbation modes in protoplanetary discs brought about by velocity shear of the disc's differential rotation and by disc self-gravity. We have concentrated on non-magnetised discs and considered three basic modes/types of perturbations – spiral density waves, vortices and vertical convection – that are primarily responsible for angular momentum transport and also contribute to planet formation process in such discs. Previously, these modes had been analysed mainly separately in different settings. Spiral density waves were studied in the context of self-gravitating thin discs, because they can be (modally) excited by gravitational instability and are able to drive angular momentum and mass transport through the disc. On the other hand, vortices and convection were commonly studied in non-self-gravitating discs. In this thesis, we have demonstrated that the shear of disc flow results in these modes becoming dynamically coupled and generating each other in the course of evolution, so they must be considered concurrently in order to gain a proper understanding of disc dynamics and evolution. This shear-induced mode coupling can also have important implications for other dynamical processes in discs, such as planet formation and planet migration. In this final chapter, we summarise our main results and outline possible future developments.

In chapter 3, we analysed the linear and non-linear dynamics of vortices in razor-thin (2D), compressible, self-gravitating gaseous discs with Keplerian rotation, using the shearing sheet model. First, in the linear regime, we showed that the vortical mode, similar to density waves, is also subject to the influence of self-gravity and as a result can experience substantial transient-exponential growth. A comparative analysis of the growth rates of the vortical and density wave modes has shown that the former grows due to self-gravity 2-3 times more than the latter, implying that the role of vortices can be as important as density waves in the overall dynamics of self-gravitating discs. This is noteworthy, because, as noted, the perturbation type commonly considered in most studies of self-gravitating disc dynamics is the density wave mode alone. We also demonstrated that vortices can generate density waves due to the shear of disc flow. The efficiency of this wave generation process increases with decreasing Toomre's Q, that is, it is more powerful in the presence of self-gravity. Density waves emitted by vortices turn into shocks afterwards, which propagate through the disc. Thus, vortices appear as a new source of density waves and thus as participants in shock manifestation processes in discs. Previous compressible simulations of vortices in non-self-gravitating discs showed that density waves (shocks) excited by vortices tend to considerably enhance angular momentum transport rates due solely to vortices (Johnson and Gammie, 2005b).

After analysing the linear dynamics of vortices, we followed their non-linear evolution under the influence of disc self-gravity using numerical simulations. The primary aim was to generally understand the effects of self-gravity on vortex development and, in particular, whether the scenario of planetesimal formation in vortex cores (Barge and Sommeria, 1995; Klahr and Bodenheimer, 2006), proposed originally for non-self-gravitating discs, can be at work in self-gravitating ones too. This could be an alternative to planet formation via disc fragmentation resulting from gravitational instability, which seems to be unlikely and not favoured in the literature (see e.g., Durisen et al., 2007). We found that in self-gravitating discs, vortices are no longer well-organised and long-lived structures as opposed to those occurring in non-self-gravitating discs. Instead, they are transient structures undergoing recurring phases of formation, growth and destruction lasting for a few orbital periods. Vortex growth in size is limited to scales comparable to the local Jeans scale, whereas in non-self-gravitating discs they can grow to sizes larger than or comparable to the disc scale height due to the inverse (i.e., towards smaller wavenumbers) cascade of energy in 2D geometry. In other words, our study thus showed that self-gravity opposes this inverse cascade, as also observed in other simulations of self-gravitating discs (Lyra et al., 2009). The coupling of spiral density waves and vortices is also at work in the non-linear regime. We showed that for each spatial Fourier harmonic it occurs qualitatively similarly to that in the linear regime. Anticyclonic vortices produce overdense/overpressure regions in the density field (cyclonic vortices quickly get sheared out), which may potentially trap dust particles, but because of the transient nature of vortices, this process may not be efficient.

Future developments:

- Investigate a coupled system – dust particles embedded in a self-gravitating 2D gaseous disc, taking into account particle self-gravity as well. The aim will be to see whether particles get trapped effectively within such irregular and short-lived transient vortices/overpressure/overdensity regions and, more importantly, if they can stay bound owing to their self-gravity and do not dissolve with the parent vortex. This will allow us to understand if the process of planetesimal formation via vortex trapping could be initiated at the early stages of disc lifetime, that is, during embedded and early T Tauri phases.
- In 2D discs, it is also interesting to investigate the influence of self-gravity on the development of the baroclinic and Rossby wave instabilities. These instabilities, generating potential vorticity, lead to the formation of large-scale organised vortices, while as we demonstrated here self-gravity tends to oppose this process. So, the aim will be to see which out of these two factors prevails and how the properties of resulting vortices differ from those in non-self-gravitating baroclinic discs.
- Consider a more general problem of formation, evolution and survival of vortices in 3D vertically stratified self-gravitating discs. The presence of stratification somewhat helps vortices to withstand the destructive effect of the elliptical instability in 3D (Barranco and Marcus, 2005; Lesur and Papaloizou, 2009). On the other hand, as we found, self-gravity does not favour long-lived vortices. But in the 3D case, the effect of self-gravity is somewhat reduced compared with that in the 2D (see chapter 5). Thus, in the 3D case, there are two competing factors – stratification and self-gravity – and it remains to be seen in future numerical experiments which of these two prevails. Another question of interest in stratified discs (either self-gravitating or not) is, as also pointed out by Johnson and Gammie (2005b), if long-lived off-midplane vortices can be generated from an initial random distribution of vorticity, as happens in the 2D case, rather than from specially chosen vortex solutions as in simulations of Barranco and Marcus (2005).

In chapter 4, we considered 3D compressible, non-self-gravitating Keplerian discs with a superadiabatic vertical stratification and investigated another manifestation of a shear-induced linear mode coupling phenomenon – generation of density waves by vertical convection – again within the shearing box formalism. We were motivated by the recent important result from simulations of incompressible convection in protoplanetary discs that vertical convection is in fact able to transport angular momentum outwards in the non-linear regime (Lesur and Ogilvie, 2010; Käpylä et al., 2010), contrary to the belief held previously that convection always transports angular momentum inwards and hence cannot be regarded as a viable mechanism driving disc accretion process. Here, we aimed to go further and examine the role of compressibility, and hence of density waves, in the dynamics of disc convection. We demonstrated that as a result of shear, the non-axisymmetric convective mode is able to excite spiral density waves during evolution, so that, like vortices, it can be regarded as a new source of density waves in discs. The efficiency of wave generation is maximal when both horizontal and vertical wavelengths of modes are comparable to the disc scale height. We also showed that, unlike the non-axisymmetric convective mode, which in the linear regime tends to transport angular momentum inwards, density waves excited by it transport angular momentum outwards. Based on these findings, we suggested that density waves generated by convective mode may aid and enhance the angular momentum transport due solely to convection in the non-linear regime, where the latter becomes outward.

Future developments:

- Generalise the above-mentioned recent simulations of incompressible convection in discs by carrying out non-linear high-resolution simulations of compressible vertical convection in the light of the described here linear phenomenon of generation of density waves by convective motions. The aim will be to see if these density waves are indeed able to sufficiently boost outward angular momentum transport due to convection alone and thus to yield larger values of the corresponding transport parameter α that will match observed disc accretion rates.
- Another point, which needs to be addressed in future simulations of compressible convection is where in the disc density waves generated by convection primarily cause dissipation through shock formation, or in other words, the question of self-sustenance of convection. In all the previous simulations of convection in discs, the latter was not generated self-consistently, but maintained by an artificially imposed heat flux. If the shocks of density waves dissipate near the surface, this may stifle convection, while if – near the midplane, this can then sustain convection.

In chapter 5, we carried out a linear analysis of vertical normal modes of axisymmetric perturbations in stratified self-gravitating gaseous discs. Although axisymmetric modes do not display shear-induced couplings among each other, their analysis provides insight into how gravitational instability develops in the three-dimensional case and what its onset criterion is. The mode dynamics in this case turned out to be more diverse than that in razor-thin 2D discs. Four principal types of modes can be distinguished in a stratified disc (apart from the stationary vortical mode): acoustic p-modes due mainly to compressibility, surface gravity f-modes due to displacements of disc surfaces, buoyancy g-modes due to vertical stratification and inertial r-modes due to rotation. We calculated the dispersion curves and eigenfunctions of these modes in the presence of self-gravity and contrasted them with those in non-self-gravitating discs. Our analysis showed that only the r-mode, in the case of subadiabatic and adiabatic stratifications, and only the g-mode, in the case of superadiabatic stratification (which in addition exhibits convective instability) can become unstable due to self-gravity, so that they determine the onset criterion and nature of the gravitational instability of a vertically stratified disc. These modes also appear as counterparts of 2D density waves in razor-thin self-gravitating discs. By contrast, the p-, f- and convectively stable g-modes, never become unstable due to self-gravity however small the value of 3D Toomre's parameter Q_{3D}. This is a consequence of the three-dimensionality of the disc. The eigenfunctions corresponding to the gravitationally unstable modes are intrinsically three-dimensional. We also compared the more exact instability criterion based on our three-dimensional model with that of density waves in 2D discs. Based on these findings, we discussed the origin of surface distortions seen in numerical simulations of self-gravitating discs. In particular, the three-dimensional behaviour of perturbations in self-gravitating discs was previously attributed solely to the f-mode dynamics (Pickett et al., 2000), whereas, as we found here, the latter mode is gravitationally stable, so that the r- or, in the case of superadiabatic stratification, g-modes, which are subject to gravitational instability, may instead cause similar effects.

Future developments:

- In vertically stratified self-gravitating discs, investigate the linear dynamics of non-axisymmetric p-, f-, g- and r-modes, for which the effects of disc flow shear become important. Specifically, the aim will be to quantify (transient) amplification and possible mutual couplings of these modes under the combined action of self-gravity and shear, as we did in 2D discs in chapter 3. Besides, in order to find where dissipation, and hence heating, can predominantly occur in a self-gravitating disc, it is also interesting to examine the propagation of the non-axisymmetric 3D modes within a self-gravitating disc, based on the analysis of Lin et al. (1990); Lubow and Pringle (1993b); Lubow and Ogilvie (1998); Bate et al. (2002) on the propagation of waves in stratified non-self-gravitating discs.

- Another related problem also of interest in this context is the interaction between self-gravity and the MRI in magnetised discs (see e.g., Fromang, 2005). In particular, the aim will be to understand how self-gravity modifies the properties of magnetic normal modes responsible for the MRI; the latter were comprehensively studied in the absence of self-gravity by Ogilvie (1998).

- Follow the non-linear development of the gravitational instability of 3D modes, described here in the linear regime, through simulations of vertically stratified self-gravitating discs in the shearing box that allows much higher numerical resolution than global disc models can afford. The aim will be: (1) to understand the criterion for fragmentation in 3D, first with constant cooling time and then with more complicated realistic radiative cooling and also how such a criterion differs from that of Gammie (2001) for razor-thin 2D discs (i.e., from Eq. 2.6), (2) to clarify the issue of convective cooling, whether it is really efficient enough to cause disc fragmentation, (3) to characterise the properties of the resulting gravitoturbulent state in the case of no fragmentation.

APPENDIX A

Numerical methods

Here we describe the specific scheme we used for numerically solving basic hydrodynamic and Poisson's equations (3.1-3.5) in the shearing sheet. We start with the description of the numerical method for solving the hydrodynamic equations (3.1-3.3) and the energy equation (3.5), which is based on the famous and widely used ZEUS code (Stone and Norman, 1992), but is specially adapted for the shearing sheet equations (Gammie, 2001; Johnson and Gammie, 2003, 2005b) and employs a different – FARGO scheme (Masset, 2000) for advection due to a large Keplerian velocity. Then we describe the method for solving Poisson's equation (3.4) in razor-thin (2D) discs by means of Fast Fourier Transform (FFT) technique.

As in the original ZEUS algorithm, a staggered grid is used here on which scalar quantities, such as the surface density, pressure, internal energy and gravitational potential are defined at the cell centres, while vector quantities, such as the velocity components are defined at the cell edges (Fig. A.1). Each grid cell with dimensions $\Delta x \times \Delta y = L/N \times L/N$ is characterised by two indices $1 \leq i, j \leq N$ corresponding, respectively, to the x- and y-coordinates, where L and N, as defined in chapter 3, are the size of the simulation domain in the (x, y)-plane and the total number of grid cells in each direction. Such a staggered distribution of quantities is very convenient as it leads to centred-difference equations that are second-order accurate in space (see below). For a grid cell represented by indices (i, j), we denote the coordinates of the left and bottom edges, respectively, as x_i and y_j and the coordinates of the right and top edges as $x_{i+1} = x_i + \Delta x$ and $y_{j+1} = y_j + \Delta y$. Accordingly, the coordinates of the cell

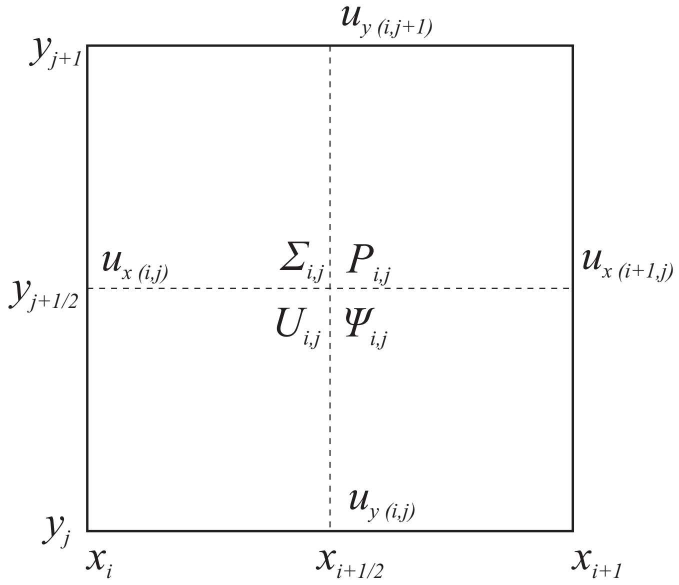

Figure A.1: Schematic diagram illustrating the staggered centred discrete quantities within a grid cell with indices (i,j).

centre are then

$$x_{i+1/2} = 0.5(x_i + x_{i+1}), \qquad y_{j+1/2} = 0.5(y_j + y_{j+1}).$$

Figure A.1 shows the distribution of discretised quantities within this cell. The discrete surface density $\Sigma_{i,j}$, pressure $P_{i,j}$, internal energy $U_{i,j}$ and gravitational potential $\psi_{i,j}$ are defined at the cell centre $(x_{i+1/2}, y_{j+1/2})$. The discrete x-component of velocity, $u_{x\ (i,j)}$, is defined at the middle point of the left edge with the coordinates $(x_i, y_{j+1/2})$ and the discrete y-component, $u_{y\ (i,j)}$, is defined at the middle point of the bottom edge with coordinates $(x_{i+1/2}, y_j)$. Similarly, $u_{x\ (i+1,j)}$ and $u_{y\ (i,j+1)}$ are defined, respectively, at the middle points of the right and top edges with coordinates $(x_{i+1}, y_{j+1/2})$ and $(x_{i+1/2}, y_{j+1})$.

Because the ZEUS code solves equations in a conservative form, we rewrite Eqs. (3.1-3.3) and (3.5) in the latter form (for non-dimensional quantities)

$$\frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma \mathbf{u}) - qx\frac{\partial \Sigma}{\partial y} = 0, \tag{A.1}$$

$$\frac{\partial p_x}{\partial t} + \nabla \cdot (p_x \mathbf{u}) - qx\frac{\partial p_x}{\partial y} = -\frac{\partial P}{\partial x} + 2p_y - \Sigma\frac{\partial \psi}{\partial x}, \tag{A.2}$$

$$\frac{\partial p_y}{\partial t} + \nabla \cdot (p_y \mathbf{u}) - qx\frac{\partial p_y}{\partial y} = -\frac{\partial P}{\partial y} + (q-2)p_x - \Sigma\frac{\partial \psi}{\partial y}. \tag{A.3}$$

$$\frac{\partial U}{\partial t} + \nabla \cdot (U \mathbf{u}) - qx\frac{\partial U}{\partial y} = -P\nabla \cdot \mathbf{u} - \frac{U}{\tau_c}, \tag{A.4}$$

where $p_x = \Sigma u_x$ and $p_y = \Sigma u_y$ are the components of momentum density. The terms on the right hand side of these equations – the pressure, Coriolis and gravity forces in Eqs. (A.2) and

(A.3) and the compressional heating and cooling terms in Eq. (A.4) – are referred to as source terms. The terms on the left hand side containing spatial partial derivatives are referred to as advection terms and describe the evolution of quantities as a result of pure transport (advection) by flow. We have explicitly decomposed the advection of mass, momentum and internal energy by the flow into two parts: the first one (with $\nabla\cdot(..\mathbf{u})$) describes the advection by the perturbed flow velocity $\mathbf{u}$ and the second one (with $-qx\cdot\partial/\partial y$) describes the advection by a much larger background Keplerian velocity $\mathbf{u_0} = (0, -q\Omega x)$.

ZEUS solves equations by an operator-splitting method. Momentum and internal energy are first updated in time by application of the source terms (the continuity equation has no source terms). Then, these updated values, together with surface density, are further updated by application of the advection terms. In each substep, the most recently obtained values of all quantities are used for further update. However, in the present case, the situation is different from that in the standard ZEUS algorithm, because as mentioned above, we also have to do advection by a fast Keplerian flow, which requires a different treatment (see below). So, to emphasise this fact, we have explicitly separated out the corresponding terms in Eqs. (A.1-A.4). Such a successive updating process due to different terms, or operators, is called an operator splitting method. In the following sections, we describe how this is realised in practice by updating the physical quantities between time levels n and $n+1$, i.e., from a moment t_n to $t_{n+1} = t_n + \Delta t$, where Δt is the time step. Partially updated values of the quantities at the end of each substep will be denoted by superscripts $n+a, n+b, n+c$, etc.

A.1 Source terms

We start by updating the velocity and internal energy due to the source terms. Momentum equations (A.2) and (A.3) involving pressure, Coriolis and gravity forces as the source terms, are updated in their non-conservative form

$$\frac{\partial u_x}{\partial t} = -\frac{1}{\Sigma}\frac{\partial P}{\partial x} + 2u_y - \frac{\partial \psi}{\partial x}, \tag{A.5}$$

$$\frac{\partial u_y}{\partial t} = -\frac{1}{\Sigma}\frac{\partial P}{\partial y} + (q-2)u_x - \frac{\partial \psi}{\partial y}, \tag{A.6}$$

where the advection terms have been omitted for the moment, because they are inactive in this substep. Discretising Eqs. (A.5) and (A.6) on our staggered grid, we find the finite difference updates of the velocity components due to the pressure and gravity forces

$$\frac{u^{n+a}_{x\ (i,j)} - u^{n}_{x\ (i,j)}}{\Delta t} = -\frac{2(\gamma-1)(U^n_{i,j} - U^n_{i-1,j})}{\Delta x(\Sigma^n_{i,j} + \Sigma^n_{i-1,j})} - \frac{\psi^n_{i,j} - \psi^n_{i-1,j}}{\Delta x}$$

$$\frac{u^{n+a}_{y\ (i,j)} - u^{n}_{y\ (i,j)}}{\Delta t} = -\frac{2(\gamma-1)(U^n_{i,j} - U^n_{i,j-1})}{\Delta y(\Sigma^n_{i,j} + \Sigma^n_{i,j-1})} - \frac{\psi^n_{i,j} - \psi^n_{i,j-1}}{\Delta y},$$

where we have used the equation of state $P_{i,j} = (\gamma - 1)U_{i,j}$ to replace the pressure with the internal energy. Because the velocity components $u_{x\ (i,j)}$ and $u_{y\ (i,j)}$ are defined, respectively, on the left and on the lower edges of the cell and the surface density, internal energy and gravitational potential are defined at the cell centres, the discrete spatial derivatives (gradients) as well as the average values of densities in two adjacent cells in these finite difference equations are all spatially centred with corresponding velocity components. The gravitational potential $\psi_{i,j}$ is obtained through $\Sigma_{i,j}$ by solving Poisson's equation, as will be described described later in section A.4. Note also that the finite difference derivatives in the source terms are second-order accurate in space because of the staggered grid, while the time update is only of first order accuracy.

We now further update velocities due to Coriolis force. Because $u_{x\ (i,j)}$ and $u_{y\ (i,j)}$ are located at different positions within the cell, in order to update $u_{x\ (i,j)}$, we need to interpolate $u_{y\ (i,j)}$ at the location of $u_{x\ (i,j)}$ and vice versa. As a result, we get

$$\frac{u^{n+b}_{x\ (i,j)} - u^{n+a}_{x\ (i,j)}}{\Delta t} = 2\frac{\left(u^{n+a}_{y\ (i,j)} + u^{n+a}_{y\ (i-1,j)} + u^{n+a}_{y\ (i,j+1)} + u^{n+a}_{y\ (i-1,j+1)}\right)}{4},$$

$$\frac{u^{n+b}_{y\ (i,j)} - u^{n+a}_{y\ (i,j)}}{\Delta t} = (q-2)\frac{\left(u^{n+b}_{x\ (i,j)} + u^{n+b}_{x\ (i+1,j)} + u^{n+b}_{x\ (i,j-1)} + u^{n+b}_{x\ (i+1,j-1)}\right)}{4}.$$

We include artificial viscosity in our problem in order to capture shocks. In ZEUS, this is done through the classical von Neumann and Richtmyer prescription for artificial viscosity, which consists in introducing a new artificial pressure-strain diffusive term, $\mathbf{Q} = (Q_x, Q_y)$, into hydrodynamical equations of motion in addition to normal pressure. In other words, we replace $(\partial P/\partial x, \partial P/\partial y)$ with $(\partial(P+Q_x)/\partial x, \partial(P+Q_y)/\partial y)$ in momentum Eqs. (A.2) and (A.3) and replace the compressional heating term $P\nabla\cdot\mathbf{u}$ in Eq. (A.4) with $P\nabla\cdot\mathbf{u} + \mathbf{Q} : \nabla\mathbf{u}$. The discrete components of $\mathbf{Q}$ are also located at the cell centres and are defined as

$$Q_{x\ (i,j)} = \begin{cases} C\Sigma_{i,j}\left(u_{x\ (i+1,j)} - u_{x\ (i,j)}\right)^2, & \text{if } u_{x\ (i+1,j)} - u_{x\ (i,j)} < 0 \\ 0, & \text{otherwise} \end{cases}$$

$$Q_{y\ (i,j)} = \begin{cases} C\Sigma_{i,j}\left(u_{y\ (i,j+1)} - u_{y\ (i,j)}\right)^2, & \text{if } u_{y\ (i,j+1)} - u_{y\ (i,j)} < 0 \\ 0, & \text{otherwise} \end{cases}$$

where C is the dimensionless artificial viscosity parameter whose value is generally determined for each problem separately by numerical experiments (in our case we choose $C = 2.5$). It measures how much (i.e., on how many grid cells) artificial viscosity smears out discontinuities (shocks). This form of $\mathbf{Q}$ ensures that artificial viscosity becomes appreciable only in contracting regions ($\nabla \cdot \mathbf{u} < 0$), if physical quantities there undergo sharp variations (jumps) on the grid cell scale (as, for example, typically occurs in shocks) and leaves expanding regions ($\nabla \cdot \mathbf{u} > 0$) of the flow unaffected (since it is zero there). In other words, artificial viscosity effectively operates near shock fronts and simulates the actual energy dissipation occurring there, without practically influencing the rest of the flow (see Stone and Norman, 1992). This dissipation is taken into account in the internal energy evolution equation (A.4) as pointed out above,

$$\frac{U^{n+a}_{i,j} - U^{n}_{i,j}}{\Delta t} = -Q^{n+b}_{x\ (i,j)} \frac{u^{n+b}_{x\ (i+1,j)} - u^{n+b}_{x\ (i,j)}}{\Delta x} - Q^{n+b}_{y\ (i,j)} \frac{u^{n+b}_{y\ (i,j+1)} - u^{n+b}_{y\ (i,j)}}{\Delta y},$$

where in $\mathbf{Q}^{n+b}_{i,j}$, the value of the surface density $\Sigma^{n}_{i,j}$, which has not been updated yet, and the velocity $\mathbf{u}^{n+b}_{i,j}$ after updating due to the Coriolis force are implied. Note also that because Neumann-Richtmyer viscosity is operative only near the regions of rapid spatial variation of physical quantities and is negligible elsewhere, it is not expected to directly affect PV field, as the latter generally varies relatively smoothly over coordinates (see e.g., Figs. 3.13 and 3.14). For this reason, we have omitted viscous terms in the PV evolution equation in section 3.2, but, of course, we are including them in the energy equation here. Then, updating velocity components due to artificial viscosity proceeds similarly to that due to pressure terms:

$$\frac{u^{n+c}_{x\ (i,j)} - u^{n+b}_{x\ (i,j)}}{\Delta t} = -\frac{2\left(Q^{n+b}_{x\ (i,j)} - Q^{n+b}_{x\ (i-1,j)}\right)}{\Delta x(\Sigma^{n}_{i,j} + \Sigma^{n}_{i-1,j})}$$

$$\frac{u^{n+c}_{y\ (i,j)} - u^{n+b}_{y\ (i,j)}}{\Delta t} = -\frac{2\left(Q^{n+b}_{y\ (i,j)} - Q^{n+b}_{y\ (i,j-1)}\right)}{\Delta y(\Sigma^{n}_{i,j} + \Sigma^{n}_{i,j-1})}.$$

The final source terms are the compressional heating and cooling terms on the right hand side of the internal energy equation (A.4). To achieve a better conservation of the overall energy, updating due to the compressional heating term is done with a time-centred implicit scheme:

$$\frac{U^{n+b}_{i,j} - U^{n+a}_{i,j}}{\Delta t} = -P^{n+1/2}_{i,j} \left(\frac{u^{n+c}_{x\ (i+1,j)} - u^{n+c}_{x\ (i,j)}}{\Delta x} + \frac{u^{n+c}_{y\ (i,j+1)} - u^{n+c}_{y\ (i,j)}}{\Delta y} \right), \qquad \text{(A.7)}$$

where $P^{n+1/2}_{i,j} = 0.5(P^{n+a}_{i,j} + P^{n+b}_{i,j})$. Using the equation of state $P = (\gamma - 1)U$, this can be rearranged to give an explicit expression for $U^{n+b}_{i,j}$,

$$U^{n+b}_{i,j} = U^{n+a}_{i,j} \left[\frac{1 - 0.5\Delta t(\gamma - 1)(\nabla \cdot \mathbf{u})^{n+c}_{i,j}}{1 + 0.5\Delta t(\gamma - 1)(\nabla \cdot \mathbf{u})^{n+c}_{i,j}} \right],$$

where the velocity divergence in this equation is calculated as in Eq. (A.7).

Lastly, we update the internal energy for the cooling term

$$\frac{U_{i,j}^{n+c} - U_{i,j}^{n+b}}{\Delta t} = -\frac{U_{i,j}^{n+b}}{\tau_c}.$$

With this last step, we have completed updating of quantities for the source terms. Now we should update the quantities for advection by the flow.

A.2 Advection terms

To update the quantities for advection, we return to Eqs. (A.1-A.4) written for conserved quantities: mass, momentum and energy, omitting now the source terms. As mentioned above, there are two types of advection in these equations. The first one with $\nabla \cdot (..\mathbf{u})$ describes advection by the perturbed flow $\mathbf{u}$ and the second one with $-qx \cdot \partial/\partial y$ advection by the background Keplerian flow, whose velocity is much larger than the perturbed velocity. For this reason, different algorithms are employed to perform these two types of advection.

As the form of Eqs. (A.1-A.4) suggests, in order to update the surface density, momentum density and internal energy density within a cell by time interval Δt, we need to know the fluxes of these quantities through the cell edges, that is, how much mass, momentum and internal energy enters and leaves the cell during this time as a result of transport by the flow. Advection of these quantities by means of corresponding fluxes in the x- and y-directions is done again in an operator-splitting manner. First, we calculate all the fluxes in the x-direction and find updates to quantities with these fluxes. Then we further update by repeating the same procedure in the y-direction. These successive x- and y-advection steps are commonly referred to as sweeps. Initially we perform advection of all the quantities by the perturbed flow. The algorithm for this is similar to that used in the standard version of ZEUS (Stone and Norman, 1992). After that, we do an additional y-sweep to update for the azimuthal Keplerian velocity.

For further use, we note that the discrete components of momentum density (which are yet to be advected) at the end of the updating process due to the source terms are

$$p_{x\ (i,j)}^{n} = 0.5(\Sigma_{i,j}^{n} + \Sigma_{i-1,j}^{n})u_{x\ (i,j)}^{n+c},$$

$$p_{y\ (i,j)}^{n} = 0.5(\Sigma_{i,j}^{n} + \Sigma_{i,j-1}^{n})u_{y\ (i,j)}^{n+c},$$

obviously defined at the same locations as $u_{x\ (i,j)}$ and $u_{y\ (i,j)}$, respectively.

We begin performing x-sweep (which is due solely to the x-component of the perturbed flow velocity) of quantities with the surface density. The amount of mass crossing the left edge of the (i, j)-cell per unit time in the x-direction, that is, the x-component of the mass flux, is given by

$$\dot{M}_{x\ (i,j)} = u^{n+c}_{x\ (i,j)} \Sigma^*_{i,j} \Delta y,$$

and is defined at the same location within the cell as $u_{x\ (i,j)}$ (note that in this formula, we have used the value of the velocity already updated due to the source terms). Here $\Sigma^*_{i,j}$ is the average value of the density passing through this edge from time level n to $n + 1$. This average is found according to van Leer algorithm as follows. Let us define the slope centred at the cell edges as

$$\Delta\Sigma_{i,j} = \frac{\Sigma^n_{i,j} - \Sigma^n_{i-1,j}}{\Delta x}.$$

Then, $\Sigma^*_{i,j}$ is calculated by means of this slope in an upwind manner for numerical stability. If $u^{n+c}_{x\ (i,j)} > 0$, this average value is determined by the density of matter that flows into the given (i, j)-cell from the cell to its left (i.e., from the $(i - 1, j)$-cell) and is given by

$$\Sigma^*_{i,j} = \Sigma^n_{i-1,j} + \left(\Delta x - u^{n+c}_{x\ (i,j)} \Delta t\right) d\Sigma_{i-1,j}. \tag{A.8}$$

Otherwise, if $u^{n+c}_{x\ (i,j)} < 0$, then it is determined by the density of matter in the given cell that flows out towards the left cell

$$\Sigma^*_{i,j} = \Sigma^n_{i,j} - \left(\Delta x + u^{n+c}_{x\ (i,j)} \Delta t\right) d\Sigma_{i,j}, \tag{A.9}$$

where $d\Sigma_{i,j}$ is the monotonised slope centred at the same location as $\Sigma_{i,j}$ and defined as a harmonic average of the two adjacent values $\Delta\Sigma_{i,j}$ and $\Delta\Sigma_{i+1,j}$,

$$d\Sigma_{i,j} = \begin{cases} \frac{\Delta\Sigma_{i,j}\Delta\Sigma_{i+1,j}}{\Delta\Sigma_{i,j}+\Delta\Sigma_{i+1,j}}, & \text{if } \Delta\Sigma_{i,j}\Delta\Sigma_{i+1,j} > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{A.10}$$

In this case, the variation in surface density within a cell is interpolated by a linear function and therefore the scheme is accurate to second order in space. However, the scheme can become unstable if in such a piecewise linear distribution spurious maxima emerge, which were not present in the initial data. To avoid this problem, van Leer (1977) proposed the monotonisation procedure reflected in (A.10) for edge-centred slopes, which reduces the slope of the linear distribution and removes any spurious maxima wherever they appear. Thus, if the density variation is monotonic across the cell then the method is indeed accurate to second order, but wherever a spurious maximum emerges, the slope becomes zero according to Eq. (A.10) and therefore the density switches to a constant value across a cell, so that the scheme changes to being first-order accurate in space (donor-cell scheme). Because of reduction in accuracy, the method becomes at the same time more diffusive. In other words, the scheme is stabilised at the expense of its accuracy by increase in its diffusivity. There are more accurate schemes, such as, for example, a piecewise parabolic method (PPM), which uses a parabolic interpolation of quantities within grid cells instead of a linear one and thus is third-order accurate in space (Colella and Woodward, 1984). However, in the present version of ZEUS, linear interpolation, i.e., the original van Leer scheme, is employed, which still does pretty well as the test case below demonstrates.

Now knowing the x-component of the mass flux $\dot{M}_{x\ (i,j)}$, we can update/advect the surface density in this direction by taking the difference between the fluxes at the left and right edges of the given cell according to continuity equation (A.1)

$$\frac{\Sigma_{i,j}^{n+a} - \Sigma_{i,j}^{n}}{\Delta t} = -\frac{\dot{M}_{x\ (i+1,j)} - \dot{M}_{x\ (i,j)}}{\Delta x \Delta y}.$$

To advect now the momentum and internal energy densities, we need to know their corresponding fluxes through cell edges. Advection does not result in the net change of mass, momentum and energy, so that these quantities are conserved. However, local conservation can be affected by the numerical diffusion associated with the finite-difference representation of derivatives and interpolation algorithms. To circumvent this difficulty and improve the local conservation, a so-called 'consistent transport' algorithm is implemented in ZEUS (see Stone and Norman, 1992). Consistent transport attempts to reduce the effect of this diffusion by making the fluxes of momentum and internal energy consistent with the mass flux (as it should be physically), so that these quantities are advected in the same way as the mass is. In other words, when a mass element moves from one grid cell to another, its momentum and energy should move along with it.

We first construct the x-component of the internal energy flux, which is tied to the mass flux found above as

$$F^{U}_{x\ (i,j)} = \left(\frac{U}{\Sigma}\right)^{*}_{i,j} \dot{M}_{x\ (i,j)},$$

where $(U/\Sigma)^{*}$ is the averaged value of the specific internal energy across the cell edge calculated in the same way as that for the surface density above (i.e., by simply replacing Σ with U/Σ in Eqs. A.8-A.10). This flux is also defined at the same location as $\dot{M}_{x\ (i,j)}$. With this flux, we can update the internal energy in the given (i,j)-cell due to x-advection according to Eq. (A.4)

$$\frac{U_{i,j}^{n+d} - U_{i,j}^{n+c}}{\Delta t} = -\frac{F^{U}_{x\ (i+1,j)} - F^{U}_{x\ (i,j)}}{\Delta x \Delta y}.$$

We now consider x-advection of the momentum density. The velocity and therefore momentum components are located on the cell edges, so in order to advect them, momentum fluxes should be calculated at the cell centres. Momentum flux, equal to velocity multiplied by mass flux, is constructed similarly to the energy flux. So, interpolating the mass flux $\dot{M}_{x\ (i,j)}$ at the cell centres, we find the x-components of fluxes for the p_x and p_y components of momentum density

$$F^{p_x}_{x\ (i,j)} = 0.5u^*_{x\ (i,j)}\left(\dot{M}_{x\ (i,j)} + \dot{M}_{x\ (i+1,j)}\right),$$

$$F^{p_y}_{x\ (i,j)} = 0.5u^*_{y\ (i,j)}\left(\dot{M}_{x\ (i,j)} + \dot{M}_{x\ (i,j-1)}\right),$$

where again asterisks over the x- and y-components of velocity mean that these components are averaged according to the same van Leer algorithm as is the density. After that, the updates to the momentum components due to the x-advection are found from Eqs. (A.2) and (A.3)

$$\frac{p^{n+a}_{x\ (i,j)} - p^{n}_{x\ (i,j)}}{\Delta t} = -\frac{F^{p_x}_{x\ (i,j)} - F^{p_x}_{x\ (i-1,j)}}{\Delta x \Delta y},$$

$$\frac{p^{n+a}_{y\ (i,j)} - p^{n}_{y\ (i,j)}}{\Delta t} = -\frac{F^{p_y}_{x\ (i+1,j)} - F^{p_y}_{x\ (i,j)}}{\Delta x \Delta y}.$$

With this step, the advection of the basic quantities in the x-direction is completed. The advected velocities are found from

$$u^{n+d}_{x\ (i,j)} = \frac{2p^{n+a}_{x\ (i,j)}}{\Sigma^{n+a}_{i,j} + \Sigma^{n+a}_{i-1,j}},$$

$$u^{n+d}_{y\ (i,j)} = \frac{2p^{n+a}_{y\ (i,j)}}{\Sigma^{n+a}_{i,j} + \Sigma^{n+a}_{i,j-1}}.$$

Next, we perform y-sweep, which, as mentioned above, consists of two parts – advection due to the y-component of the perturbed flow velocity and to the azimuthal background Keplerian shear flow. We start with the former. As before, we first calculate mass flux in the y-direction and then, using it, construct the fluxes of energy and momentum. Because of operator-splitting, we carry on updating from the values of the quantities already advected in the x-direction. The y-component of the mass flux is found similarly to its x-component

$$\dot{M}_{y\ (i,j)} = u^{n+d}_{y\ (i,j)}\Sigma^*_{i,j}\Delta x,$$

which is located at the same position within the cell as $u_{y\ (i,j)}$ is. The asterisk here and in other flux expressions below denotes, as before, time-average of a quantity across a cell edge, but now in the y-direction. In other words, in Eqs. (A.8-A.10), we need to change $u_x \rightarrow u_y$

and act on the index j instead of i. Following from Eq. (A.1), the update to the surface density due to y-advection is

$$\frac{\Sigma_{i,j}^{n+b} - \Sigma_{i,j}^{n+a}}{\Delta t} = -\frac{\dot{M}_{y\ (i,j+1)} - \dot{M}_{y\ (i,j)}}{\Delta x \Delta y}.$$

The energy flux in the y-direction is constructed similarly to its x-component,

$$F^U_{y\ (i,j)} = \left(\frac{U}{\Sigma}\right)^*_{i,j} \dot{M}_{y\ (i,j)}$$

and we find the y-advected internal energy from Eq. (A.4)

$$\frac{U_{i,j}^{n+e} - U_{i,j}^{n+d}}{\Delta t} = -\frac{F^U_{y\ (i,j+1)} - F^U_{y\ (i,j)}}{\Delta x \Delta y}.$$

The y-components of the momentum fluxes are constructed in a similar way as their x-components

$$F^{p_x}_{y\ (i,j)} = 0.5 u^*_{x\ (i,j)} \left(\dot{M}_{y\ (i,j)} + \dot{M}_{y\ (i-1,j)}\right),$$

$$F^{p_y}_{y\ (i,j)} = 0.5 u^*_{y\ (i,j)} \left(\dot{M}_{y\ (i,j+1)} + \dot{M}_{y\ (i,j)}\right)$$

and we update the momentum components according to Eqs. (A.2) and (A.3)

$$\frac{p^{n+b}_{x\ (i,j)} - p^{n+a}_{x\ (i,j)}}{\Delta t} = -\frac{F^{p_x}_{y\ (i,j+1)} - F^{p_x}_{y\ (i,j)}}{\Delta x \Delta y},$$

$$\frac{p^{n+b}_{y\ (i,j)} - p^{n+a}_{y\ (i,j)}}{\Delta t} = -\frac{F^{p_y}_{y\ (i,j)} - F^{p_y}_{y\ (i,j-1)}}{\Delta x \Delta y}.$$

With this step, the advection of the quantities due to the perturbed flow in the y-direction is completed. The corresponding values of the velocity components are

$$u^{n+e}_{x\ (i,j)} = \frac{2p^{n+b}_{x\ (i,j)}}{\Sigma^{n+b}_{i,j} + \Sigma^{n+b}_{i-1,j}},$$

$$u^{n+e}_{y\ (i,j)} = \frac{2p^{n+b}_{y\ (i,j)}}{\Sigma^{n+b}_{i,j} + \Sigma^{n+b}_{i,j-1}}.$$

These updated values from the y-sweep due to the perturbed velocity still need to be advected due to the fast Keplerian shear flow $\mathbf{u_0}$. To do this, one could, in principle, apply the same technique as that used above for doing advection due to the perturbed velocity. However, this will result in a severe limitation on the time step (see below), that, in turn, could considerably slow down simulations. To circumvent this problem, in the present version of ZEUS, advection by Keplerian flow is done by means of the different – the so-called FARGO – algorithm developed by Masset (2000). The essence of this scheme is the following. At a fixed x-coordinate, advection due to Keplerian flow during the time interval Δt in Eqs. (A.1-A.4) results in a shift of all the quantities by the distance $-qx\Delta t$ in the y-direction, which can be greater than the domain size, L, because of the large value of Keplerian velocity compared to the perturbed velocity $\mathbf{u}$. We can represent this shift as a sum $-qx\Delta t = kL + l\Delta y + \delta y$, where $0 \leq k$ and $0 \leq l < N$ are some integer numbers and $\delta y < \Delta y$ is the remainder. Because of the periodic boundary conditions in the y-direction, shift by the distance kL does not change the distribution of quantities and can be ignored. Advection by an integer number of grid cells, $l\Delta y$, is done by means of a simple re-mapping procedure, i.e., the values of the quantities at the (i, j)-cell are put into the $(i, j - \mathrm{sign}(x)l)$-cell.[1] Finally, shift by δy is equivalent to the advection with the velocity $\delta y/\Delta t$, which is comparable to the y-component of the perturbed velocity u_y. So, we do this shift by applying the same method as that used above for advection due to the perturbed velocity, except that we replace the latter with $\delta y/\Delta t$. Thus, from the numerical point of view, advection by Keplerian flow has in fact been reduced to the advection due to much smaller velocity, comparable to the perturbed velocity. The re-mapping procedure, which actually does the majority of advection due to the Keplerian flow, is quite straightforward not requiring any special algorithm. Below we will see the advantage of this separation of advection steps in the presence of Keplerian flow when setting the time step. With this last step, we obtain the fully updated values of the quantities at time level $n + 1$: $\Sigma^{n+1}_{i,j}, u^{n+1}_{x\ (i,j)}, u^{n+1}_{y\ (i,j)}, U^{n+1}_{i,j}$.

A.3 Time step

We should also discuss the requirements for the time step Δt used above. This time step is chosen such that to satisfy the Courant-Friedrichs-Lewy condition. There are several characteristic time-scales in the problem. Thanks to the FARGO scheme, the bulk of advection due to Keplerian velocity, being done by re-mapping over an integer number of grid cells, does not put any limitations on the time step. So, only advection by the perturbed flow can constrain the time step. The characteristic time-scales related to this advection in the x- and y-directions for the (i, j)-cell are

$$\Delta t_{1\ (i,j)} = \frac{\Delta x}{|u_{x\ (i,j)}| + c_{s\ (i,j)}}, \quad \Delta t_{2\ (i,j)} = \frac{\Delta y}{|u_{y\ (i,j)}| + c_{s\ (i,j)}}.$$

If we did advection without using FARGO, these time-scales would contain, instead of the perturbed velocity $\mathbf{u}$, the full velocity $\mathbf{u_0} + \mathbf{u}$ that would severely reduce them and slow

---

[1] If $|j - \mathrm{sign}(x)l| > N$, we subtract from or add to j multiples of N (which we can do because of periodicity in N) until $|j - \mathrm{sign}(x)l| \leq N$.

down simulations. There are also the other time-scales to be considered, which are related to artificial viscosity (Stone and Norman, 1992):

$$\Delta t_{3\ (i,j)} = \frac{\Delta x}{4C|u_{x\ (i+1,j)} - u_{x\ (i,j)}|}, \quad \Delta t_{4\ (i,j)} = \frac{\Delta y}{4C|u_{y\ (i,j+1)} - u_{y\ (i,j)}|},$$

which are obviously defined when artificial viscosity is not zero, i.e., for convergent flows with $u_{x\ (i+1,j)} - u_{x\ (i,j)} < 0$ and/or $u_{y\ (i,j+1)} - u_{y\ (i,j)} < 0$). From these two viscous time-scales we choose the smallest

$$\Delta t_{av\ (i,j)} = \min(\Delta t_{3\ (i,j)}, \Delta t_{4\ (i,j)}).$$

In ZEUS, the final time step, Δt, used in simulations is then composed as

$$\Delta t = \frac{A}{\sqrt{\max_{1\leq i,j\leq N}\left(\Delta t_{1\ (i,j)}^{-2} + \Delta t_{2\ (i,j)}^{-2} + \Delta t_{av\ (i,j)}^{-2}\right)}},$$

where the maximum is taken over all grid cells and A is the so-called Courant number – a safety factor less than one ensuring that during one time step fluid elements do not travel a distance larger than the cell size. This is a necessary condition for the numerical stability of any explicit hydrodynamical scheme and is known as the Courant-Friedrichs-Lewy condition. In our simulations, we take $A = 0.3$, which gives fairly good results.

A.4 Solving Poisson's equation

To express $\psi_{i,j}$ via $\Sigma_{i,j}$, we solve Poisson's equation for disc self-gravity in razor-thin (2D) geometry with a method paralleling that of Johansen et al. (2007) and differing from the original method of Gammie (2001) in the following respect. In Gammie (2001), the surface density is first linearly interpolated onto a grid of shearing coordinates ($x' = x, y' \equiv y + qx(t - t_p)$) with $t_p = (q\Omega)^{-1}NINT(q\Omega t)$, where it is exactly periodic, and then Fourier transformed with a standard Fast Fourier Transform (FFT) technique. After that the Fourier transform of the gravitational potential is found from the Poisson's equation rewritten in the wavenumber **K**-plane

$$\psi(K_x, K_y, t)|_{z=0} = -\frac{2}{Q\sqrt{(K_x + qK_y(t - t_p))^2 + K_y^2}}\Sigma(K_x, K_y, t),$$

where $\psi(K_x, K_y, t)|_{z=0}$ and $\Sigma(K_x, K_y, t)$ are the Fourier transforms of the gravitational potential and surface density with the background constant value subtracted. The potential is then transformed back into the grid of shearing coordinates and finally linearly interpolated onto the grid of (x, y) coordinates. We instead fourier transform directly from the (x, y) plane to the (K_x, K_y) plane and vice versa without doing an intermediate transformation

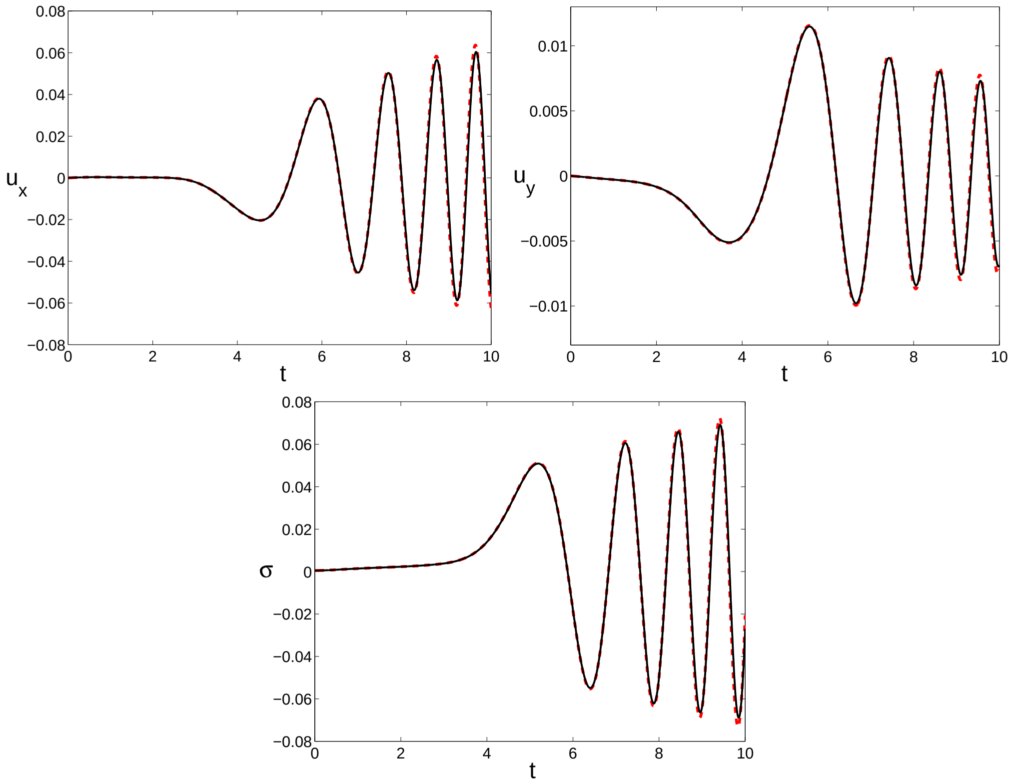

Figure A.2: Evolution of the velocity and surface density amplitudes of a single, small amplitude (i.e., linear) shearing wave with $K_x(0) = -18\pi/L, K_y = 4\pi/L$ for $Q = 1$ as given by the code (solid line) and by more exact time-integration of the system of first order differential equations with respect to time (dashed line) obtained from linearised Eqs. (3.1-3.5) after substituting shearing wave solution. The domain size is $L = 15$ and the resolution is $N = 512$. As clearly seen, an agreement between these two solutions is very good.

to the shearing coordinates. However, in this case we take into account the fact that the radial wavenumber K_x of each spatial fourier harmonic is no longer constant, but changes with time according to $K_x(t) = K_x(0) + qK_y t$ in order to remain consistent with the shearing sheet boundary conditions (see section 3.4). As a consequence, at each time t, in the corresponding computational domain in the **K**-plane, there are only fourier harmonics with wavenumbers $K_y = 2\pi n_y/L, K_x = 2\pi n_x/L + qt'(2\pi n_y/L)$, where $t' = mod(t, 1/(q|n_y|))$ (here $mod(a, b)$ is modulus after dividing a by b) and integer numbers n_x, n_y lie in the interval $-N/2 \leq n_x, n_y \leq N/2$. So, when doing Fourier (inverse Fourier) transform we decompose (sum) into (over) these wavenumbers. We accordingly modified the standard FFT routine to perform this decomposition (summation). In some sense, this method is equivalent to changing from the (x, y) plane to the (x', y') plane via Fourier transformation, which is more exact than that done by linear interpolation.[2].

---

[2] See also website http://imp.mcmaster.ca/∼colinm/ism/rotfft.html

A.5 Test

To test this version of the ZEUS code and our new solver for self-gravity, we evolved with them a small amplitude shearing wave from leading to trailing phases at the resolution $N = 512$. In Fig. A.2, we show the temporal evolution of the velocity and surface density amplitudes according to the basic non-linear Eqs. (3.1-3.5) (with cooling term omitted) as given by the code. We compare this solution with the evolution of the same amplitudes resulting from a direct time-integration of the system of first order differential equations, which is obtained by substituting spatial shearing waves (3.10) into linearised Eqs. (3.1-3.5) (again without cooling term). It is clearly seen that the code results are in an excellent agreement with those of the linear theory until the decreasing radial wavelength of shearing waves becomes comparable to the grid cell size. More similar tests for this version of ZEUS in the (non)-self-gravitating shearing sheet, involving vortical and wave mode shearing waves, were performed by Gammie (2001) and Johnson and Gammie (2005b). Thus, the code has been well tested previously and we can be confident in its reliable performance when applied to our problem.

Bibliography

Adams, F. C., Lada, C. J., and Shu, F. H.: 1987, *ApJ,* **312**, 788

Adams, F. C., Lada, C. J., and Shu, F. H.: 1988, *ApJ,* **326**, 865

Adams, F. C., Ruden, S. P., and Shu, F. H.: 1989, *ApJ,* **347**, 959

Adams, F. C. and Watkins, R.: 1995, *ApJ,* **451**, 314

Afshordi, N., Mukhopadhyay, B., and Narayan, R.: 2005, *ApJ,* **629**, 373

Andrews, S. M. and Williams, J. P.: 2005, *ApJ,* **631**, 1134

Andrews, S. M. and Williams, J. P.: 2007a, *ApJ,* **671**, 1800

Andrews, S. M. and Williams, J. P.: 2007b, *ApJ,* **659**, 705

Andrews, S. M., Wilner, D. J., Hughes, A. M., Qi, C., and Dullemond, C. P.: 2009, *ApJ,* **700**, 1502

Andrews, S. M., Wilner, D. J., Hughes, A. M., Qi, C., and Dullemond, C. P.: 2010, *ApJ,* **723**, 1241

Armitage, P. J.: 1998, *ApJ,* **501**, L189

Armitage, P. J.: 2010, *Astrophysics of Planet Formation*, Cambridge University Press, Cambridge, UK

Armitage, P. J., Livio, M., and Pringle, J. E.: 2001, *MNRAS,* **324**, 705

Baggett, J. S., Driscoll, T. A., and Trefethen, L. N.: 1995, *Physics of Fluids,* **7**, 833

Bakas, N. A. and Farrell, B. F.: 2009, *Journal of Physical Oceanography,* **39**, 497

Balbus, S. A.: 2000, *ApJ,* **534**, 420

Balbus, S. A.: 2003, *ARA&A,* **41**, 555

Balbus, S. A. and Hawley, J. F.: 1991, *ApJ,* **376**, 214

Balbus, S. A. and Hawley, J. F.: 1992, *ApJ,* **400**, 610

Balbus, S. A. and Hawley, J. F.: 1998, *Reviews of Modern Physics,* **70**, 1

Balbus, S. A., Hawley, J. F., and Stone, J. M.: 1996, *ApJ,* **467**, 76

Balbus, S. A. and Papaloizou, J. C. B.: 1999, *ApJ,* **521**, 650

Barge, P. and Sommeria, J.: 1995, *A&A,* **295**, L1

Barranco, J. A. and Marcus, P. S.: 2005, *ApJ,* **623**, 1157

Bate, M. R., Ogilvie, G. I., Lubow, S. H., and Pringle, J. E.: 2002, *MNRAS,* **332**, 575

Bayly, B. J., Orszag, S. A., and Herbert, T.: 1988, *Annual Review of Fluid Mechanics,* **20**, 359

Bech, K. H., Tillmark, N., Alfredsson, P. H., and Andersson, H. I.: 1995, *Journal of Fluid Mechanics,* **286**, 291

Beckwith, S., Sargent, A. I., Scoville, N. Z., Masson, C. R., Zuckerman, B., and Phillips, T. G.: 1986, *ApJ,* **309**, 755

Beckwith, S. V. W., Sargent, A. I., Chini, R. S., and Guesten, R.: 1990, *AJ,* **99**, 924

Bell, K. R., Cassen, P. M., Klahr, H. H., and Henning, T.: 1997, *ApJ,* **486**, 372

Bell, K. R. and Lin, D. N. C.: 1994, *ApJ,* **427**, 987

Bertin, G., Lin, C. C., Lowe, S. A., and Thurstans, R. P.: 1989, *ApJ,* **338**, 104

Bertin, G. and Lodato, G.: 2001, *A&A,* **370**, 342

Bidokhti, A. A. and Tritton, D. J.: 1992, *Journal of Fluid Mechanics,* **241**, 469

Binney, J. and Tremaine, S.: 1987, *Galactic dynamics*, Princeton University Press, Princeton, NJ

Blaes, O. M.: 1987, *MNRAS,* **227**, 975

Blaes, O. M. and Balbus, S. A.: 1994, *ApJ,* **421**, 163

Bodenheimer, P.: 1995, *ARA&A,* **33**, 199

Bodo, G., Chagelishvili, G., Murante, G., Tevzadze, A., Rossi, P., and Ferrari, A.: 2005, *A&A,* **437**, 9

Bodo, G., Mignone, A., Cattaneo, F., Rossi, P., and Ferrari, A.: 2008, *A&A,* **487**, 1

Bodo, G., Poedts, S., Rogava, A., and Rossi, P.: 2001, *A&A,* **374**, 337

Bodo, G., Tevzadze, A., Chagelishvili, G., Mignone, A., Rossi, P., and Ferrari, A.: 2007, *A&A,* **475**, 51

Boley, A. C.: 2009, *ApJ,* **695**, L53

Boley, A. C. and Durisen, R. H.: 2006, *ApJ,* **641**, 534

Boley, A. C., Durisen, R. H., Nordlund, Å., and Lord, J.: 2007, *ApJ,* **665**, 1254

Boley, A. C., Hayfield, T., Mayer, L., and Durisen, R. H.: 2010, *Icarus,* **207**, 509

Boley, A. C., Mejía, A. C., Durisen, R. H., Cai, K., Pickett, M. K., and D'Alessio, P.: 2006, *ApJ,* **651**, 517

Boss, A. P.: 1998, *ApJ,* **503**, 923

Boss, A. P.: 2002, *ApJ,* **576**, 462

Boss, A. P.: 2003, *ApJ,* **599**, 577

Boss, A. P.: 2004, *ApJ,* **610**, 456

Boss, A. P.: 2007, *ApJ,* **661**, L73

Boss, A. P.: 2008, *ApJ,* **677**, 607

Bottin, S., Dauchot, O., and Daviaud, F.: 1997, *Physical Review Letters,* **79**, 4377

Bracco, A., Chavanis, P. H., Provenzale, A., and Spiegel, E. A.: 1999, *Physics of Fluids,* **11**, 2280

Brandenburg, A. and Dintrans, B.: 2006, *A&A,* **450**, 437

Brandenburg, A., Nordlund, A., Stein, R. F., and Torkelsson, U.: 1995, *ApJ,* **446**, 741

Brosa, U. and Grossmann, S.: 1999, *European Physical Journal B,* **9**, 343

Butler, K. M. and Farrell, B. F.: 1992, *Physics of Fluids,* **4**, 1637

Cabot, W.: 1996, *ApJ,* **465**, 874

Cabot, W. and Pollack, J. B.: 1992, *Geophysical and Astrophysical Fluid Dynamics,* **64**, 97

Cai, K., Durisen, R. H., Michael, S., Boley, A. C., Mejía, A. C., Pickett, M. K., and D'Alessio, P.: 2006, *ApJ,* **636**, L149

Cai, K., Pickett, M. K., Durisen, R. H., and Milne, A. M.: 2010, *ApJ,* **716**, L176

Calvet, N., Hartmann, L., and Strom, S. E.: 2000, *Protostars and Planets IV,* p. 377

Calvet, N., Magris, G. C., Patino, A., and D'Alessio, P.: 1992, *RMxAA,* **24**, 27

Cameron, A. G. W.: 1978, *Moon and Planets,* **18**, 5

Cassen, P. and Moosman, A.: 1981, *Icarus,* **48**, 353

Chagelishvili, G. D., Rogava, A. D., and Tsiklauri, D. G.: 1996, *Phys. Rev. E,* **53**, 6028

Chagelishvili, G. D., Tevzadze, A. G., Bodo, G., and Moiseev, S. S.: 1997, *Physical Review Letters,* **79**, 3178

Chagelishvili, G. D., Zahn, J., Tevzadze, A. G., and Lominadze, J. G.: 2003, *A&A,* **402**, 401

Chandrasekhar, S.: 1960, *Proceedings of the National Academy of Science,* **46**, 253

Chavanis, P. H.: 2000, *A&A,* **356**, 1089

Chiang, E. I. and Goldreich, P.: 1997, *ApJ,* **490**, 368

Chiang, E. I., Joung, M. K., Creech-Eakman, M. J., Qi, C., Kessler, J. E., Blake, G. A., and van Dishoeck, E. F.: 2001, *ApJ,* **547**, 1077

Clarke, C. J.: 2009, *MNRAS,* **396**, 1066

Colella, P. and Woodward, P. R.: 1984, *J. Comp. Phys.,* **54**, 174

Cossins, P., Lodato, G., and Clarke, C.: 2010, *MNRAS,* **401**, 2587

Cossins, P., Lodato, G., and Clarke, C. J.: 2009, *MNRAS,* **393**, 1157

D'Alessio, P., Canto, J., Calvet, N., and Lizano, S.: 1998, *ApJ,* **500**, 411

Dauchot, O. and Daviaud, F.: 1995, *Physics of Fluids,* **7**, 901

Davis, S. S.: 2002, *ApJ,* **576**, 450

Davis, S. S., Sheehan, D. P., and Cuzzi, J. N.: 2000, *ApJ,* **545**, 494

Davis, S. W., Stone, J. M., and Pessah, M. E.: 2010, *ApJ,* **713**, 52

Desch, S. J.: 2004, *ApJ,* **608**, 509

Dodson-Robinson, S. E., Veras, D., Ford, E. B., and Beichman, C. A.: 2009, *ApJ,* **707**, 79

Drazin, P. G. and Reid, W. H.: 1981, *Hydrodynamic Stability*, Cambridge University Press, Cambridge, UK

Dubrulle, B.: 1993, *Icarus,* **106**, 59

Dubrulle, B., Marié, L., Normand, C., Richard, D., Hersant, F., and Zahn, J.: 2005, *A&A,* **429**, 1

Dubrulle, B. and Zahn, J.: 1991, *Journal of Fluid Mechanics,* **231**, 561

Dullemond, C. P., Hollenbach, D., Kamp, I., and D'Alessio, P.: 2007, *Protostars and Planets V,* pp 555–572

Dunham, M. M., Evans, N. J., Terebey, S., Dullemond, C. P., and Young, C. H.: 2010, *ApJ,* **710**, 470

Durisen, R. H., Boss, A. P., Mayer, L., Nelson, A. F., Quinn, T., and Rice, W. K. M.: 2007, *Protostars and Planets V,* pp 607–622

Dutrey, A.: 2000, in G. Garzón, C. Eiroa, D. de Winter, & T. J. Mahoney (ed.), *Disks, Planetesimals, and Planets*, Vol. 219 of *Astronomical Society of the Pacific Conference Series*, p. 128

Dutrey, A., Guilloteau, S., Duvert, G., Prato, L., Simon, M., Schuster, K., and Menard, F.: 1996, *A&A,* **309**, 493

Dutrey, A., Guilloteau, S., and Simon, M.: 1994, *A&A,* **286**, 149

Eckhardt, B. and Mersmann, A.: 1999, *Phys. Rev. E,* **60**, 509

Eisner, J. A. and Carpenter, J. M.: 2006, *ApJ,* **641**, 1162

Eisner, J. A., Hillenbrand, L. A., Carpenter, J. M., and Wolf, S.: 2005, *ApJ,* **635**, 396

Farrell, B. F. and Ioannou, P. J.: 1993, *Physics of Fluids,* **5**, 1390

Farrell, B. F. and Ioannou, P. J.: 2000, *Physics of Fluids,* **12**, 3021

Fleming, T. and Stone, J. M.: 2003, *ApJ,* **585**, 908

Forgan, D., Rice, K., Stamatellos, D., and Whitworth, A.: 2009, *MNRAS,* **394**, 882

Frank, J., King, A., and Raine, D. J.: 2002, *Accretion Power in Astrophysics: Third Edition*, Cambridge University Press, Cambridge, UK

Fromang, S.: 2005, *A&A,* **441**, 1

Fromang, S. and Nelson, R. P.: 2005, *MNRAS,* **364**, L81

Fromang, S. and Papaloizou, J.: 2007, *A&A,* **476**, 1113

Fromang, S., Papaloizou, J., Lesur, G., and Heinemann, T.: 2007, *A&A,* **476**, 1123

Fromang, S., Terquem, C., and Balbus, S. A.: 2002, *MNRAS,* **329**, 18

Gammie, C. F.: 1996, *ApJ,* **457**, 355

Gammie, C. F.: 1999, in J. A. Sellwood & J. Goodman (ed.), *Astrophysical Discs - an EC Summer School*, Vol. 160 of *Astronomical Society of the Pacific Conference Series,*, p. 122

Gammie, C. F.: 2001, *ApJ,* **553**, 174

Gammie, C. F. and Menou, K.: 1998, *ApJ,* **492**, L75

Gebhardt, T. and Grossmann, S.: 1994, *Phys. Rev. E,* **50**, 3705

Godon, P. and Livio, M.: 1999a, *ApJ,* **521**, 319

Godon, P. and Livio, M.: 1999b, *ApJ,* **523**, 350

Godon, P. and Livio, M.: 2000, *ApJ,* **537**, 396

Goldreich, P. and Lynden-Bell, D.: 1965a, *MNRAS,* **130**, 97

Goldreich, P. and Lynden-Bell, D.: 1965b, *MNRAS,* **130**, 125

Goldreich, P. and Tremaine, S.: 1978, *ApJ,* **222**, 850

Goodman, A. A., Benson, P. J., Fuller, G. A., and Myers, P. C.: 1993, *ApJ,* **406**, 528

Greene, T. P., Wilking, B. A., Andre, P., Young, E. T., and Lada, C. J.: 1994, *ApJ,* **434**, 614

Grossmann, S.: 2000, *Reviews of Modern Physics,* **72**, 603

Guilloteau, S. and Dutrey, A.: 1998, *A&A,* **339**, 467

Gullbring, E., Hartmann, L., Briceno, C., and Calvet, N.: 1998, *ApJ,* **492**, 323

Gustavsson, L. H.: 1991, *Journal of Fluid Mechanics,* **224**, 241

Haghighipour, N. and Boss, A. P.: 2003, *ApJ,* **583**, 996

Haisch, Jr., K. E., Lada, E. A., and Lada, C. J.: 2001, *ApJ,* **553**, L153

Hamilton, J. M., Kim, J., and Waleffe, F.: 1995, *Journal of Fluid Mechanics,* **287**, 317

Hartmann, L., Calvet, N., Gullbring, E., and D'Alessio, P.: 1998, *ApJ,* **495**, 385

Hartmann, L., D'Alessio, P., Calvet, N., and Muzerolle, J.: 2006, *ApJ,* **648**, 484

Hartmann, L., Hewett, R., Stahler, S., and Mathieu, R. D.: 1986, *ApJ,* **309**, 275

Hartmann, L. and Kenyon, S. J.: 1996, *ARA&A,* **34**, 207

Hawley, J. F.: 1991, *ApJ,* **381**, 496

Hawley, J. F.: 2001, *ApJ,* **554**, 534

Hawley, J. F., Balbus, S. A., and Winters, W. F.: 1999, *ApJ,* **518**, 394

Hawley, J. F., Gammie, C. F., and Balbus, S. A.: 1995, *ApJ,* **440**, 742

Hayashi, C.: 1981, *Progress of Theoretical Physics Supplement,* **70**, 35

Heinemann, T. and Papaloizou, J. C. B.: 2009a, *MNRAS,* **397**, 52

Heinemann, T. and Papaloizou, J. C. B.: 2009b, *MNRAS,* **397**, 64

Herbig, G. H.: 1977, *ApJ,* **217**, 693

Hueso, R. and Guillot, T.: 2005, *A&A,* **442**, 703

Hughes, A. M., Wilner, D. J., Qi, C., and Hogerheijde, M. R.: 2008, *ApJ,* **678**, 1119

Inaba, S. and Barge, P.: 2006, *ApJ,* **649**, 415

Ioannou, P. J. and Kakouris, A.: 2001, *ApJ,* **550**, 931

Isella, A., Carpenter, J. M., and Sargent, A. I.: 2009, *ApJ,* **701**, 260

Isella, A., Testi, L., Natta, A., Neri, R., Wilner, D., and Qi, C.: 2007, *A&A,* **469**, 213

Ji, H., Burin, M., Schartman, E., and Goodman, J.: 2006, *Nature,* **444**, 343

Johansen, A., Andersen, A. C., and Brandenburg, A.: 2004, *A&A,* **417**, 361

Johansen, A., Oishi, J. S., Mac Low, M., Klahr, H., Henning, T., and Youdin, A.: 2007, *Nature,* **448**, 1022

Johnson, B. M. and Gammie, C. F.: 2003, *ApJ,* **597**, 131

Johnson, B. M. and Gammie, C. F.: 2005a, *ApJ,* **626**, 978

Johnson, B. M. and Gammie, C. F.: 2005b, *ApJ,* **635**, 149

Johnson, B. M. and Gammie, C. F.: 2006, *ApJ,* **636**, 63

Kalas, P., Graham, J. R., Chiang, E., Fitzgerald, M. P., Clampin, M., Kite, E. S., Stapelfeldt, K., Marois, C., and Krist, J.: 2008, *Science,* **322**, 1345

Käpylä, P. J., Brandenburg, A., Korpi, M. J., Snellman, J. E., and Narayan, R.: 2010, *ApJ,* **719**, 67

Kenyon, S. J., Calvet, N., and Hartmann, L.: 1993, *ApJ,* **414**, 676

Kenyon, S. J. and Hartmann, L.: 1987, *ApJ,* **323**, 714

Kenyon, S. J. and Hartmann, L.: 1995, *ApJS,* **101**, 117

Kenyon, S. J., Hartmann, L. W., Strom, K. M., and Strom, S. E.: 1990, *AJ,* **99**, 869

Kerswell, R. R.: 2002, *Annual Review of Fluid Mechanics,* **34**, 83

Kim, W. and Ostriker, E. C.: 2001, *ApJ,* **559**, 70

King, A. R., Pringle, J. E., and Livio, M.: 2007, *MNRAS,* **376**, 1740

Kitamura, Y., Momose, M., Yokogawa, S., Kawabe, R., Tamura, M., and Ida, S.: 2002, *ApJ,* **581**, 357

Klahr, H.: 2004, *ApJ,* **606**, 1070

Klahr, H. and Bodenheimer, P.: 2006, *ApJ,* **639**, 432

Klahr, H. H. and Bodenheimer, P.: 2003, *ApJ,* **582**, 869

Klahr, H. H., Henning, T., and Kley, W.: 1999, *ApJ,* **514**, 325

Kley, W., Papaloizou, J. C. B., and Lin, D. N. C.: 1993, *ApJ,* **416**, 679

Korycansky, D. G.: 1992, *ApJ,* **399**, 176

Korycansky, D. G. and Pringle, J. E.: 1995, *MNRAS,* **272**, 618

Kratter, K. M., Matzner, C. D., and Krumholz, M. R.: 2008, *ApJ,* **681**, 375

Kratter, K. M., Murray-Clay, R. A., and Youdin, A. N.: 2010, *ApJ,* **710**, 1375

Lada, C. J.: 1987, in M. Peimbert & J. Jugaku (ed.), *Star Forming Regions*, Vol. 115 of *IAUS*, pp 1–17

Lada, C. J. and Wilking, B. A.: 1984, *ApJ,* **287**, 610

Latter, H. N. and Balbus, S. A.: 2009, *MNRAS,* **399**, 1058

Laughlin, G. and Bodenheimer, P.: 1994, *ApJ,* **436**, 335

Laughlin, G., Korchagin, V., and Adams, F. C.: 1997, *ApJ,* **477**, 410

Laughlin, G. and Rozyczka, M.: 1996, *ApJ,* **456**, 279

Lerche, I. and Parker, E. N.: 1967, *ApJ,* **149**, 559

Lerner, J. and Knobloch, E.: 1988, *Journal of Fluid Mechanics,* **189**, 117

Lesur, G. and Longaretti, P.: 2005, *A&A,* **444**, 25

Lesur, G. and Ogilvie, G. I.: 2010, *MNRAS,* **404**, L64

Lesur, G. and Papaloizou, J. C. B.: 2009, *A&A,* **498**, 1

Lesur, G. and Papaloizou, J. C. B.: 2010, *A&A,* **513**, 60

Li, H., Colgate, S. A., Wendroff, B., and Liska, R.: 2001, *ApJ,* **551**, 874

Li, H., Finn, J. M., Lovelace, R. V. E., and Colgate, S. A.: 2000, *ApJ,* **533**, 1023

Li, L., Goodman, J., and Narayan, R.: 2003, *ApJ,* **593**, 980

Lin, D. N. C. and Papaloizou, J.: 1980, *MNRAS,* **191**, 37

Lin, D. N. C. and Papaloizou, J.: 1985, in D. C. Black & M. S. Matthews (ed.), *Protostars and Planets II*, pp 981–1072

Lin, D. N. C. and Papaloizou, J. C. B.: 1993, in E. H. Levy & J. I. Lunine (ed.), *Protostars and Planets III*, pp 749–835

Lin, D. N. C., Papaloizou, J. C. B., and Kley, W.: 1993, *ApJ,* **416**, 689

Lin, D. N. C., Papaloizou, J. C. B., and Savonije, G. J.: 1990, *ApJ,* **364**, 326

Lithwick, Y.: 2007, *ApJ,* **670**, 789

Lithwick, Y.: 2009, *ApJ,* **693**, 85

Lodato, G.: 2007, *Nuovo Cimento Rivista Serie,* **30**, 293

Lodato, G. and Rice, W. K. M.: 2004, *MNRAS,* **351**, 630

Lodato, G. and Rice, W. K. M.: 2005, *MNRAS,* **358**, 1489

Longaretti, P.: 2002, *ApJ,* **576**, 587

Longaretti, P. and Lesur, G.: 2010, *A&A,* **516**, 51

Longaretti, P.-Y. and Dauchot, O.: 2005, in Kerswell & Mullin (ed.), *Proc. Bristol 2004 Symposium on the laminar-turbulent transition*, Kluwer

Lovelace, R. V. E., Li, H., Colgate, S. A., and Nelson, A. F.: 1999, *ApJ,* **513**, 805

Lubow, S. H. and Ogilvie, G. I.: 1998, *ApJ,* **504**, 983

Lubow, S. H. and Pringle, J. E.: 1993a, *MNRAS,* **263**, 701

Lubow, S. H. and Pringle, J. E.: 1993b, *ApJ,* **409**, 360

Lynden-Bell, D. and Kalnajs, A. J.: 1972, *MNRAS,* **157**, 1

Lynden-Bell, D. and Pringle, J. E.: 1974, *MNRAS,* **168**, 603

Lyra, W., Johansen, A., Zsom, A., Klahr, H., and Piskunov, N.: 2009, *A&A,* **497**, 869

Mamatsashvili, G. R. and Chagelishvili, G. D.: 2007, *MNRAS,* **381**, 809

Mamatsashvili, G. R. and Rice, W. K. M.: 2009, *MNRAS,* **394**, 2153

Mamatsashvili, G. R. and Rice, W. K. M.: 2010, *MNRAS,* **406**, 2050

Marois, C., Macintosh, B., Barman, T., Zuckerman, B., Song, I., Patience, J., Lafrenière, D., and Doyon, R.: 2008, *Science,* **322**, 1348

Masset, F.: 2000, *A&AS,* **141**, 165

Mayer, L., Lufkin, G., Quinn, T., and Wadsley, J.: 2007, *ApJ,* **661**, L77

Mejía, A. C., Durisen, R. H., Pickett, M. K., and Cai, K.: 2005, *ApJ,* **619**, 1098

Menou, K. and Quataert, E.: 2001, *ApJ,* **552**, 204

Mukhopadhyay, B., Afshordi, N., and Narayan, R.: 2005, *ApJ,* **629**, 383

Nakagawa, Y. and Sekiya, M.: 1992, *MNRAS,* **256**, 685

Narayan, R., Goldreich, P., and Goodman, J.: 1987, *MNRAS,* **228**, 1

Natta, A., Grinin, V., and Mannings, V.: 2000, *Protostars and Planets IV,* p. 559

Nelson, A. F.: 2006, *MNRAS,* **373**, 1039

Nelson, A. F., Benz, W., and Ruzmaikina, T. V.: 2000, *ApJ,* **529**, 357

Nelson, R. P.: 2005, *A&A,* **443**, 1067

Nelson, R. P. and Papaloizou, J. C. B.: 2004, *MNRAS,* **350**, 849

Ogilvie, G. I.: 1998, *MNRAS,* **297**, 291

Ogilvie, G. I. and Lubow, S. H.: 1999, *ApJ,* **515**, 767

Oishi, J. S. and Mac Low, M.: 2009, *ApJ,* **704**, 1239

Oishi, J. S., Mac Low, M., and Menou, K.: 2007, *ApJ,* **670**, 805

Orszag, S. A. and Patera, A. T.: 1980, *Physical Review Letters,* **45**, 989

Orszag, S. A. and Patera, A. T.: 1983, *Journal of Fluid Mechanics,* **128**, 347

Osorio, M., D'Alessio, P., Muzerolle, J., Calvet, N., and Hartmann, L.: 2003, *ApJ,* **586**, 1148

Padgett, D. L., Brandner, W., Stapelfeldt, K. R., Strom, S. E., Terebey, S., and Koerner, D.: 1999, *AJ,* **117**, 1490

Papaloizou, J. C. and Savonije, G. J.: 1991, *MNRAS,* **248**, 353

Papaloizou, J. C. B. and Lin, D. N. C.: 1989, *ApJ,* **344**, 645

Papaloizou, J. C. B. and Nelson, R. P.: 2003, *MNRAS,* **339**, 983

Papaloizou, J. C. B. and Pringle, J. E.: 1984, *MNRAS,* **208**, 721

Papaloizou, J. C. B. and Pringle, J. E.: 1985, *MNRAS,* **213**, 799

Petersen, M. R., Julien, K., and Stewart, G. R.: 2007a, *ApJ,* **658**, 1236

Petersen, M. R., Stewart, G. R., and Julien, K.: 2007b, *ApJ,* **658**, 1252

Pickett, B. K., Cassen, P., Durisen, R. H., and Link, R.: 2000, *ApJ,* **529**, 1034

Pickett, B. K., Mejía, A. C., Durisen, R. H., Cassen, P. M., Berry, D. K., and Link, R. P.: 2003, *ApJ,* **590**, 1060

Piétu, V., Dutrey, A., and Guilloteau, S.: 2007, *A&A,* **467**, 163

Piétu, V., Guilloteau, S., and Dutrey, A.: 2005, *A&A,* **443**, 945

Pringle, J. E.: 1981, *ARA&A,* **19**, 137

Rafikov, R. R.: 2005, *ApJ,* **621**, L69

Rafikov, R. R.: 2007, *ApJ,* **662**, 642

Rayleigh: Lord 1880, *Scientific Papers*, Vol. 1, 474, Cambridge University Press, Cambridge, UK

Reddy, S., Schmid, P. J., and Henningson, D.: 1993, *SIAM J. Appl. Math.,* **53**, 15

Reddy, S. C. and Henningson, D. S.: 1993, *Journal of Fluid Mechanics,* **252**, 209

Reshotko, E.: 2001, *Physics of Fluids,* **13**, 1067

Rice, W. K. M., Armitage, P. J., Bate, M. R., and Bonnell, I. A.: 2003, *MNRAS,* **339**, 1025

Rice, W. K. M., Lodato, G., and Armitage, P. J.: 2005, *MNRAS,* **364**, L56

Rice, W. K. M., Lodato, G., Pringle, J. E., Armitage, P. J., and Bonnell, I. A.: 2004, *MNRAS,* **355**, 543

Rice, W. K. M., Lodato, G., Pringle, J. E., Armitage, P. J., and Bonnell, I. A.: 2006, *MNRAS,* **372**, L9

Rice, W. K. M., Mayo, J. H., and Armitage, P. J.: 2010, *MNRAS,* **402**, 1740

Richard, D.: 2001, *Ph.D. thesis*, Observatoire de Paris GIT/SPEC CEA Saclay

Richard, D., Dauchot, O., and Zahn, J.-P.: 2001, in *12th Couette-Taylor Workshop, Evaston, USA*

Richard, D. and Zahn, J.: 1999, *A&A,* **347**, 734

Rincon, F., Ogilvie, G. I., and Cossu, C.: 2007, *A&A,* **463**, 817

Romeo, A. B.: 1992, *MNRAS,* **256**, 307

Romeo, A. B.: 1994, *A&A,* **286**, 799

Rozyczka, M. and Spruit, H. C.: 1993, *ApJ,* **417**, 677

Ruden, S. P. and Lin, D. N. C.: 1986, *ApJ,* **308**, 883

Ruden, S. P., Papaloizou, J. C. B., and Lin, D. N. C.: 1988, *ApJ,* **329**, 739

Ruden, S. P. and Pollack, J. B.: 1991, *ApJ,* **375**, 740

Rüdiger, G., Tschäpe, R., and Kitchatinov, L. L.: 2002, *MNRAS,* **332**, 435

Ryu, D. and Goodman, J.: 1992, *ApJ,* **388**, 438

Salmeron, R. and Wardle, M.: 2003, *MNRAS,* **345**, 992

Sano, T., Miyama, S. M., Umebayashi, T., and Nakano, T.: 2000, *ApJ,* **543**, 486

Sano, T. and Stone, J. M.: 2002, *ApJ,* **577**, 534

Sargent, A. I. and Beckwith, S.: 1987, *ApJ,* **323**, 294

Schmid, P. J.: 2000, *Physics of Plasmas,* **7**, 1788

Schmid, P. J. and Henningson, D. S.: 2000, *Stability and transition in shear flows*, Springer-Verlag, Berlin

Schmiegel, A. and Eckhardt, B.: 1997, *Physical Review Letters,* **79**, 5250

Sellwood, J. A. and Carlberg, R. G.: 1984, *ApJ,* **282**, 61

Shakura, N. I. and Sunyaev, R. A.: 1973, *A&A,* **24**, 337

Shakura, N. I., Sunyaev, R. A., and Zilitinkevich, S. S.: 1978, *A&A,* **62**, 179

Shen, Y., Stone, J. M., and Gardiner, T. A.: 2006, *ApJ,* **653**, 513

Shtemler, Y. M., Mond, M., Rüdiger, G., Regev, O., and Umurhan, O. M.: 2010, *MNRAS,* **406**, 517

Shu, F.: 1968, *Ph.D. thesis*, Harvard University, Cambridge, Massachusetts, USA

Shu, F., Najita, J., Galli, D., Ostriker, E., and Lizano, S.: 1993, in E. H. Levy & J. I. Lunine (ed.), *Protostars and Planets III*, pp 3–45

Shu, F. H.: 1974, *A&A,* **33**, 55

Shu, F. H., Adams, F. C., and Lizano, S.: 1987, *ARA&A,* **25**, 23

Sicilia-Aguilar, A., Hartmann, L., Calvet, N., Megeath, S. T., Muzerolle, J., Allen, L., D'Alessio, P., Merín, B., Stauffer, J., Young, E., and Lada, C.: 2006, *ApJ,* **638**, 897

Simon, M., Dutrey, A., and Guilloteau, S.: 2000, *ApJ,* **545**, 1034

Simon, M. and Prato, L.: 1995, *ApJ,* **450**, 824

Stamatellos, D. and Whitworth, A. P.: 2008, *A&A,* **480**, 879

Stone, J. M. and Balbus, S. A.: 1996, *ApJ,* **464**, 364

Stone, J. M., Hawley, J. F., Gammie, C. F., and Balbus, S. A.: 1996, *ApJ,* **463**, 656

Stone, J. M. and Norman, M. L.: 1992, *ApJS,* **80**, 753

Strom, K. M., Newton, G., Strom, S. E., Seaman, R. L., Carrasco, L., Cruz-Gonzalez, I., Serrano, A., and Grasdalen, G. L.: 1989, *ApJS,* **71**, 183

Taylor, G. I.: 1936, *Royal Society of London Proceedings Series A,* **157**, 546

Terebey, S., Shu, F. H., and Cassen, P.: 1984, *ApJ,* **286**, 529

Tevzadze, A. G.: 2006, *Ph.D. thesis*, Katholieke Universiteit Leuven

Tevzadze, A. G., Chagelishvili, G. D., Bodo, G., and Rossi, P.: 2010, *MNRAS,* **401**, 901

Tevzadze, A. G., Chagelishvili, G. D., and Zahn, J.: 2008, *A&A,* **478**, 9

Tevzadze, A. G., Chagelishvili, G. D., Zahn, J., Chanishvili, R. G., and Lominadze, J. G.: 2003, *A&A,* **407**, 779

Thomson, W.: 1887, *Philosophical Magazine,* **24**, 188

Tillmark, N. and Alfredsson, P.: 1996, in S. Gavrilakis, L. Machiels & P. Monkewitz (ed.), *Advances in Turbulence VI*, p. 391, Kluwer

Toomre, A.: 1964, *ApJ,* **139**, 1217

Toomre, A.: 1981, in S. M. Fall & D. Lynden-Bell (ed.), *Structure and Evolution of Normal Galaxies*, pp 111–136

Trefethen, L. N., Trefethen, A. E., Reddy, S. C., and Driscoll, T. A.: 1993, *Science,* **261**, 578

Turner, N. J. and Sano, T.: 2008, *ApJ,* **679**, L131

Turner, N. J., Sano, T., and Dziourkevitch, N.: 2007, *ApJ,* **659**, 729

Ulrich, R. K.: 1976, *ApJ,* **210**, 377

Umebayashi, T.: 1983, *Progress of Theoretical Physics,* **69**, 480

Umebayashi, T. and Nakano, T.: 1981, *PASJ,* **33**, 617

Umurhan, O. M. and Regev, O.: 2004, *A&A,* **427**, 855

van Leer, B.: 1977, *J. Comp. Phys.,* **23**, 276

Vanneste, J.: 2008, *Journal of Atmospheric Sciences,* **65**, 1622

Vanneste, J. and Yavneh, I.: 2004, *Journal of Atmospheric Sciences,* **61**, 211

Velikhov, E. P.: 1959, *J. Exp. Theor. Phys. (USSR),* **36**, 1398

von Weizsäcker, C. F.: 1948, *Z. Naturforschung,* **3**, 524

Vorobyov, E. I. and Basu, S.: 2006, *ApJ,* **650**, 956

Vorobyov, E. I. and Basu, S.: 2007, *MNRAS,* **381**, 1009

Vorobyov, E. I. and Basu, S.: 2008, *ApJ,* **676**, L139

Vorobyov, E. I. and Basu, S.: 2010, *ApJ,* **714**, L133

Wada, K., Meurer, G., and Norman, C. A.: 2002, *ApJ,* **577**, 197

Waleffe, F.: 1997, *Physics of Fluids,* **9**, 883

Waleffe, F.: 2003, *Physics of Fluids,* **15**, 1517

Weidenschilling, S. J.: 1977, *MNRAS,* **180**, 57

Whitney, B. A., Kenyon, S. J., and Gomez, M.: 1997, *ApJ,* **485**, 703

Williams, J. P. and Cieza, L. A.: 2011, *Arxiv e-prints: astro-ph/1103.0556*

Wilner, D. J., Ho, P. T. P., Kastner, J. H., and Rodríguez, L. F.: 2000, *ApJ,* **534**, L101

Wolk, S. J. and Walter, F. M.: 1996, *AJ,* **111**, 2066

Yecko, P. A.: 2004, *A&A,* **425**, 385

Yorke, H. W. and Bodenheimer, P.: 1999, *ApJ,* **525**, 330

Yorke, H. W., Bodenheimer, P., and Laughlin, G.: 1993, *ApJ,* **411**, 274

Yoshida, Z.: 2005, *Physics of Plasmas,* **12**, 024503

Zahn, J., Toomre, J., Spiegel, D. O., and Gough, E. A.: 1974, *Journal of Fluid Mechanics,* **64**, 319

Zhu, Z., Hartmann, L., and Gammie, C.: 2009, *ApJ,* **694**, 1045