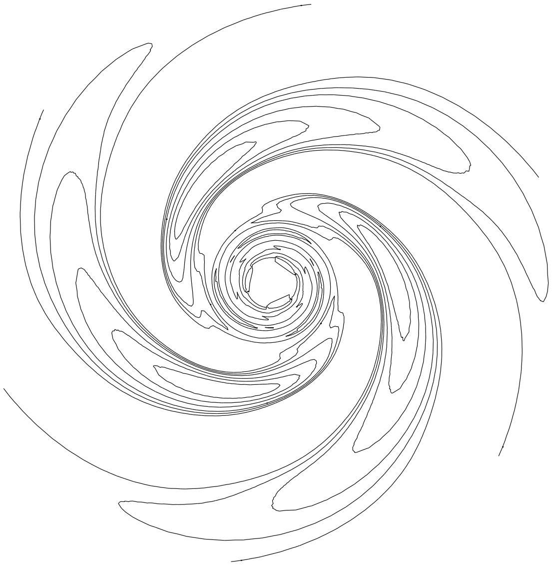
The Stability of Model Disk Galaxies

Jennifer Read,
Merton College



Theoretical Physics,
University of Oxford

The Stability of Model Disk Galaxies

Jennifer C. A. Read, Merton College, Oxford

A thesis submitted for the degree of Doctor of Philosophy

September 1997

Abstract

A linear, normal-mode stability analysis is carried out for a particular class of galaxy models, the power-law disks. These are infinitesimally thin, axisymmetric stellar disks of infinite extent. The rotation curve can be rising, falling or flat (this last case was investigated by Toomre and Zang in 1976). The self-consistent power-law disks are scale-free: all quantities scale as a power of radius. This symmetry leads to interesting but unusual stability properties, which are described in some detail. Additionally, the thesis studies cut-out power-law disks, in which stars close to the origin (and sometimes also stars at great distance) are prevented from participating in any disturbance. A homogeneous Fredholm integral equation is derived relating the response of such disks to a perturbation imposed upon them. When the response density is equal to the imposed density, the solution of the integral equation is a normal mode. Numerical algorithms to isolate the growth rates and the pattern speeds of the normal modes are developed.

The stability to axisymmetric perturbations is found to agree well with the local criterion derived by Toomre in 1964. The cut-out disks are remarkably stable to bisymmetric perturbations, but they are very susceptible to one-armed ($m = 1$) modes. Physical reasons for the prevalence of one-armed instabilities are put forward, together with a description of the modes in terms of swing-amplified, reflected travelling waves. Growing three-armed and four-armed modes occur only at very low temperatures. However, neutral $m = 3$ and $m = 4$ modes are possible at higher temperatures for some disks. They are suppressed by the existence of closed m-lobed orbits. The rotation curve has a marked effect on the stability. For every azimuthal wavenumber, disks with rising rotation curves are less stable than those with falling. This is because disks with rising rotation curves exhibit less shear (or, equivalently, are closer to solid-body rotation). The maintenance of any spiral pattern is helped as the shear is minimised. The abruptness with which the central stars are carved out also influences the stability. In every case, disks where the cut-out is gentler are more stable.

Acknowledgements

First of all I would like to thank my supervisor, Wyn Evans, for his kindness and patience over the past three years. He has spent so much time and trouble guiding me through this project. I have really enjoyed our long discussions. I have also greatly appreciated his advice on many occasions when I was pondering what to do after this D. Phil. I am very grateful to Jim Collett for all his help, for welcoming me when I arrived and for always being encouraging. My thanks to David Earn and Jerry Sellwood for kindly loaning us their code, and assisting us in its use. I am grateful to the Particle Physics and Astrophysics Research Council for supporting me during my D. Phil. I would never have got to grips with the computer system without the assistance of Andrew McMurry, who would be a rich man had he charged me consultation fees by the hour. I must also thank Lory Rice for resuscitating my computer on numerous occasions, and for being generous with disk quota.

My other office-mates and colleagues in the Theoretical Physics department – especially Tetsuo Amaya, Andy Green, Neil Griffiths, Ralf Häfner, Carole Jordan, Keith MacPherson, John Magorrian, Debbie Philippides, Magnus Richardson, Stuart Sim, Serge Tabachnik and Mark Wilkinson – have enlivened many a coffee-time and generally enhanced my time here. I thank my old tutorial partner Mark Venables for many supportive emails. I must also mention the other members of the Aylesbury Five, and indeed the rest of Taruithorn.

It is impossible for me adequately to thank my family – my parents, sister and parents-in-law –for all they have done for me over the years. Most of all, I have to thank my dear husband Marc, for generally being wonderful, cheering me up and patiently learning a lot more than he ever wished to know about power-law disks.

> THE HEAVENS ARE TELLING
> THE GLORY OF GOD; AND THE
> FIRMAMENT PROCLAIMS HIS
> HANDIWORK. DAY TO DAY POURS
> FORTH SPEECH, AND NIGHT TO
> NIGHT DECLARES KNOWLEDGE.
> THERE IS NO SPEECH, NOR ARE
> THERE WORDS; THEIR VOICE IS
> NOT HEARD: YET THEIR VOICE
> GOES OUT THROUGH ALL THE
> EARTH, AND THEIR WORDS
> UNTO THE ENDS OF THE WORLD.
>
> PSALM XIX — COELI ENARRANT — 1-4

Contents

Chapter 1
Introduction

1.1 The observations

Spiral structure in nebulae was first observed by William Parsons, Third Earl of Rosse (1850). Many people had already speculated that the nebulae were island universes, galaxies like our own, but it was not until 1923 that Hubble settled the matter. He was able to resolve stars in M31 and M33; assuming that they were comparable to stars in our Galaxy, he concluded that these systems must be vast distances away.

Hubble developed the most widely used system for the classification of galaxies. He divided spiral galaxies into two groups: the barred spirals SB, and S (or SA), in which there is no bar. He subdivided each group into a, b, and c. Galaxies of type a have smoother and more tightly wound arms, less gas, a bigger nuclear bulge and more total mass (e.g. Rubin *et al.* 1978) than those of type c. Hubble referred to type a as early, and type c as late, since he thought that spirals evolved from a to c. These terms are still used, although the temporal interpretation has been abandoned. Various refinements to Hubble's scheme have been introduced, such as that of Bruce and Debbie Elmegreen (1982), which classifies spirals according to the degree of grand design visible in their spiral patterns.

The surface brightness of many disk galaxies is an exponential function of radius: $I = I_0 \exp(-R/R_d)$. The scale length R_d is typically in the range 1-50 kpc; our own Galaxy's is about 4kpc. Most stars have velocities from 100-400 km/s; very little of this motion is out of the plane of the disk, so spiral disks are very thin. The thickness varies for different populations; in our Galaxy it is 50pc for young stars and 700pc for older ones (Mihalas & Binney 1987, p. 252). Within the plane, the stars' motion is mainly circular; the deviation from mean circular motion is only about 10-50km/s or about 10%. High-velocity objects such as globular clusters and old stars are also observed. These belong to the old, metal-poor spheroidal halo.

The stellar disk is formed out of stars which condensed from the cold gas component at different times.

The brightest stars (O, B) are those which formed most recently (10-15 Myr ago); their distribution and kinematics are thus similar to those of the current cold interstellar medium. Gas and young stars are together known as Population I. This appears bluer and exists in a very thin layer in the disk. Older stars (Population II) appear redder and are distributed more thickly throughout the disk. The spiral patterns of the two populations differs, When a galaxy is viewed in blue light (Population I), its arms appear thin and ragged, with many clumps and knots. In red light (Population II), the image is much smoother, and the arms appear broader.

The total mass in a spiral galaxy is typically 10^{10} - 10^{12} solar masses, of which less than 5% is in the form of gas. Observations of the Doppler shift of emission lines in the gas can be used to derive the galaxy's rotation curve. Optical transitions in warm gas, or the 21cm radio line in cold hydrogen can both be used. The latter is probably more accurate, and can trace the rotation curve outside the optical disk. Rotation curves deduced in this way are proportional to radius in the central region of the galaxy, indicating solid-body rotation in the core. At a certain radius the curves turn over, and become remarkably flat. They continue to be so well outside the optical disk, indicating the presence of a great deal more mass than is visible in the disk stars. Many disk galaxies – such as NGC1300 (Sandage 1961, p. 45) – are barred, i.e. the spiral arms extend from the ends of a long bar which passes through the nucleus.

1.2 Theories of spiral structure

Disk galaxies are not rigid bodies, but show differential rotation. Stars further out tend to circle more slowly than stars close to the centre of the galaxy. Thus at first sight spirality seems natural. Any inhomogeneity which arises in the disk is sheared out into a spiral, like cream stirred into a cup of coffee. In both coffee and galaxies, this spirality rapidly disappears as the differential rotation continues to wind up the inhomogeneity. The possibility was briefly considered that the different Hubble types represent galaxies at different stages of this winding process, with loosely-wound galaxies of Hubble type c being wound up into tightly-wound galaxies of Hubble type a. However, it became clear that other determinants of Hubble type, such as gas content and bulge size, do not change on the rapid timescale implied by the winding.

Flocculent or chaotic spiral structure may well be formed by the local instability introduced by Goldreich & Lynden-Bell (1965a), Julian & Toomre (1966), and termed "swing amplification" by Toomre (1981). These authors noticed that a disk responds vigorously to the non-axisymmetric gravity field of orbiting mass concentrations and gas clumps. Such lumps gather a wake, both fore and aft, which is sheared into a spiral arm. The clumping is periodically regenerated – perhaps by gravitational instability, perhaps by infall of gas. So, the inhomogeneities continually arise and contribute briefly to the ragged spiral structure before growing dim and being smeared out by differential rotation.

However, some galaxies exhibit grand design spirals, with arms extending across the entire galaxy. In these, some sort of global mechanism seems to be at work. To avoid the winding problem, a natural step is to suppose that the spiral pattern is caused by a density wave. This rotates as a solid body, while stars move into and out of the pattern. This explains why bright young stars, molecular clouds and HII regions are found in the spiral arms. The gas is compressed as it passes through the arms, triggering star formation. The lifetimes of the bright objects is short compared to the rotation period of the arms, so they fade away as the density wave moves on.

It is still unclear how such a wave might be caused, or whether it could persist as a self-consistent mode of oscillation in the disk for tens of rotations, without regeneration (Toomre 1969). The approach of Lin and Shu (1964) was simply to hypothesise that a density wave existed and was quasi-stationary, and to examine the consequences of this hypothesis. This has been valuable in stimulating research, and has had a number of successes. For instance, Visser compared the observed kinematics of atomic hydrogen in M81 with those predicted by a model based on the Lin-Shu hypothesis. Using only the pattern speed of the spiral as a fitting parameter, he was able to obtain impressive agreement (Visser 1980). It seems beyond dispute that the spiral structure in many galaxies represents a density wave of some kind.

However, the hypothesis that the density wave is quasi-stationary cannot be regarded as proved. Models in which the density wave is a transient phenomenon can be equally successful in matching the observations. One possibility that goes back at least as far as Chamberlin (1901) is that the spiral arms are caused by tidal forces from a companion. The density wave is then just a transient wave generated by the rattling and shaking of the galactic disk caused by the close approach. The classic example is M51, which has dramatic grand-design spiral arms and a small companion NGC 5195 (Toomre & Toomre 1972; Toomre 1995). Toomre (1981; 1995) has also modelled the spirality of M81 by a tidal encounter with the nearby M82. Numerical integrations show that the swing-amplified growth of the transient wave can lead to a very fine, if temporary, grand-design spiral. Toomre's match to the observations is as impressive as that of Visser (1980) with his quasi-stationary density wave. It remains to be shown whether suitable encounters with other galaxies are common enough for tidal models to explain all the grand-design spirals. However, their success indicates that the quasi-stationary hypothesis is assuredly not the only possibility.

1.3 Global stability analyses

The best hope for generating quasi-steady spiral patterns of finite amplitude seems to be mildly unstable discrete spiral modes (if they exist). This motivates the study of the overall or global stability of galactic disks. Research workers have used two approaches – first, direct N-body work, and second, normal mode analyses.

The first approach led to a big surprise in the early 70s. Let us recall that cold disks of stars – i.e. disks in which all the stars are on circular orbits – are violently unstable to the Jeans instability: they break up into rings. Toomre (1964), approximating a disk as an infinitesimally thin rotating sheet of infinite extent, showed that a minimum amount of random motion is needed for the disk to be at least locally stable to these modes. This is conveniently expressed in terms of his stability parameter Q, which is the ratio of the actual velocity dispersion at a point in the disk to that needed for local axisymmetric stability. Thus disk galaxies must have $Q > 1$ to avoid breaking up into hoops. Studies (Julian & Toomre 1966) seemed to indicate that a rotating sheet is always locally stable to self-consistent non-axisymmetric perturbations. In the late 60s it was presumed that local theory is a reliable indicator of global stability, i.e. disks with $Q \geq 1$ should be totally stable.

However, the surprise came when numerical studies indicated that disks with $Q = 1$ retain violent bisymmetric instabilities. A classic study is that of Hohl (1971), who studied a disk of stars with $Q = 1$. The disk evolved rapidly into a striking two-armed spiral structure, and subsequently settled into an axisymmetric state with $Q >> 1$. High values of Q were required to stabilise the disk to bisymmetric perturbations. However, real galaxies appear to be stable with much lower values of Q, at least judging from the observed velocity dispersions. This tendency towards global bar-making has been confirmed by extensive N-body work throughout the subsequent decades (Athanassoula & Sellwood 1986; Sellwood & Wilkinson 1993). N-body simulations have driven much of our understanding of the stability of galactic disks. Even so, the noise in N-body simulations is some orders of magnitude greater than in real stellar systems, so care is needed in interpreting the results of these studies.

An alternative approach is to study the instabilities of model disks via a linear normal mode analysis, as has been advocated particularly by Kalnajs (1971). An analogy may help here. The normal modes of a system such as a set of coupled pendula are defined to be those oscillations in which all the pendula have the same frequency. In a linear analysis, one obtains a set of coupled linear differential equations describing the motion of the system in the small-amplitude regime where $\sin\theta \approx \theta$. One then substitutes in a solution of the form $\theta_i = A_i \exp(i\omega t)$. The allowed values of ω represent the normal modes of the system. Associated with each mode are the relative values of the amplitudes A_i. A general solution is given by a superposition of many normal modes. Similarly, in a galactic stability analysis, one first obtains a linear differential equation describing the evolution of the system. This is the linearised collisionless Boltzmann equation, which, like $\sin\theta \approx \theta$ for the pendula, is valid for small perturbations from equilibrium. The normal modes are then found by substituting in a solution with time-dependence given by $\exp(i\omega t)$. As for the pendula, only certain values of ω are allowed, and they are each associated with a particular shape of density perturbation.

For the set of coupled pendula, there are as many normal modes as there are degrees of freedom in the system. We can be confident that the normal modes represent a complete set of solutions, and that any possible motion of the pendula can be represented by a suitable combination of the normal modes. For a galaxy, we do not have this confidence. Toomre (1977) cautioned that the spectrum of normal modes of stellar systems must include delta-functions or van Kampen modes (van Kampen 1955; Clemmow & Dougherty 1969, p. 258) near any resonance radii. For the closely-related problem of the stability of differentially-rotating stars, the completeness of the normal modes, together with the van Kampen modes, has been proved (Dyson & Schutz 1979). In stellar dynamics, the normal modes have not been shown to be complete.

Few normal-mode studies have been carried out. The first was the stability analysis of the Maclaurin spheroids, carried out over a century ago by Bryan (1889). These are rigidly-rotating homogeneous fluid bodies. They are remarkably stable, becoming unstable only when spinning very rapidly and thus highly flattened. The first dynamical instability to which they fall prey is a bar-like $m = 2$ mode. Similar results were found for the flat Maclaurin fluid disks (Hunter 1963). Kalnajs (1972) analysed the stellar analogues of the Maclaurin disks. These are rigidly rotating axisymmetric disks with sharp edges. Like the fluid disks, they become unstable to $m = 2$ modes if they rotate too rapidly. The instability sets in at much lower angular velocities: the stellar disks are considerably more unstable than the fluid disks. All these systems are comparatively simple to analyse because they are rigidly rotating. Differentially-rotating disks are much harder (Vauterin & Dejonghe 1996).

The present thesis is based on the pioneering work of Toomre and Zang (1976). They carried out a detailed linear stability analysis of a class of differentially-rotating, infinite disks in which the circular velocity is independent of radius. The equilibrium disk was self-similar, but stars in its central regions were prevented from taking part in the disturbance. Effectively the centre of the disk was cut out. In accordance with numerical studies, they found that Toomre's local stability criterion was remarkably accurate in predicting the global stability to axisymmetric modes. However, somewhat surprisingly, they found that the disk was very stable to bisymmetric modes. Random motion corresponding to $Q = 1$ was already sufficient to stabilise the disk to all $m = 2$ disturbances. This greatly surprised Toomre and Zang, since almost all other studies had suggested the bisymmetric modes would be the hardest to stabilise. In fact, the Toomre-Zang disks turned out to be highly susceptible to one-armed instabilities. These modes showed few signs of disappearing as the temperature was increased, although their growth rate slowed to vanishing.

1.4 Purpose of the thesis

The aim of this thesis is to investigate the large-scale, global, linear modes of a family of horizontally hot, vertically cold, idealised disk galaxies, with rising, falling or flat rotation curves. These include as a special case the model studied by Toomre and Zang.

In any study of this kind, the degree of realism enjoyed by a galaxy model must be balanced with the requirement of analytical tractability. Thus analytical approaches often use less physically plausible models than numerical studies. The two approaches complement each other in their common aim of gaining insight into the types of instabilities to which galaxies may be prone. Where they agree, the analytical work enables one to be confident that the numerical results are not an artefact of the various procedures used to simulate the stars' behaviour; conversely, the numerical work confirms that the analytical result does not change significantly when a more realistic representation is used.

The first simplification imposed by our analysis is to describe the billions of stars in a galaxy by a single distribution function. We thus deal in probabilities rather than actualities – a reasonable approximation where such huge numbers are concerned. Chapter 2 of this thesis introduces the details of the family under consideration, the so-called power-law disks. We make the assumption of zero thickness usually employed in modelling galactic disks (Toomre 1966). This seems a good working hypothesis, given the extreme flatness of real disk galaxies - our own Galaxy's radius is at least a hundred times its thickness.

The equilibrium power-law disks are self-similar, so all dynamical quantities vary as a power of radius. The rotation curve is described by $v_{\rm circ} \propto R^{-\beta/2}$. We shall principally consider disks with $-0.5 < \beta < 0.5$. The rotation curves of real galaxies are approximately flat across a wide extent of the galaxy, but both rising and falling rotation curves are also observed. An aim of this thesis is to understand how the slope of the rotation curve influences the stability.

The self-similarity of the equilibrium disks is analytically extremely important. It enables much of the calculation to be carried out exactly. However, the perfectly self-similar disk is something of a special case. Following Toomre and Zang, we have therefore concentrated on a modified version of the pure power-law disks, in which the central regions of the disk are immobilised or cut out. This was achieved by reducing the fraction of active stars – those able to respond to a disturbance – from unity at moderate radii to zero at the centre. In order to aid comparison of our results with the numerical simulations of Earn (1993), we have also tapered the fraction of active stars to zero at large radii. Our originally infinite disk has thus become an annulus. Provided that the cut-out is not too abrupt, this is perhaps more realistic than an infinite disk with a surface density that diverges at the origin. The immobile core may be regarded as representing the hot bulge of real disk galaxies. The immobile components still contribute to the potential experienced by the active stars. Thus the cut-out disks are not self-consistent: the potential they experience is greater than that due to their own density.

Having investigated the equilibrium properties of the power-law disks, we proceed in Chapter 3 to examine their response to disturbance. We arrive at an integral equation which must be satisfied by self-consistent modes. This is the fundamental equation in the thesis. It enables us to deduce whether a given disk is stable or unstable to growing disturbances. We then pause, in Chapter 4, to consider the special case of the self-consistent, scale-free disks. We derive a response function describing the response of the self-consistent disk to neutral perturbations (that is, perturbations which neither grow nor rotate). Chapter 5 describes the various numerical methods adopted for the evaluation of the response function and the solution of the integral equation, a homogeneous Fredholm integral equation of the second kind. Checks on the numerics are described. We discuss the results of the analysis, investigating modes with each azimuthal symmetry in turn. Chapter 6 describes the stability of both the self-similar and the cut-out disks to axisymmetric ($m = 0$) modes. Chapter 7 describes the stability to bisymmetric ($m = 2$) modes. Chapters 8 and 9 discuss the disks stability to one-armed ($m = 1$) modes, and to higher angular harmonics ($m = 3, 4$).

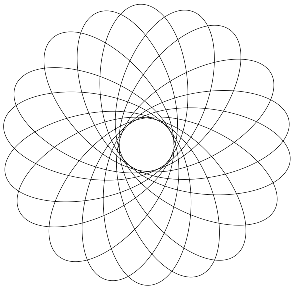
Questions to be answered include:

- How much random motion is required to stabilise a disk to modes of each angular harmonic?
- How does stability vary with rotation curve? Specifically, are galaxies with rising rotation curves more or less stable than galaxies with falling curves?
- Do the scale-free self-consistent disks admit modes? There are contradictory speculations in the literature (Zang 1976; Lynden-Bell & Lemos 1996).

- What kinds of density laws can be carved out to render models stable (or unstable)?
- How does stability vary with outer truncation?
- How do rules of thumb such as Toomre's local criterion and the Ostriker-Peebles global $T_{\rm rot}/|W|$ criterion fare in comparison with an analytical stability analysis?
- What kind of modes occur when the random motion is too small for stability?

Chapter 2

Equilibrium Properties of the Power-Law Disks



2.1 Introduction

In this chapter, we derive the basic properties of our galaxy model. These are infinitesimally thin disks in which the circular velocity varies as a power of R. The special case in which the rotation curve is flat and the stars move on circular orbits is the Mestel (1963) disk. The hot stellar dynamical disks with the potential and density of the Mestel disk are referred to as the Toomre-Zang disks in honour of the pioneering work in Zang's (1976) Ph.D. thesis. The self-similarity of these disks simplifies the analysis considerably. Extending the work of Zang, we introduce new variables (functions of the energy and angular momentum), which can be used to characterise orbits according to their shape and size. We can then exploit the self-similarity to deal with all orbits of the same shape together. It is also useful to "cut out" the centre of the disk, and to taper the disk at large radii. Both these require changes to the distribution function. We discuss the modifications necessary to the distribution function and their effect on the surface density.

Many quantities will be introduced in this chapter and used throughout this thesis. A summary of all the dynamical quantities and their dimensionless analogues, as well as the distribution functions used, is provided in Appendix A.

2.2 The self-similar power-law disks

The general power-law disks are infinitesimally thin disks, of infinite extent. For a disk which has surface density Σ_0 at a reference radius R_0, we have

$$\Sigma_{\rm eq} = \Sigma_0 \left(\frac{R_0}{R}\right)^{1+\beta}. \tag{2.1}$$

Schmitz & Ebert (1987) and Lemos *et al.* (1991) have studied the global stability to axisymmetric perturbations of gaseous disks with this density profile. This work has been extended to non-axisymmetric distortions by Syer & Tremaine (1996) and Lynden-Bell & Lemos (1996).

Fig. 2.1 shows the surface density for seven values of β.

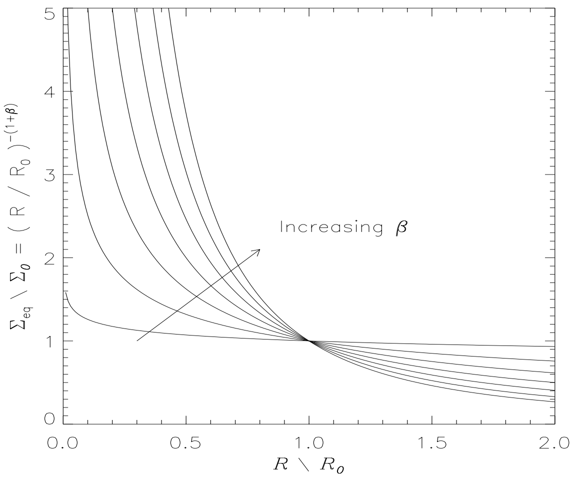


Figure 2.1: Surface density against radius for $\beta = -0.9, -0.6, -0.3, 0.0, +0.3, +0.6, +0.9$

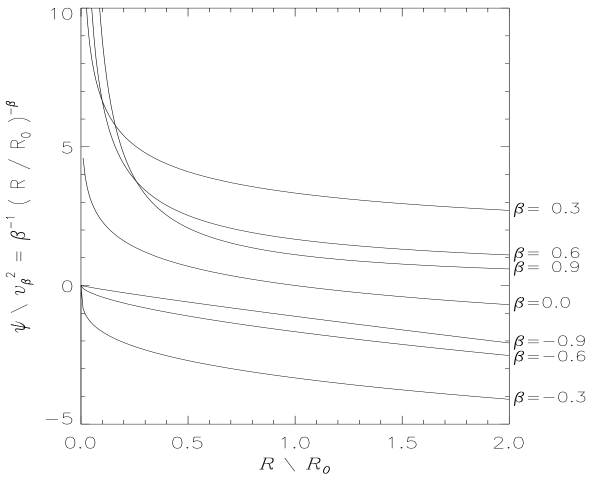


Figure 2.2: Potential against radius for $\beta = -0.9, -0.6, -0.3, 0.0, +0.3, +0.6, +0.9$

The case $\beta = 0$ is the Mestel (1963) disk, which was investigated by Toomre (1975) and Zang (1976). In the following chapters, we frequently wish to compare our results to those of Zang. The notation Z followed by an equation number is used as a convenient shorthand for reference to Zang's (1976) thesis.

We use the sign convention in which the potential ψ has opposite sign to the potential energy ϕ. In this convention, Poisson's equation is $\nabla^2\psi = -4\pi G\rho$, where ρ is the density. To find the self-consistent potential corresponding to this density, we first look for solutions satisfying Laplace's equation $\nabla^2\psi = 0$ outside the disk. We obtain a set of solutions $\psi_k(R, z \neq 0) = e^{-k|z|}J_0(kR)$, where k can take any value from 0 to ∞ (see e.g. Binney & Tremaine 1987). The self-consistent potential is constructed from these solutions, and must generate the required surface density in the disk. Applying Gauss' theorem to a small box enclosing part of the disk, we require

$$\left.\frac{d\psi_{z>0}}{dz}\right|_{z=0} - \left.\frac{d\psi_{z<0}}{dz}\right|_{z=0} = -4\pi G\Sigma_{\rm eq}(R, z=0). \tag{2.2}$$

In this way, the potential in the plane of the disk is found to be (Schmitz & Ebert 1987; Lemos *et al.* 1991; Evans 1994)

$$\psi(R) = \frac{v_\beta^2}{\beta}\left(\frac{R_0}{R}\right)^\beta, \tag{2.3}$$

where the *reference velocity* v_β is defined by (cf. Z2.3)

$$v_\beta^2 = 2\pi G\Sigma_0 R_0 \frac{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}. \tag{2.4}$$

Clearly (2.3) fails for $\beta = 0$; in this case we obtain (Mestel 1963; Z2.5)

$$\psi(R) = -v_0^2 \ln\left(\frac{R}{R_0}\right). \tag{2.5}$$

This can also be derived from (2.3) by adjusting the zero of potential and using the result

$$\lim_{\beta\to 0} \frac{1}{\beta}\left(x^{-\beta} - 1\right) = -\ln x.$$

Henceforth it is to be understood that when an expression fails at $\beta = 0$, the corresponding result for the Mestel or Toomre-Zang disk must be used. It can often be derived by taking the limit $\beta \to 0$ with l'Hôpital's theorem. Fig. 2.2 shows the potential for seven values of β.

With the definition of potential given in (2.3), the zero of potential is different for rising ($\beta < 0$) and falling ($\beta > 0$) rotation curves. This has consequences for the energy E of a star. We have $E = T + \phi(R)$, where T is the kinetic energy, and $\phi(R) = -\psi(R)$ is the potential energy. The range of R open to the star is limited by the requirement

$$\phi(R) \le E. \tag{2.6}$$

For the falling rotation curve ($\beta > 0$), $\phi(R)$ is negative for all R. If E is positive, then the inequality (2.6) is satisfied for all R. The star can roam freely from the origin to infinity, and is not bound to the disk. Conversely, if E is negative, then (2.6) cannot be satisfied for all R. The star is restricted to a finite area of the disk: it is said to be in a bound orbit. Since we require our stars to be bound to the galaxy, we shall use only negative energies when β is positive. For the rising rotation curve ($\beta < 0$), $\phi(R)$ is positive for all R, so the inequality (2.6) can be satisfied only for positive energy. The rule is, then, that a bound star orbiting in a disk with parameter β has energy of opposite sign to β. For the Mestel disk, in which $\beta = 0$, the orbit is bound whatever the sign of the energy.

The circular velocity is (cf. Z2.1)

$$v_{\text{circ}}^2 = \left(\frac{R_0}{R}\right)^{\beta} v_\beta^2. \tag{2.7}$$

Thus the reference velocity v_β is the circular velocity at the reference radius R_0. The circular velocity depends on the radius as $R^{-\beta}$. Disks with *positive* beta have *falling* rotation curves, whereas disks with *negative* beta have *rising* rotation curves.

Galaxies typically have fairly flat rotation curves (e.g. Rubin *et al.* 1978; Mihalas & Binney 1987), and so the potential in the plane can be described at least roughly by the Mestel disk. Though flattish, the rotation curves are often slowly rising or falling at large radii (Casertano & van Gorkom 1991), and this motivates our study of the general power-law disk.

We may choose $R_0 = 1$ without loss of generality. When comparing results between disks with different β, we may choose to keep v_β constant. The reference density Σ_0 is then determined according to (2.4). Alternatively, we may choose to keep Σ_0 constant, in which case (2.4) defines the reference velocity v_β. Figs. 2.3 and 2.4 show these alternative choices. They show the circular velocity $v_{\rm circ}$ plotted against radius for different values of β. In fig. 2.3, $v_\beta = R_0 = 1$; in fig. 2.4, $\Sigma_0 = R_0 = 1$. How quantities change as β is varied depends on the choice of units. Circular velocity is relatively easy to measure observationally, whereas surface density is almost impossible to determine. It thus seems more useful to compare disks with the same circular velocity at a particular radius, rather than with the same surface density there. Throughout this thesis, then, we shall employ units in which $v_\beta = R_0 = 1$.

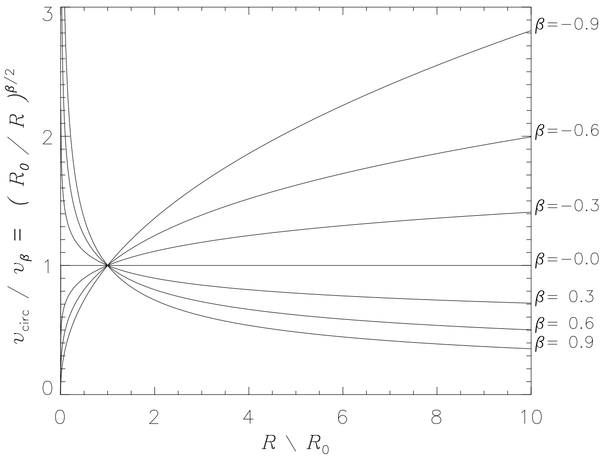


Figure 2.3: Circular velocity against radius, in units where $v_\beta = R_0 = 1$, for β = -0.9, -0.6, -0.3, 0, +0.3, +0.6, +0.9

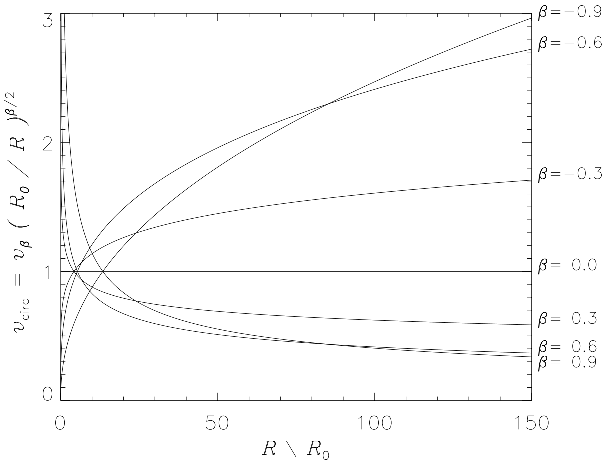


Figure 2.4: Circular velocity against radius, in units where $\Sigma_0 = R_0 = 1$, for β = -0.9, -0.6, -0.3, 0, +0.3, +0.6, +0.9

The mass enclosed within a radius R is

$$M(R) = \frac{2\pi\Sigma_0 R_0^2}{1-\beta}\left(\frac{R}{R_0}\right)^{1-\beta} \tag{2.8}$$

For a flat rotation curve, the enclosed mass depends linearly on radius. For rising rotation curves, the enclosed mass rises more steeply with R; for falling rotation curves, it rises steeply at low radii but then increases only slowly. The total mass of the disk is infinite for all β in the range $-1 < \beta < 1$.

The orbits of stars in the disk

The Lagrangian for the general power-law disk is

$$\mathcal{L} = \frac{1}{2}\dot{R}^2 + \frac{1}{2}R^2\dot{\theta}^2 + \frac{v_\beta^2}{\beta}\left(\frac{R_0}{R}\right)^{\beta}. \tag{2.9}$$

A star moving in this potential obeys the equations of motion

$$\ddot{R} = R\dot{\theta}^2 - \frac{v_\beta^2}{R}\left(\frac{R_0}{R}\right)^\beta, \qquad \frac{d}{dt}\left(R^2\dot{\theta}\right) = 0. \tag{2.10}$$

In terms of the radial velocity $u = \dot{R}$ and the tangential velocity $v = R\dot{\theta}$, the angular momentum is

$$L_z = Rv, \tag{2.11}$$

and the energy is

$$E = \frac{1}{2}\left(u^2 + v^2\right) - \frac{v_\beta^2}{\beta}\left(\frac{R_0}{R}\right)^\beta. \tag{2.12}$$

Following Zang (1976), we define the *home radius* R_H to be the radius at which the tangential velocity is equal to the circular velocity. By conservation of angular momentum we have:

$$L_z = Rv = R_\mathrm{H} v_\mathrm{circ}\left(R_\mathrm{H}\right). \tag{2.13}$$

Substituting for the circular velocity, we have (cf. Z2.26)

$$R_\mathrm{H} = R_0\left(\frac{L_z}{v_\beta R_0}\right)^{\frac{2}{2-\beta}}. \tag{2.14}$$

Again employing Zang's (1976) terminology, we define the *eccentric velocity* U to be the maximum radial speed reached during an orbit. U is thus positive by definition.

We rearrange the above expression for the energy (2.12) to obtain the radial velocity

$$u^2 = 2E - \frac{L_z^2}{R^2} + 2\frac{v_\beta^2}{\beta}\left(\frac{R_0}{R}\right)^\beta. \tag{2.15}$$

Differentiating this with respect to R, we find that the maximum radial velocity occurs at the home radius (2.14), and that its value is given by

$$U^2 = 2E + \left(\frac{2}{\beta} - 1\right)\left(\frac{v_\beta^2 R_0^\beta}{L_z^\beta}\right)^{\frac{2}{2-\beta}}. \tag{2.16}$$

A similar derivation using the potential of the Mestel disk gives (Z2.29)

$$U^2 = 2E - v_0^2\left(1 + 2\ln\frac{L_z}{v_0 R_0}\right). \tag{2.17}$$

As an illustration, let us consider the following orbits, where in each case a star is followed through ten radial oscillations. The star on the left has low U, and so a more nearly circular orbit, whereas the one on the right has high U, and so a more eccentric orbit.

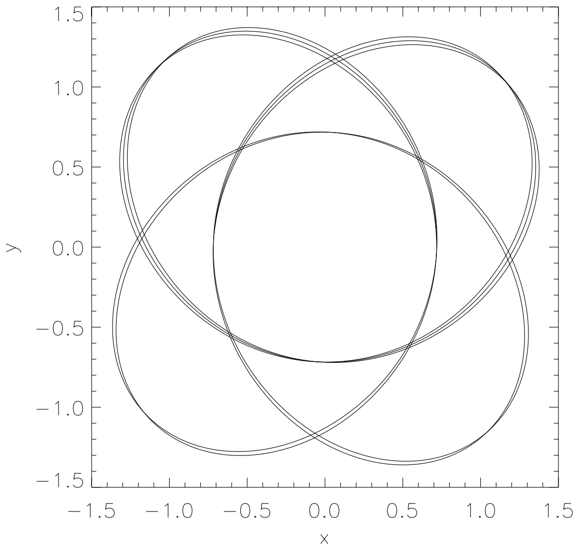


Figure 2.5: The orbit of a star with low eccentric velocity: $E = -3.375$ and $L = 1$, $U = 0.5$ and $R_{\rm H} = 1$, in the $\beta = 0.25$ disk, with $R_0 = v_\beta = 1$

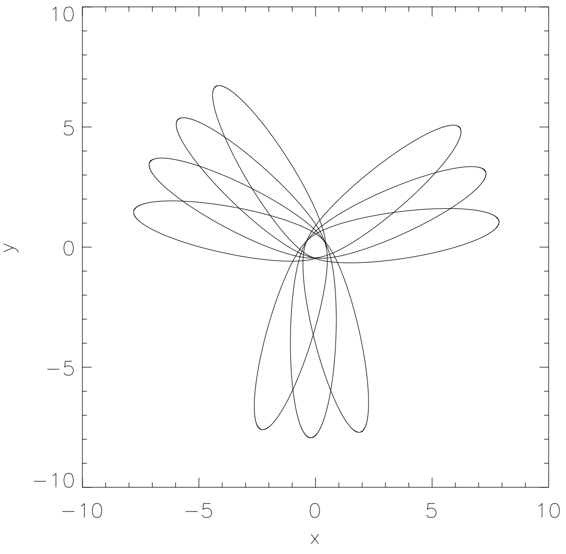


Figure 2.6: The orbit of a star with high eccentric velocity: $E = -2.375$ and $L = 1$, $U = 1.5$ and $R_{\rm H} = 1$, in the $\beta = 0.25$ disk, with $R_0 = v_\beta = 1$

It will frequently be convenient to work in dimensionless co-ordinates. We define the following dimensionless integrals of motion:

$$\tilde{U}^2 = \frac{U^2}{v_\beta^2}\left(\frac{R_{\rm H}}{R_0}\right)^\beta, \qquad \tilde{R}_{\rm H} = \frac{R_{\rm H}}{R_0}, \tag{2.18}$$

$$\tilde{E} = \frac{E}{v_\beta^2}\left(\frac{R_{\rm H}}{R_0}\right)^\beta, \qquad \tilde{L}_z = \frac{L_z}{v_\beta R_0}. \tag{2.19}$$

We define the dimensionless radial and tangential velocities similarly:

$$\tilde{u}^2 = \frac{u^2}{v_\beta^2}\left(\frac{R_{\rm H}}{R_0}\right)^\beta, \qquad \tilde{v}^2 = \frac{v^2}{v_\beta^2}\left(\frac{R_{\rm H}}{R_0}\right)^\beta. \tag{2.20}$$

We shall also use dimensionless radial and time coordinates:

$$\tilde{R} = \frac{R}{R_{\rm H}}, \qquad \tilde{t} = \frac{v_\beta}{R_{\rm H}}\left(\frac{R_0}{R_{\rm H}}\right)^{\beta/2} t. \tag{2.21}$$

These definitions throw some light on the significance of the home radius. As suggested by the definition $\tilde{R} = R/R_{\rm H}$, it is a scale radius for the orbit. The self-similarity of the disk means that orbits of any size are simply scaled copies of each other. All dependence on radius can thus be factored out: the equations of motion become

$$\frac{d^2\tilde{R}}{d\tilde{t}^2} = \frac{1}{\tilde{R}^3} - \frac{1}{\tilde{R}^{1+\beta}}, \qquad \frac{d\theta}{d\tilde{t}} = \frac{1}{\tilde{R}^2}. \tag{2.22}$$

The scaled energy is

$$\tilde{E} = \frac{1}{2}\left(\tilde{u}^2 + \tilde{v}^2\right) - \frac{1}{\beta\tilde{R}^\beta}. \tag{2.23}$$

Equations (2.14) and (2.16) become

$$\tilde{R}_{\mathrm{H}} = \tilde{L}_z^{\frac{2}{2-\beta}}, \qquad \tilde{U}^2 = 2\tilde{E} - 1 + \frac{2}{\beta}. \tag{2.24}$$

We shall also need expressions for the radial and tangential velocities:

$$\tilde{u}^2 = \tilde{U}^2 + 1 - \tilde{R}^{-2} + \frac{2}{\beta}\left(\tilde{R}^{-\beta} - 1\right), \qquad \tilde{v} = \frac{1}{\tilde{R}}. \tag{2.25}$$

As a star orbits in the equilibrium disk, the accessible phase space is limited by its energy and angular momentum (or equivalently its eccentric velocity and home radius). Because of the azimuthal symmetry of the disk, the star circulates through all values of ϕ. It does not explore all values of R, but moves within an annulus of the disk bounded by $R_{\min}$ and $R_{\max}$. For certain values of the eccentric velocity, the orbit closes, and the star then traverses a one-dimensional manifold within the annulus; in general the orbit does not close, and the star eventually passes arbitrarily close to every region of the annulus (see e.g. Landau & Lifshitz 1960).

The turning-points of the orbit occur when the radial velocity $u = 0$. In terms of the dimensionless co-ordinates, this condition corresponds to

$$\tilde{U}^2 + 1 - \tilde{R}^{-2} + \frac{2}{\beta}\left(\tilde{R}^{-\beta} - 1\right) = 0. \tag{2.26}$$

For given $\tilde{U}$ and β, there are two solutions, $\tilde{R}_{\min}$ corresponding to pericentre, and $\tilde{R}_{\max}$ corresponding to apocentre. We solved this equation numerically using the Newton-Raphson method. The choice of initial guess depends on the size of $\tilde{U}$.

To find the maximum, we use the Newton-Raphson method directly on (2.26). For large $\tilde{U}$, the star's orbit is highly eccentric, and $\tilde{R}_{\max}$ is large. We can thus neglect $\tilde{R}^{-2}$, and use $\tilde{R}_{\max} \approx \left[1 - \beta(\tilde{U}^2 + 1)/2\right]^{-1/\beta}$ as an initial approximation. For small $\tilde{U}$, $\tilde{R}_{\max}$ will not greatly exceed unity. We write $\tilde{R} = 1 + \epsilon$, and obtain $\tilde{R}_{\max} \approx 1 + \tilde{U}/\sqrt{2-\beta}$.

To find the minimum, we rewrite (2.26) in terms of $q = 1/\tilde{R}$ and look for the larger root of q. Again the choice of initial value depends on the size of $\tilde{U}$. For large $\tilde{U}$, $\tilde{R}_{\min}$ is small, so $q_{\max}$ is large; we neglect q^{β}, and use $q_{\max} \approx \sqrt{\tilde{U}^2 + 1}$ as a starting-point. For small $\tilde{U}$, we write $q = 1 + \epsilon$, and obtain $q_{\max} \approx 1 + \tilde{U}/\sqrt{2-\beta}$. We chose to take $\tilde{U} = 1$ as the dividing line between "large" and "small" $\tilde{U}$.

The following table presents some values of the extrema for various β and $\tilde{U}$:

β	$\tilde{U}$	$\tilde{R}_{\min}$	$\tilde{R}_{\max}$
-0.50	0.10	0.939647	1.06636
	1.0	0.582508	2.07998
	1.5	0.471179	3.19709
-0.25	0.10	0.936703	1.07035
	1.0	0.573797	2.25494
	1.5	0.463844	3.81605
0.00	0.10	0.933250	1.07510
	1.0	0.563777	2.51105
	1.5	0.455322	4.97694
0.25	0.10	0.929120	1.08091
	1.0	0.552085	2.92590
	1.5	0.445289	7.93950
0.50	0.10	0.924055	1.08823
	1.0	0.538191	3.72685
	1.5	0.433277	28.3503

Integrals of motion

A star orbiting in the disk has four constants of motion (for example, its initial position and velocity). *Integrals of motion* are functions of the phase-space coordinates $(\mathbf{x}, \mathbf{v})$ which are unchanged as the star moves in its orbit (Landau & Lifshitz 1960). In our equilibrium disk, there are two symmetries, and hence two integrals of motion: E stems from the time-independence, and L_z from the azimuthal symmetry. E and L_z are called *isolating integrals*, a term introduced by Lynden-Bell (1962): together they confine the orbit of a star to a two-dimensional manifold within the four-dimensional phase space. Throughout this thesis, we shall generally use the alternative isolating integrals U and R_{H}, which are functions of E and L_z.

Periods and frequencies

To begin with, we shall define a useful *auxiliary integral*

$$\mathcal{J}_n(\tilde{U}) = 2\int_{\tilde{R}_{\min}(\tilde{U})}^{\tilde{R}_{\max}(\tilde{U})} \frac{d\tilde{R}}{\tilde{R}^n\left(\tilde{U}^2 + 1 - \tilde{R}^{-2} + \frac{2}{\beta}\left(\tilde{R}^{-\beta} - 1\right)\right)^{\frac{1}{2}}}, \tag{2.27}$$

where $\tilde{R}_{\min}$ and $\tilde{R}_{\max}$ are the solutions of (2.26).

For the $\beta = 0$ case, this becomes (Z2.39)

$$\mathcal{J}_n(\tilde{U}) = 2\int_{\tilde{R}_{\min}(\tilde{U})}^{\tilde{R}_{\max}(\tilde{U})} \frac{d\tilde{R}}{\tilde{R}^n\left(\tilde{U}^2 + 1 - \tilde{R}^{-2} - 2\ln\tilde{R}\right)^{\frac{1}{2}}}. \tag{2.28}$$

To evaluate $\mathcal{J}_n$ numerically, we remove the singularities at either end of the integrand by transforming to a variable θ. We define $\tilde{R} = m + a\sin\theta$, where m is the midpoint of the radial motion, $m = \frac{1}{2}(\tilde{R}_{\min} + \tilde{R}_{\max})$ and a is its amplitude, $a = \frac{1}{2}(\tilde{R}_{\min} - \tilde{R}_{\max})$. The integration can then be carried out using the midpoint method (Press *et al.* 1989, ch. 4).

The following table gives values for some auxiliary integrals:

β	$\tilde{U}$	$\mathcal{J}_0$	$\mathcal{J}_1$	$\mathcal{J}_2$
-0.50	0.10	3.98427	3.97433	3.97235
	1.0	5.08256	4.00796	3.84033
	1.5	6.63669	4.01663	3.71130
-0.25	0.10	4.20340	4.19058	4.18709
	1.0	5.87455	4.35169	4.03084
	1.5	8.69860	4.51015	3.86754
0.00	0.10	4.46330	4.44659	4.44103
	1.0	7.08918	4.81884	4.26144
	1.5	12.9549	5.29286	4.05457
0.25	0.10	4.77838	4.75622	4.74773
	1.0	9.19215	5.50467	4.54829
	1.5	25.6789	6.81201	4.27913
0.50	0.10	5.17141	5.14127	5.12842
	1.0	13.7003	6.65529	4.91932
	1.5	155.287	12.1669	4.52611

We shall need $\mathcal{J}_0$ and $\mathcal{J}_2$ in order to find the frequencies of the orbit.

The radial frequency

The radial period T is the time taken for the star to travel between two successive pericentres: i.e. to move out and in again. Using the symmetry of the orbit, we have

$$T = 2\int_{R=R_{\mathrm{min}}}^{R=R_{\mathrm{max}}} dt = 2\int_{R_{\mathrm{min}}}^{R_{\mathrm{max}}} \frac{dR}{\dot{R}}. \tag{2.29}$$

In terms of dimensionless variables, this is

$$T = \frac{2R_{\mathrm{H}}}{v_\beta}\left(\frac{R_{\mathrm{H}}}{R_0}\right)^{\frac{\beta}{2}} \int_{\tilde{R}_{\mathrm{min}}}^{\tilde{R}_{\mathrm{max}}} \frac{d\tilde{R}}{\tilde{u}}. \tag{2.30}$$

Using the expression for the radial velocity given in (2.25), along with the definition (2.27), we find

$$T = \frac{R_{\mathrm{H}}}{v_\beta}\left(\frac{R_{\mathrm{H}}}{R_0}\right)^{\frac{\beta}{2}} \mathcal{J}_0. \tag{2.31}$$

The radial frequency κ is

$$\kappa = \frac{2\pi}{T} = \frac{v_\beta}{R_{\mathrm{H}}}\left(\frac{R_0}{R_{\mathrm{H}}}\right)^{\frac{\beta}{2}} \frac{2\pi}{\mathcal{J}_0}. \tag{2.32}$$

We shall find it useful to define the dimensionless radial frequency $\tilde{\kappa}$:

$$\tilde{\kappa} = \frac{R_{\mathrm{H}}}{v_\beta}\left(\frac{R_{\mathrm{H}}}{R_0}\right)^{\frac{\beta}{2}} \kappa = \frac{2\pi}{\mathcal{J}_0}. \tag{2.33}$$

The angular frequency

The angular period Θ is defined to be the angle through which a star moves during the time taken to complete one radial oscillation:

$$\Theta = \int_{t=0}^{t=T} d\theta = 2\int_{R_{\min}}^{R_{\max}} \frac{\dot{\theta}}{\dot{R}} dR = 2\int_{R_{\min}}^{R_{\max}} \frac{v dR}{uR}. \tag{2.34}$$

In terms of dimensionless variables, this is

$$\Theta = 2\int_{\tilde{R}_{\min}}^{\tilde{R}_{\max}} \frac{\tilde{v} d\tilde{R}}{\tilde{u}\tilde{R}} = \mathcal{J}_2, \tag{2.35}$$

where $\mathcal{J}_2$ is defined in (2.27).

The angular frequency Ω is defined to be the mean angular speed of the star:

$$\Omega = \frac{\Theta}{T} = \frac{v_\beta}{R_{\mathrm{H}}} \left(\frac{R_0}{R_{\mathrm{H}}}\right)^{\frac{\beta}{2}} \frac{\mathcal{J}_2}{\mathcal{J}_0}. \tag{2.36}$$

Once again, we shall henceforth use the dimensionless angular frequency $\tilde{\Omega}$, defined by

$$\tilde{\Omega} = \frac{R_{\mathrm{H}}}{v_\beta} \left(\frac{R_{\mathrm{H}}}{R_0}\right)^{\frac{\beta}{2}} \Omega = \frac{\mathcal{J}_2}{\mathcal{J}_0}. \tag{2.37}$$

The dependence of $\tilde{\kappa}$ and $\tilde{\Omega}$ on $\tilde{U}$ and β is shown in the following two graphs.

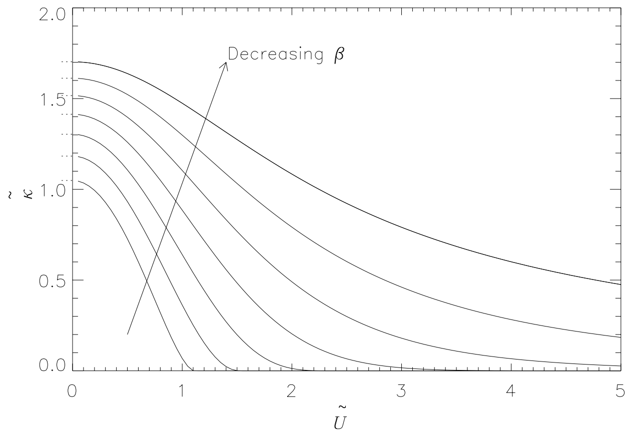


Figure 2.7: $\tilde{\kappa}$ plotted against $\tilde{U}$, for β = -0.9, -0.6, -0.3, 0, +0.3, +0.6, +0.9. The dotted lines on the vertical axis mark $\sqrt{2-\beta}$, the calculated value of $\tilde{\kappa}$ as $\tilde{U} \to 0$.

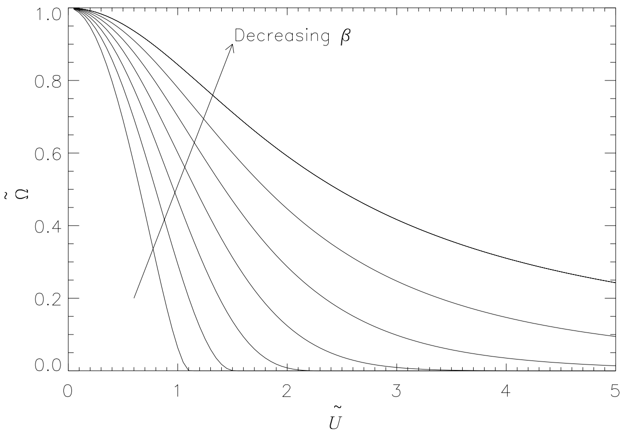


Figure 2.8: $\tilde{\Omega}$ plotted against $\tilde{U}$, for β = -0.9, -0.6, -0.3, 0, +0.3, +0.6, +0.9. Note that $\tilde{\Omega} \to 1$ as $\tilde{U} \to 0$

Some numerical values are:

β	$\tilde{U}$	$\tilde{\kappa}$	$\tilde{\Omega}$
-0.50	0.10	1.576996	0.997007
	1.00	1.236223	0.755589
	1.50	0.946734	0.559210
-0.25	0.10	1.494785	0.996120
	1.00	1.069561	0.686153
	1.50	0.722322	0.444616
0.00	0.10	1.407386	0.995006
	1.00	0.885537	0.600740
	1.50	0.484038	0.312416
0.25	0.10	1.314920	0.993586
	1.00	0.683538	0.494801
	1.50	0.244682	0.166640
0.50	0.10	1.214986	0.991687
	1.00	0.458618	0.359068
	1.50	0.040462	0.029147

In general the integral (2.27) has to be evaluated numerically for given β and U. However, we can calculate its behaviour in the limit $U \to 0$. The eccentric velocity U is, by definition, the maximum radial velocity. If this is zero, the star has no radial motion: it is on a circular orbit. Then the tangential velocity must be equal to the circular velocity: i.e. the star is at the home radius. So $\tilde{R}_{\text{min}} = \tilde{R}_{\text{max}} = 1$. As U increases from zero, $\tilde{R}_{\text{min}}$ will drop below 1 and $\tilde{R}_{\text{max}}$ will rise above 1. We therefore write $\tilde{R} = 1 + \epsilon$, and substitute this into (2.26) to obtain

$$\tilde{U}^2 + (\beta - 2)\,\epsilon^2 = 0. \tag{2.38}$$

and thus

$$\epsilon = \pm\Delta, \qquad \Delta = \frac{\tilde{U}}{\sqrt{(2-\beta)}}. \tag{2.39}$$

The two values of ϵ correspond to $\tilde{R}_{\text{min}}$ and $\tilde{R}_{\text{max}}$. Then (retaining terms to second order in ϵ) the general integral (2.27) becomes

$$\lim_{\tilde{U}\to 0} \mathcal{J}_n(\tilde{U}) = 2\int_{-\Delta}^{+\Delta} \frac{d\epsilon}{\left(\tilde{U}^2 - (2-\beta)\,\epsilon^2\right)^{\frac{1}{2}}} = \frac{2\pi}{\sqrt{2-\beta}}. \tag{2.40}$$

Therefore the dimensionless epicyclic and circular frequencies are

$$\tilde{\kappa}_0 \equiv \lim_{\tilde{U}\to 0} \tilde{\kappa} = \sqrt{2-\beta}, \qquad \tilde{\Omega}_0 \equiv \lim_{\tilde{U}\to 0} \tilde{\Omega} = 1. \tag{2.41}$$

The dimensional frequencies are

$$\kappa_0 \equiv \lim_{\tilde{U}\to 0} \kappa = \sqrt{2-\beta} \times \frac{v_\beta}{R}\left(\frac{R_0}{R}\right)^{\beta/2}, \qquad \Omega_0 \equiv \lim_{\tilde{U}\to 0} \Omega = \frac{v_\beta}{R}\left(\frac{R_0}{R}\right)^{\beta/2}. \tag{2.42}$$

As expected, then, the angular frequency is that of a star in a circular orbit at R_{H}: $\Omega_0 = v_{\text{circ}}/R_{\text{H}}$. At low but non-zero $\tilde{U}$, there is superimposed on this circular motion a small radial oscillation, which can be viewed as due to the epicyclic motion of the stars. Note that the epicyclic frequency can also be obtained via the formula (Binney & Tremaine 1987)

$$\kappa_0^2 = \left(R \frac{d\Omega_0^2}{dR} + 4\Omega_0^2 \right).$$

It should be emphasised that throughout this thesis we shall use the general frequencies κ and Ω; we do not rely on the epicyclic approximations κ_0 and Ω_0.

2.3 The structure of phase space

We have chosen a surface density for our disk; we have calculated the gravitational potential due to this density, and we have investigated some properties of the orbits of stars in this disk. We now need to consider how the stars make up such a disk. We must position the stars so as to build up the correct density distribution. Further, the stars are not stationary, but are orbiting in the potential of the disk. Some of the orbits are circular, some may be highly elongated; some are closed, some open rosettes; but they must all combine to give an unchanging density distribution with exactly the radial dependence required. On the face of it, this seems impossibly complicated to arrange, and yet it can be done!

The mathematical object that positions the stars correctly is the *distribution function* $f(R, \theta, u, v, t)$. This is normalised so that the integral of the distribution function over all velocities gives the surface density at that point:

$$\Sigma(R, \theta, t) = \iint_{\text{all velocities}} f(R, \theta, u, v, t)\, du\, dv. \tag{2.43}$$

The distribution function thus has units of surface density over velocity-squared. f describes how the stars are located in phase space, which for a disk is four-dimensional. f can be regarded as a *phase space density*, or the density of a fluid in phase space. If the distribution of the stars changes with time, the fluid will move to reflect this. Stars cannot jump through phase space, nor can they change their velocity discontinuously. In real galaxies, the relaxation time – the time taken for a star's velocity to be significantly changed by collisions with other stars – is much greater than the age of the galaxy (Jeans 1919; Binney & Tremaine 1987, ch. 4). This means that collisions can be neglected. The flow is then incompressible. Under these conditions, f satisfies a continuity equation in phase space

$$\frac{\partial f}{\partial t} + \frac{\partial}{\partial v_i}(f \dot{v}_i) + \frac{\partial}{\partial x_i}(f \dot{x}_i) = 0. \tag{2.44}$$

where $\mathbf{v}$ is the star's velocity and $\mathbf{x}$ its position, and the Einstein summation convention applies. From this (Binney & Tremaine 1987, ch. 4), we obtain the *collisionless Boltzmann equation*:

$$\frac{df}{dt} \equiv \frac{\partial f}{\partial t} + v_i \frac{\partial f}{\partial x_i} + \frac{\partial \psi}{\partial x_i} \frac{\partial f}{\partial v_i} = 0. \tag{2.45}$$

df/dt is the *convective* or *Lagrangian* derivative, which describes how the local phase-space density changes for an observer travelling with a star as it moves through phase-space. When df/dt vanishes, the observer notices no change in the density: i.e. the fluid is incompressible.

The collisionless Boltzmann equation leads to an important constraint on the possible form of the equilibrium distribution function f, first derived by Poincaré (1911), and popularised by Jeans (1915). In general, f is a function of R, θ, u, v and t. But the equilibrium distribution function does not depend explicitly on t. Thus f is a function of the four phase-space coordinates alone. Also, the collisionless Boltzmann equation tells us that f does not change over the orbit of a star: $df/dt = 0$. These two conditions together mean that f is an integral of motion. Conversely, any function of integrals of motion will yield a time-independent distribution function that satisfies the collisionless Boltzmann equation

$$\frac{d}{dt} f\left(I_1(\mathbf{x},\mathbf{v})\ldots I_n(\mathbf{x},\mathbf{v})\right) = \frac{\partial f}{\partial I_n}\frac{dI_n}{dt} = 0. \tag{2.46}$$

This result is known as the Jeans theorem. Lynden-Bell (1962) pointed out that only isolating integrals should be used in the Jeans theorem. In this thesis, we build our equilibrium distribution function out of the energy E and angular momentum L_z. A simple set of distribution functions has already been provided by Evans (1994). They form the basis for the equilibrium models studied in this thesis. But they are assuredly not the only possible choices. An example of an alternative set of distribution functions is provided in Appendix B. These are perhaps curiosities, as the disks are built from orbits of one shape only.

Evans (1994) followed the method of Fricke (1951) and Miyamoto (1971; 1974) to find distribution functions for power-law disks. The surface density is expanded in powers of ψ and R, $\Sigma_{\rm eq} = R^k\psi^j$, and distribution functions of the form $f = CL_z^l E^e$ are sought. The surface density must be the integral of the distribution function over velocity space: $\Sigma_{\rm eq}(R,\theta) = \iint f(E,L_z)\,du\,dv$. Substituting in the assumed distribution function $CL_z^l E^e$, the integration is carried out to give $\Sigma_{\rm eq}$ as a product of powers of R and ψ. Thus the powers k and j are related to e and l.

The distribution function obtained in this way is *self-consistent*. When integrated over velocity space, it yields the surface density which sets up the potential contained in the distribution function through $E = \frac{1}{2}\left(u^2+v^2\right) - \psi(R)$. This is illustrated by the closed loop in the following diagram:

$$\begin{tikzcd}
 & \begin{array}{c}\nabla^2\psi = -4\pi G\Sigma\delta(z)\\ \psi(R,\theta)\end{array} \arrow[dr] & \\
\Sigma_{\rm eq}(R,\theta) = \iint f_{\rm s}du\,dv \arrow[ur] & & E = T(u,v) - \psi(R,\theta) \arrow[dl] \\
 & f_{\rm s}(E,L) \arrow[ul] &
\end{tikzcd}$$

Figure 2.9: Relation between surface density and potential for a self-consistent distribution function.

Later we introduce distribution functions that are not self-consistent. This self-consistent distribution function is therefore denoted by f_{s}.

For the positive β models, Evans (1994) obtained

$$\Sigma_{\mathrm{eq}} = R^{2k}\psi^{l}, \qquad f_{\mathrm{s}}(E, L_z) = L_z^{2k}(-E)^{l-k-1}\frac{\Gamma\left[l+1\right]}{2^{k+1}\sqrt{\pi}\Gamma\left[k+\frac{1}{2}\right]\Gamma\left[l-k\right]}, \tag{2.47}$$

where $l > k$ and $k > -\frac{1}{2}$. For negative β he obtained:

$$\Sigma_{\mathrm{eq}} = \frac{R^{2k}}{(-\psi)^{l}}, \qquad f_{\mathrm{s}}(E, L_z) = \frac{L_z^{2k}}{E^{l+k+1}}\frac{\Gamma\left[l+k+1\right]}{2^{k+1}\sqrt{\pi}\Gamma\left[k+\frac{1}{2}\right]\Gamma\left[l\right]}, \tag{2.48}$$

where $l > 0$ and $k > -\frac{1}{2}$.

To relate these results to our model, we need to use the equations for the surface density (2.1) and potential (2.3) of our general power-law disk to carry out a *surface density partition* (Evans & Collett 1993). We write

$$\Sigma_{\mathrm{eq}}(R) = \Sigma_0\left(\frac{R}{R_0}\right)^{\gamma}\left(\frac{\psi}{\psi_0}\right)^{1+\frac{1}{\beta}+\frac{\gamma}{\beta}}, \tag{2.49}$$

where $\psi_0 = v_\beta^2/\beta$ is the potential at the reference radius, and γ is arbitrary.

Falling rotation curve: positive β

For the positive β models, we find

$$f_{\mathrm{s}}(E, L_z) = \frac{C_{\beta\gamma}\Sigma_0\beta^{1+\frac{1}{\beta}+\frac{\gamma}{\beta}}}{2^{\gamma/2}\sqrt{\pi}R_0^{\gamma}v_\beta^{2\left(1+\frac{1}{\beta}+\frac{\gamma}{\beta}\right)}}L_z^{\gamma}(-E)^{1/\beta+\gamma/\beta-\gamma/2}, \tag{2.50}$$

provided $\gamma > -1$. The normalisation constant $C_{\beta\gamma}$ is

$$C_{\beta\gamma} = \frac{\Gamma\left[2+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]}{\Gamma\left[\frac{1}{2}(\gamma+1)\right]\Gamma\left[1+\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}\right]}. \tag{2.51}$$

Rising rotation curve: negative β

For the negative β models, the analogous result is:

$$f_{\mathrm{s}}(E, L_z) = \frac{C_{\beta\gamma}\Sigma_0 v_\beta^{2\left(-1-\frac{1}{\beta}-\frac{\gamma}{\beta}\right)}}{2^{\gamma/2}\sqrt{\pi}R_0^{\gamma}\left(-\beta\right)^{-1-\frac{1}{\beta}-\frac{\gamma}{\beta}}}\frac{L_z^{\gamma}}{E^{\frac{\gamma}{2}-\frac{1}{\beta}-\frac{\gamma}{\beta}}}, \tag{2.52}$$

provided $\gamma > -\beta - 1$. The normalisation constant $C_{\beta\gamma}$ is now

$$C_{\beta\gamma} = \frac{\Gamma\left[\frac{\gamma}{2}-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}{\Gamma\left[\frac{1}{2}(\gamma+1)\right]\Gamma\left[-1-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}. \tag{2.53}$$

(Recall that for bound orbits the energy must have the opposite sign to that of β, so we shall never encounter problems with taking the root of a negative number.)

In summary, then:

$$0<|\,\beta\,|<1: \qquad f_{\mathrm{s}}(E,L_z)=\tilde{C}L_z^{\gamma}|E|^{1/\beta+\gamma/\beta-\gamma/2}, \tag{2.54}$$

$$\text{where} \qquad \tilde{C}=\frac{C_{\beta\gamma}\Sigma_0|\beta|^{1+\frac{1}{\beta}+\frac{\gamma}{\beta}}}{2^{\gamma/2}\sqrt{\pi}R_0^{\gamma}v_{\beta}^{2\left(1+\frac{1}{\beta}+\frac{\gamma}{\beta}\right)}}; \tag{2.55}$$

$$\beta>0: \qquad C_{\beta\gamma}=\frac{\Gamma\left[2+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]}{\Gamma\left[\frac{1}{2}(\gamma+1)\right]\Gamma\left[1+\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}\right]}, \tag{2.56}$$

$$\text{and} \qquad \gamma>-1; \tag{2.57}$$

$$\beta<0: \qquad C_{\beta\gamma}=\frac{\Gamma\left[\frac{\gamma}{2}-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}{\Gamma\left[\frac{1}{2}(\gamma+1)\right]\Gamma\left[-1-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}, \tag{2.58}$$

$$\text{and} \qquad \gamma>-\beta-1. \tag{2.59}$$

Note that these formulae differ by a factor of 2 from those given by Evans (1994). This is because Evans' results are for a bi-directional disk, where the stars rotate in both senses. In this thesis, we shall consider only uni-directional disks.

Flat rotation curve: zero β

Similar results can be obtained for the Toomre-Zang disk (Toomre 1972; Zang 1976; Toomre 1977; Binney & Tremaine 1987, ch. 4)

$$f_{\mathrm{s}}(E,L_z)=\tilde{C}L_z^{\gamma}\exp\left(-(\gamma+1)\frac{E}{v_0^2}\right), \tag{2.60}$$

where

$$\tilde{C}=\frac{C_{0\gamma}\Sigma_0}{2^{\gamma/2}\sqrt{\pi}R_0^{\gamma}v_0^{\gamma+2}}, \qquad\qquad C_{0\gamma}=\frac{(\gamma+1)^{1+\gamma/2}}{\Gamma\left[\frac{1}{2}(\gamma+1)\right]}. \tag{2.61}$$

Just as the potential for the Toomre-Zang disk involves a logarithm instead of powers of R, so the distribution function involves an exponential instead of powers of E.

The mean streaming velocity

The mean streaming velocity is defined as

$$\langle v\rangle=\frac{\int vf\,du\,dv}{\int f\,du\,dv}. \tag{2.62}$$

For $\beta > 0$, we work in terms of the binding energy $\mathcal{E} = -E > 0$ and the potential $\psi(R) > 0$. The minimum value of $\mathcal{E}$ is zero; the maximum possible is $\mathcal{E} = -\psi(R) = v_\beta^2 R_0^\beta / \beta R^\beta$. Then the streaming velocity is

$$\langle v \rangle = \frac{1}{R} \int_{\mathcal{E}=0}^{\psi(R)} d\mathcal{E}\, \mathcal{E}^{1/\beta + \gamma/\beta - \gamma/2} \int_{L=0}^{R\sqrt{2(\psi - \mathcal{E})}} \frac{L_z^{\gamma+1} dL}{\sqrt{2R^2(\psi - \mathcal{E}) - L^2}}. \tag{2.63}$$

For $\beta < 0$, we work in terms of the energy $E > 0$ and $\phi(R) = -\psi(R) > 0$. The minimum possible value of E is $E = \phi$; the maximum is $E = \infty$. The streaming velocity is given by

$$\langle v \rangle = \frac{1}{R} \int_{E=\phi(R)}^{\infty} dE\, E^{1/\beta + \gamma/\beta - \gamma/2} \int_{L=0}^{R\sqrt{2(E - \phi)}} \frac{L_z^{\gamma+1} dL}{\sqrt{2R^2(E - \phi) - L^2}}. \tag{2.64}$$

These integrations can be carried out analytically to yield

$$\langle v \rangle = v_\beta \sqrt{\frac{2}{\beta}} \frac{\Gamma\left[1 + \frac{\gamma}{2}\right]}{\Gamma\left[\frac{1}{2} + \frac{\gamma}{2}\right]} \frac{\Gamma\left[2 + \frac{1}{\beta} + \frac{\gamma}{\beta}\right]}{\Gamma\left[\frac{5}{2} + \frac{1}{\beta} + \frac{\gamma}{\beta}\right]} \left(\frac{R_0}{R}\right)^{\beta/2}, \qquad \beta > 0, \tag{2.65}$$

$$\langle v \rangle = v_\beta \sqrt{\frac{2}{-\beta}} \frac{\Gamma\left[1 + \frac{\gamma}{2}\right] \Gamma\left[-\frac{3}{2} - \frac{1}{\beta} - \frac{\gamma}{\beta}\right]}{\Gamma\left[\frac{1}{2} + \frac{\gamma}{2}\right] \Gamma\left[-1 - \frac{1}{\beta} - \frac{\gamma}{\beta}\right]} \left(\frac{R}{R_0}\right)^{-\beta/2}, \qquad \beta < 0, \tag{2.66}$$

$$\langle v \rangle = v_0 \sqrt{\frac{2}{1+\gamma}} \frac{\Gamma\left[1 + \frac{\gamma}{2}\right]}{\Gamma\left[\frac{1}{2} + \frac{\gamma}{2}\right]}, \qquad \beta = 0. \tag{2.67}$$

The radial velocity dispersion

The radial velocity dispersion is the root-mean-squared radial velocity, defined by

$$\sigma_u^2 = \langle u^2 \rangle = \frac{\int u^2 f(R, u, v)\, du\, dv}{\int f(R, u, v)\, du\, dv}. \tag{2.68}$$

Evans (1994) finds that, for the self-consistent distribution function (2.54), this is

$$\sigma_u^2 = \frac{v_\beta^2}{1 + \gamma + 2\beta} \left(\frac{R_0}{R}\right)^\beta. \tag{2.69}$$

Conveniently, this result holds for both positive and negative β, even though different distribution functions must be used in each case! We define the dimensionless radial velocity dispersion

$$\tilde{\sigma}_u^2 = \left(\frac{R}{R_0}\right)^\beta \frac{\sigma_u^2}{v_\beta^2} = \frac{1}{1 + \gamma + 2\beta}. \tag{2.70}$$

If all the stars are on circular orbits, we see from (2.68) that the radial velocity dispersion is identically zero. Such a disk is traditionally said to be "cold", by analogy with the motion of molecules in gases. As the eccentricity of the orbits increases, the stars acquire random motion, and thus the "temperature" of the disk increases. High values of the anisotropy parameter γ correspond to low velocity dispersions, i.e. cold disks. We can see why this is so from the distribution function: $f_{\rm s} \propto L_z^\gamma$. When γ is large, more of the stars have high angular momentum, i.e. approach circular orbits, so the disk is cool. The isotropic model is given by $\gamma = 0$. In this thesis, a disk is described as "hot" or "cold" according to whether its velocity dispersion σ_u is large or small. Similarly, a disk's "temperature" refers to its value of σ_u (or equivalently γ).

The tangential velocity dispersion

We define the tangential velocity dispersion as

$$\sigma_v^2 = \frac{\int (v - \langle v \rangle)^2 f(R,u,v)\, du\, dv}{\int f(R,u,v)\, du\, dv} = \langle v^2 \rangle - \langle v \rangle^2. \tag{2.71}$$

Evans (1994) finds that both positive and negative β give the same answer for $\langle v^2 \rangle$:

$$\langle v^2 \rangle = \frac{v_\beta^2 (1+\gamma)}{1+\gamma+2\beta} \left(\frac{R_0}{R}\right)^\beta = (1+\gamma)\sigma_u^2. \tag{2.72}$$

As $\gamma \to \infty$, $\sigma_u^2 \to 0$ and $\sqrt{\langle v^2 \rangle} \to v_{\rm circ}$, so the disk is rotationally supported. The power-law disks have the property that, at any spot, the ratio of radial velocity dispersion to mean squared tangential velocity is constant. Such models were first considered by Hénon (1973).

The kinetic and potential energies

We can also work out the total energy stored within a radius R in different forms. We define

$$K = \frac{1}{2}\int_0^R \Sigma(R')(\langle u^2 \rangle + \langle v^2 \rangle)\, dA, \qquad W = \int_0^R \Sigma(R') R' \frac{\partial \psi}{\partial R'}\, dA, \tag{2.73}$$

$$T = \frac{1}{2}\int_0^R \Sigma(R') \langle v \rangle^2\, dA, \qquad \Pi = \int_0^R \Sigma(R')(\sigma_u^2 + \sigma_v^2)\, dA, \tag{2.74}$$

where dA is an element of area. K is the total kinetic energy, and W is the total gravitational potential energy. K can be divided into contributions from ordered (rotational) and random motion, T and Π respectively: $K = T + \frac{1}{2}\Pi$ (Binney & Tremaine 1987, p. 212). These integrals can be evaluated to yield

$$K(R) = \pi \Sigma_0 v_\beta^2 R_0^2 \frac{2+\gamma}{(1+\gamma+2\beta)(1-2\beta)} \left(\frac{R}{R_0}\right)^{1-2\beta}, \tag{2.75}$$

and

$$W(R) = -\frac{2\pi \Sigma_0 v_\beta^2 R_0^2}{1-2\beta} \left(\frac{R}{R_0}\right)^{1-2\beta}. \tag{2.76}$$

For the power-law disks, the virial theorem takes the form

$$2K(R) + \frac{2+\gamma}{1+\gamma+2\beta} W(R) = 0. \tag{2.77}$$

Note that the disk does not, in general, satisfy the standard virial theorem $2K + W = 0$. This is because it is not possible to "enclose" the system by drawing a sufficiently large container around it. No matter how large the diameter of the container, if the disk is warm, some stars will always cross its surface. When the disk is perfectly cold, the stars have no radial motion and thus do not cross the surface. In this case, as we see from eq. (2.77), the standard virial theorem holds.

The rotational kinetic energy is

$$T(R) = \frac{2\pi\Sigma_0 v_\beta^2 R_0^2}{\beta(1-2\beta)} \frac{\Gamma^2\left[1+\frac{\gamma}{2}\right]\Gamma^2\left[2+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]}{\Gamma^2\left[\frac{1}{2}+\frac{\gamma}{2}\right]\Gamma^2\left[\frac{5}{2}+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]} \left(\frac{R}{R_0}\right)^{1-2\beta}, \qquad \beta > 0, \tag{2.78}$$

$$T(R) = -\frac{2\pi\Sigma_0 v_\beta^2 R_0^2}{\beta(1-2\beta)} \frac{\Gamma^2\left[1+\frac{\gamma}{2}\right]\Gamma^2\left[-\frac{3}{2}-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}{\Gamma^2\left[\frac{1}{2}+\frac{\gamma}{2}\right]\Gamma^2\left[-1-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]} \left(\frac{R}{R_0}\right)^{1-2\beta}, \qquad \beta < 0, \tag{2.79}$$

$$T(R) = \frac{2\pi\Sigma_0 v_0^2 R_0^2}{1+\gamma} \frac{\Gamma^2\left[1+\frac{\gamma}{2}\right]}{\Gamma^2\left[\frac{1}{2}+\frac{\gamma}{2}\right]} \left(\frac{R}{R_0}\right), \qquad \beta = 0. \tag{2.80}$$

2.4 The cut-out distribution functions

So far we have considered the self-consistent case, where the surface density generates a potential in which stars orbit with a distribution $f_{\rm s}$ so as to recover the original surface density. It is also convenient to examine disks where parts of the central density are carved out. This is very much in the spirit of Zang's (1976) pioneering investigations. The cut-out mass is still present, in the sense that it contributes to the potential experienced by the remaining stars, but it is not now free to participate in the perturbation. The disk is thus divided into "active" and "inactive" components. The following diagram illustrates this state of affairs.

$$\begin{tikzcd}
\Sigma_{\rm inactive}(R,\theta) \arrow[d] & \begin{array}{c}\nabla^2\psi = -4\pi G\Sigma\delta(z)\\ \psi(R,\theta)\end{array} \arrow[dr] & \\
\Sigma_{\rm eq}(R,\theta) \arrow[ur] & & E = T(u,v) - \psi(R,\theta) \arrow[ddl] \\
\Sigma_{\rm active}(R,\theta) = \iint f_{\rm cutout} du\, dv \arrow[u] & & \\
& f_{\rm cutout}(E,L) \arrow[ul] &
\end{tikzcd}$$

Figure 2.10: Relation between surface density and potential for a cut-out distribution function.

Although motivated partly by mathematical convenience, this is also a physically reasonable step to take.

Stars in galactic disks are subject not merely to their own potential, but also to forces from the halo and bulge. A self-consistent distribution function, such as f_{s}, is realistic only when the disk's self-gravity overwhelms the gravitational potential of the other components. The immobile central mass can be interpreted physically as the hot bulge at the centre of disk galaxies, which contributes (along with the halo) to the potential in the disk. Another possibility, suggested to us by Scott Tremaine (1997, private communication), is to interpret the rigid density as being caused by stars on highly elongated radial orbits. They pass through the centre of the disk, but they spend most of their time sufficiently far away from the disk that they do not respond to the changing potential.

There is another advantage in thus modifying the mass distribution. We wish to compare our results with those from N-body studies such as Earn (1993). Numerical simulations cannot cope with infinite forces, so in these studies the singularity at the origin is removed. Also, they cannot track the behaviour of the disk to infinite distances, so the disk must be truncated at some finite radius. For comparison with these studies, we also introduce an outer cut-out.

We multiply our self-consistent distribution function f_{s} (2.54) by a *cut-out function* $H(L_z)$:

$$\boxed{f_{\mathrm{cutout}}(E, L_z) = H(L_z) f_{\mathrm{s}}(E, L_z). \qquad (2.81)}$$

We shall carry out our analysis for three $H(L_z)$:

the *self-consistent* (scale-free) disk,

$$\boxed{H(L_z) = 1, \qquad (2.82)}$$

the *inner cut-out* disk,

$$\boxed{H(L_z) = \frac{L_z^{N_\beta}}{L_z^{N_\beta} + (v_\beta R_0)^{N_\beta}}, \qquad (2.83)}$$

and the *doubly cut-out* disk

$$\boxed{H(L_z) = \frac{L_z^{N_\beta} L_{\mathrm{c}}^{M_\beta}}{\left[L_z^{N_\beta} + (v_\beta R_0)^{N_\beta}\right]\left[L_z^{M_\beta} + L_{\mathrm{c}}^{M_\beta}\right]}. \qquad (2.84)}$$

In these formulae,

$$N_\beta = \frac{2+\beta}{2-\beta} N, \qquad M_\beta = \frac{2+\beta}{2-\beta} M, \qquad (2.85)$$

where N and M are the inner and outer cut-out indices, respectively; they must be positive integers.

We frequently use an equivalent form of the cut-out function expressed in dimensionless variables. This is defined by $\tilde{H}(\tilde{L}_z) \equiv H(L)$. The choice of the inner cut-out reduces to Zang's (1976, (Z2.57)) when $\beta \to 0$. This generalisation seems to come out of thin air, but we shall see in Appendix C that it provides a considerable simplification in our calculation of a contour integral. Earn (1993) uses a doubly cut-out function of the same form as (2.84) to carry out his numerical simulations (although, not having to perform any contour integrals, Earn is free to choose non-integral N and M). The inner cut-out function means that matter in the central parts of the disk is prevented from participating in the disturbance. This frozen matter could be regarded as representing the bulge of a physical galaxy. The outer cut-out function reduces the active surface density to zero at large radii. The potential here could be regarded as due to an extended dark halo.

The inner cut-out

The form of the inner cut-out function $\tilde{H}(\tilde{L}_z)$ is shown in fig. 2.11.

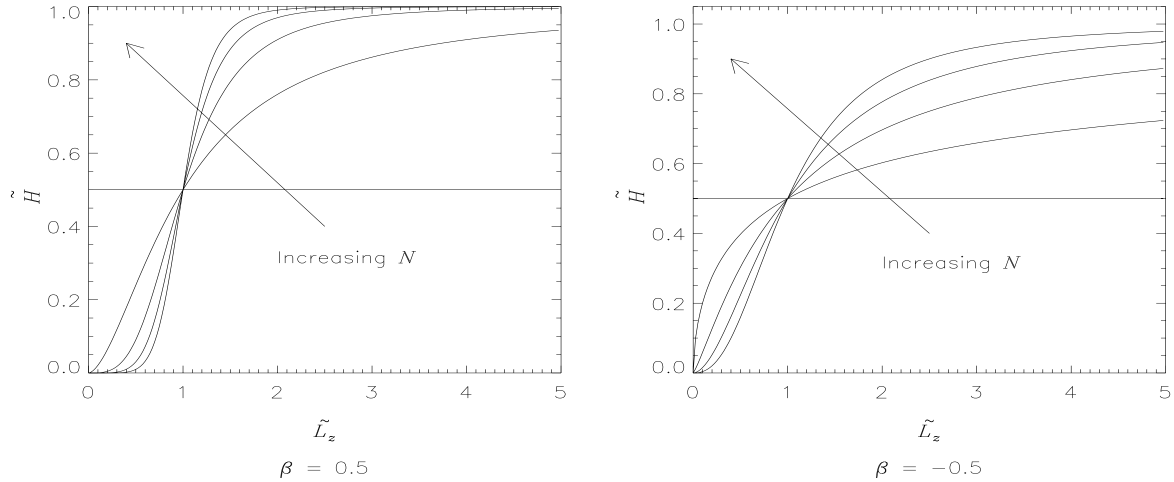


Figure 2.11: The inner cut-out function $\tilde{H}(\tilde{L}_z)$ for $\beta = +0.5$ (left) and $\beta = -0.5$ (right). Five values of N are shown, from $N = 0$ to $N = 4$. The horizontal line is the $N = 0$ curve; the arrows show the direction of increasing N.

For $N \neq 0$, $\tilde{H}(\tilde{L}_z)$ rises from zero at small $\tilde{L}_z$ ($\lesssim 1$) to unity at large $\tilde{L}_z$ ($\gtrsim 1$). We see that N controls the steepness of the rise. It is steeper for large N.

What effect does this modification of the distribution function have on the surface density? The cut-out functions depend on the dimensionless angular momentum $\tilde{L}_z$. By removing orbits with low $\tilde{L}_z$, we are discarding some of the stars with a small home radius, $\tilde{R}_{\mathrm{H}} = \tilde{L}_z^{2/(2-\beta)}$, and thus removing matter from the centre of the disk. In a cold disk, all the stars are on circular orbits, for which the angular momentum is $\tilde{L}_z = (R/R_0)^{1-\beta/2}$. The cut-out function thus depends on R only. The active surface density of a cold disk is then (c.f. Z2.66)

$$\Sigma_{\text{active}} = \Sigma_0 \left(\frac{R_0}{R}\right)^{1+\beta} \frac{R^{\frac{2+\beta}{2}N}}{R^{\frac{2+\beta}{2}N} + R_0^{\frac{2+\beta}{2}N}}. \qquad (2.86)$$

For a cold disk, the active density tends to zero at the origin whenever $(2+\beta)(N-1) > \beta$. This is true *except* when $N = 1$ and β is positive; then the active density tends to infinity at the origin. This is demonstrated in the above plot. When $\beta = 0$, the surface density at the origin is zero for $N \geq 2$, and unity for $N = 1$.

The active surface density of a hot disk must be calculated by numerical integration. Fig. 2.12 compares the active surface density for a hot disk (right-hand plot) with that of a cold disk (left-hand plot). Heating the disk makes little difference to the active density. For instance, for a cold disk, eq. (2.86) shows that the density falls to half the equilibrium value at $R = R_0$; fig. 2.12 shows that this holds true, to good accuracy, for a hotter disk.

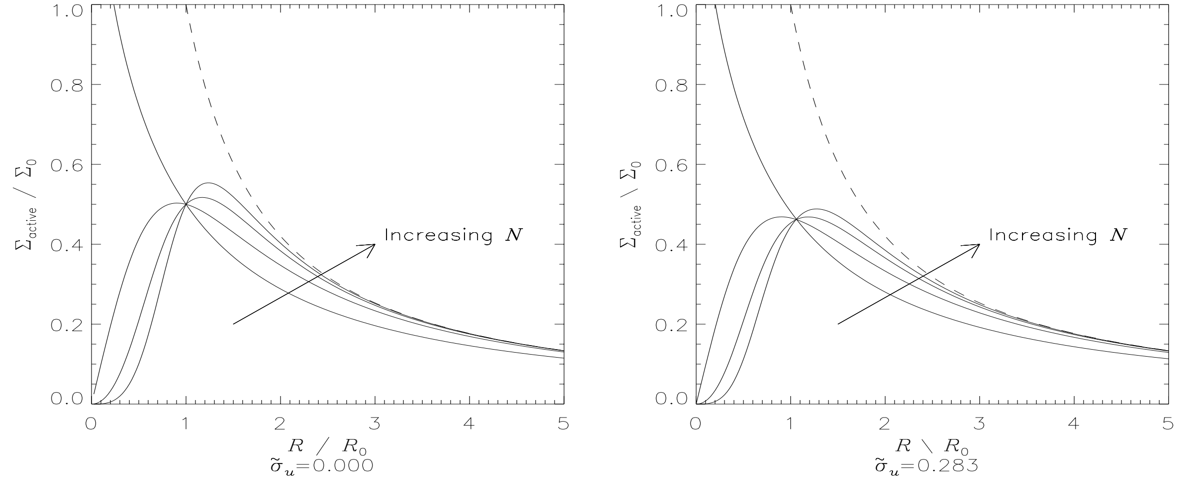


Figure 2.12: The active surface density of a $\beta = 0.25$ disk with an inner cut-out function. The left-hand plot is for a cold disk, where the density is given by eq. (2.86). The right-hand plot is for $\tilde{\sigma}_u = 0.283$, which is the temperature at which the self-consistent disk is locally stable to axisymmetric disturbances. In each case, the solid lines are the density for $N = 1$, 2, 3, 4; the arrows indicate the direction of increasing N. The dashed line is the equilibrium surface density.

In fig. 2.13, the active surface density of a hot disk is considered as a fraction of the equilibrium surface density at each radius. The form of these curves is close to those of $\tilde{H}(\tilde{L}_z)$ shown in fig. 2.11. Thus the value of $\tilde{H}(\tilde{L}_z)$ is a good approximation to the proportion of density which is active at $R = R_0\tilde{L}_z$. At $R = R_0$, approximately half the density is active; at higher radii, less matter is removed and the active density approaches that of the equilibrium disk.

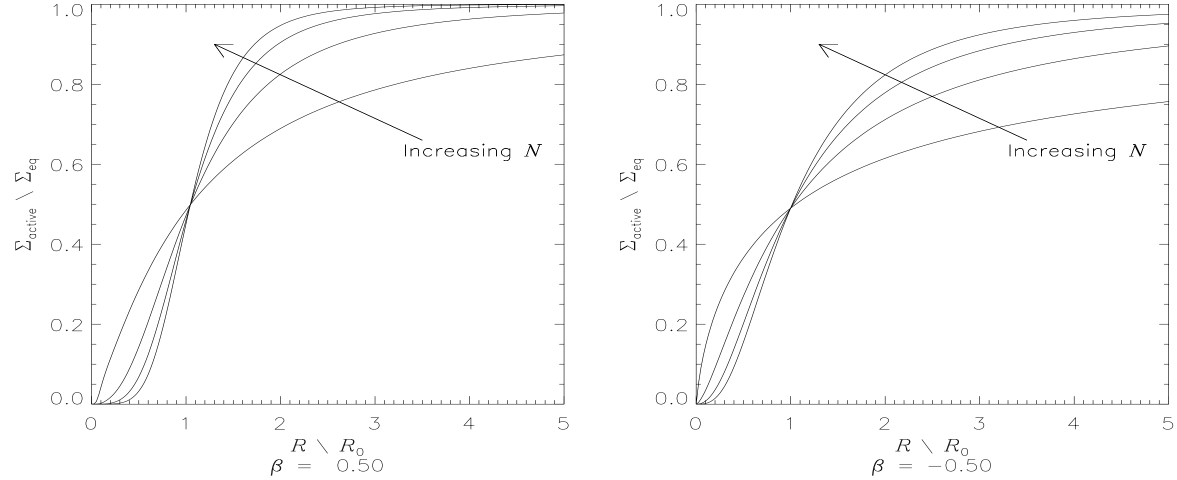


Figure 2.13: The active surface density of a hot disk with an inner cut-out function, shown as a fraction of the equilibrium density. The solid lines are the density for $N = 1$, 2, 3, 4; the arrows indicate the direction of increasing N. For $\beta = +0.5$, $\tilde{\sigma}_u = 0.199$; for $\beta = -0.5$, $\tilde{\sigma}_u = 0.740$. These are the temperatures at which the self-consistent disk is locally stable to axisymmetric disturbances.

The outer cut-out

In the doubly cut-out disk, the self-consistent distribution function is multiplied not only by the inner cut-out function (2.83) but also by the outer cut-out function $\left[\left(\tilde{L}_z/\tilde{L}_\mathrm{c}\right)^{M_\beta}+1\right]^{-1}$.

Fig. 2.14 shows the dependence of the doubly cut-out function $\tilde{H}(\tilde{L}_z)$ on $\tilde{L}_z$ for different M and $\tilde{L}_\mathrm{c}$.

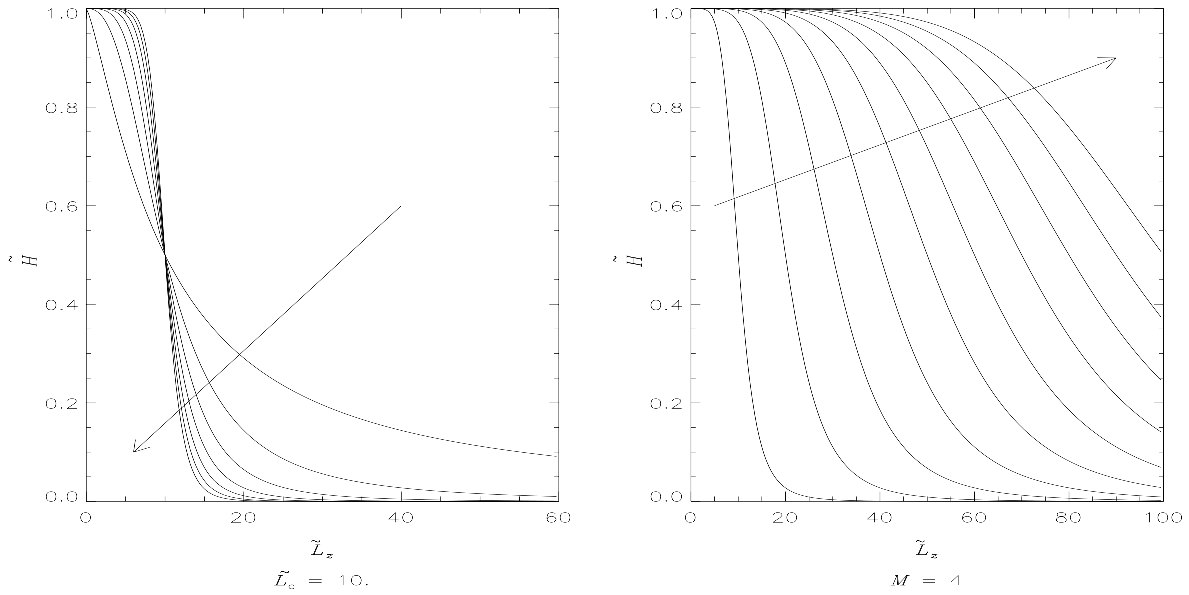


Figure 2.14: The outer cut-out function $\tilde{H}(\tilde{L}_z)$ plotted against $\tilde{L}_z$,for $\beta = +0.25$. The left-hand plot shows $\tilde{H}(\tilde{L}_z)$ for $\tilde{L}_\mathrm{c} = 10$ and eight values of M, from $M = 0$ to $M = 7$. The horizontal line is the $M = 0$ curve; the arrow shows the direction of increasing M. The right-hand plot shows $\tilde{H}(\tilde{L}_z)$ for $M = 4$. Ten values of $\tilde{L}_\mathrm{c}$ are shown: $\tilde{L}_\mathrm{c} = 10, 20, \ldots 100$. The arrow shows the direction of increasing $\tilde{L}_\mathrm{c}$.

Again, for a cold disk, we can write the outer cut-out function as a function of radius only, so that the active surface density is simply the equilibrium density multiplied by the cut-out function. The active surface density for a doubly cut-out cold disk is

$$\Sigma_\mathrm{active} = \Sigma_0 \left(\frac{R_0}{R}\right)^{1+\beta} \frac{R^{\frac{2+\beta}{2}N}}{\left[R^{\frac{2+\beta}{2}N} + R_0^{\frac{2+\beta}{2}N}\right]} \frac{(R_0\tilde{R}_\mathrm{c})^{\frac{2+\beta}{2}M}}{\left[R^{\frac{2+\beta}{2}M} + (R_0\tilde{R}_\mathrm{c})^{\frac{2+\beta}{2}M}\right]}, \tag{2.87}$$

where the *truncation radius* $\tilde{R}_\mathrm{c}$ is given by $\tilde{R}_\mathrm{c} = \tilde{L}_\mathrm{c}^{2/(2-\beta)}$.

Fig. 2.15 compares the surface density in cold and hot disks. The dotted lines mark the position of $\tilde{R}_\mathrm{c}$. At $\tilde{R}_\mathrm{c}$, roughly half the density is active.

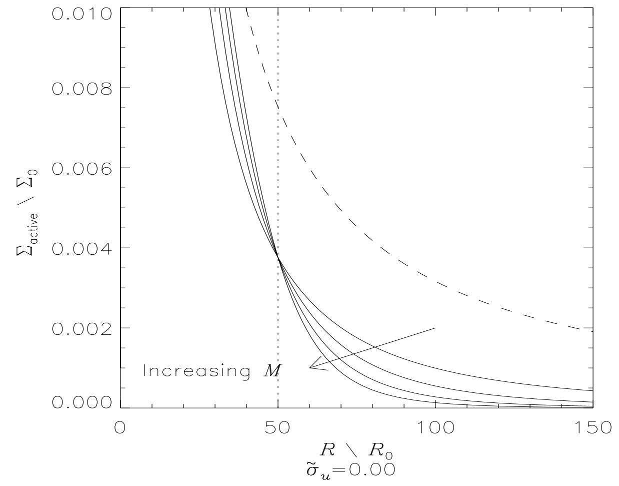


Figure 2.15: The active surface density of a $\beta = 0.25$ disk with a doubly cut-out function. The left-hand plot is for a cold disk, where the density is given by eq. (2.86). The right-hand plot is for $\tilde{\sigma}_u = 0.283$, which is the temperature at which the self-consistent disk is locally stable to axisymmetric disturbances. The solid lines are the density for $M = 1$, 2, 3, 4; the arrows indicate the direction of increasing M. In each case $N = 2$, $\tilde{R}_{\rm c} = 50$. The dashed line is the equilibrium surface density. The dotted line marks the position of $\tilde{R}_{\rm c}$.

Fig. 2.16 shows the surface density as a fraction of the equilibrium density. Once again, this is similar to the cut-out function $H(L_z)$.

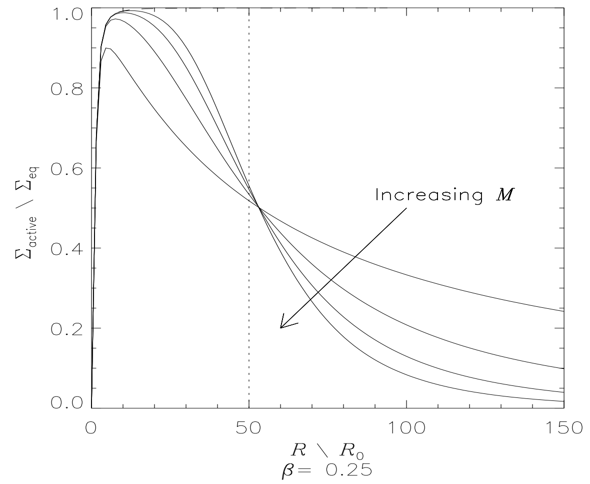
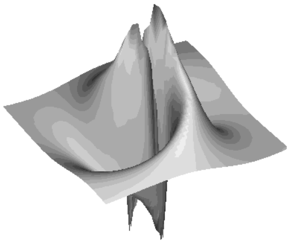


Figure 2.16: The active surface density of hot disks with a doubly cut-out function, shown as a fraction of the equilibrium density. The solid lines are the density for $M = 1$, 2, 3, 4; the arrows indicate the direction of increasing M. In each case $N = 2$, $\tilde{R}_{\rm c} = 50$. The dashed line just visible at the top of the plots is the corresponding inner cut-out disk. The dotted line marks the position of $\tilde{R}_{\rm c}$. For $\beta = +0.25$, $\tilde{\sigma}_u = 0.283$; for $\beta = -0.25$, $\tilde{\sigma}_u = 0.509$. These are the temperatures at which the self-consistent disk is locally stable to axisymmetric disturbances.

Our motivation for introducing the outer cut-off is to enable our results to be directly compared against N-body work. In practice, however, the outer cut-out does not have a significant effect on the stability properties of the disk (unless $\tilde{L}_{\rm c}$ and $\tilde{R}_{\rm c}$ are very small).



Chapter 3

The Integral Equation

3.1 Introduction

In this chapter we investigate the consequences of imposing a small density perturbation Σ_{imp} on the equilibrium disk. If the density of the disk changes, the potential and hence the forces on the stars alter, and the stars' orbits shift. This in turn leads to a change in the density Σ_{res}, which itself causes a change in the potential and so on. For such a disturbance to be a normal mode, it must be *self-consistent*[1] – that is, the imposed density disturbance must generate a response exactly equal to the imposed density. If there are no growing self-consistent disturbances, the disk is stable. If there is a self-consistent disturbance with vanishingly small growth rate, the disk is said to be marginally stable. One of the goals of stability analyses is to find the condition of marginal stability, which separates the stable and unstable modes. In this thesis, we make the step of *linearising* the Boltzmann equation, which effectively assumes that the perturbation amplitude is small. This is a fair assumption, since many galaxies are axisymmetric to a first approximation, with their spiral pattern superimposed as a small ($\approx 10\%$) perturbation (Bertin & Lin 1996).

We begin by deriving a general relationship between a change in density Σ_{imp} and the consequent change in potential ψ_{imp}. We then use the linearised collisionless Boltzmann equation to find how this change in potential ψ_{imp} affects the distribution function. The change in the distribution function f_{imp} depends on the equilibrium distribution function$f$. Integrating f_{imp} over velocity is then our route back to finding the change in density Σ_{res} caused by ψ_{imp}. We look for self-consistent solutions by requiring that Σ_{res} be equal to Σ_{imp}. In practice, we implement this general scheme by expanding Σ_{imp} in terms of a particular set of functions called logarithmic spirals, effectively taking the Fourier transform of the density distribution. We follow the scheme outlined above for a single log-spiral component $\Sigma_{\text{imp}}^{\alpha m}$ of the density, and derive the change in potential $\psi_{\text{imp}}^{\alpha m}$ and in distribution function $f_{\text{imp}}^{\alpha m}$ caused by each density component $\Sigma_{\text{imp}}^{\alpha m}$.

---

[1] We do not require our distribution functions to be self-consistent, since we are happy to ascribe the missing potential to an inactive halo component. However, in our stability analysis, we are interested in self-sustaining modes of the disk. We therefore search for self-consistent disturbances. Note that it is perfectly possible to have a self-consistent mode in a disk with a cut-out distribution function.

We can integrate this over velocity space to find the change in density $\Sigma_{\text{res}}^{\alpha m}$ caused by each log-spiral component. We then add up all the log-spiral components in order to find the total response density Σ_{res}, and compare this to Σ_{imp}. However, in practice it is more useful to consider the *transforms* of these two densities. These must be equal for a self-consistent mode. The goal of this chapter is therefore to derive the integral equation relating the transform of the response density to the transform of the imposed density. In the course of this work, several new variables and functions will be introduced. These are summarised in the tables in Appendix A.

3.2 Dynamics of perturbed disks

We have a disk with surface density $\Sigma_{\text{eq}}(R)$ and potential $\psi(R)$ satisfying Poisson's equation

$$\nabla^2\psi = 4\pi G\Sigma_{\text{eq}}(R)\cdot\delta(z), \tag{3.1}$$

where δ is the Dirac delta-function. Now imagine that a small perturbation Σ_{imp} is imposed on the surface density, so that the total density at any point is now $\Sigma_{\text{eq}} + \Sigma_{\text{imp}}$. This in turn brings about a small change ψ_{imp} in the potential. The new surface density $\Sigma_{\text{eq}} + \Sigma_{\text{imp}}$ and the new potential $\psi + \psi_{\text{imp}}$ must still obey Poisson's equation. The linearity of Poisson's equation implies

$$\nabla^2\psi_{\text{imp}} = 4\pi G\Sigma_{\text{imp}}(R)\cdot\delta(z), \tag{3.2}$$

This has the solution

$$\psi_{\text{imp}}\left(R,\theta\right) = G\int_{\theta'=0}^{\theta'=2\pi}\int_{R'=0}^{R'=\infty}\frac{\Sigma_{\text{imp}}\left(R',\theta'\right)R'\,dR'\,d\theta'}{\sqrt{R^2+R'^2-2RR'\cos\left(\theta-\theta'\right)}}. \tag{3.3}$$

We transform to $X = \ln R$ and introduce the *reduced potential* Y and the *reduced surface density* S (Binney & Tremaine 1987, ch. 2, sec. 6):

$$Y\left(X,\theta\right) = R^{\frac{1}{2}}\psi_{\text{imp}}\left(R,\theta\right), \tag{3.4}$$

$$S\left(X,\theta\right) = R^{\frac{3}{2}}\Sigma_{\text{imp}}\left(R,\theta\right). \tag{3.5}$$

Then (3.3) becomes

$$Y\left(X,\theta\right) = G\int_{\theta'=0}^{\theta'=2\pi}\int_{X'=-\infty}^{X'=+\infty}K'\left(X-X',\theta-\theta'\right)S\left(X',\theta'\right)dX'\,d\theta', \tag{3.6}$$

where

$$K'\left(X-X',\theta-\theta'\right) = \left(2\left[\cosh\left(X-X'\right)-\cos\left(\theta-\theta'\right)\right]\right)^{-\frac{1}{2}}. \tag{3.7}$$

The stars move in response to the change in potential, and so the distribution function has changed from its equilibrium value f to a new value $f + f_{\rm imp}$. This new distribution function is not an integral of motion, because it changes with time. However, it must still satisfy the collisionless Boltzmann equation (2.45), so we must have

$$\frac{\partial}{\partial t}\left(f + f_{\rm imp}\right) + \mathbf{v}\cdot\nabla\left(f + f_{\rm imp}\right) + \nabla\left(\psi + \psi_{\rm imp}\right)\frac{\partial}{\partial \mathbf{v}}\left(f + f_{\rm imp}\right) = 0. \tag{3.8}$$

Unlike Poisson's equation, the Boltzmann equation is not linear in the perturbation. So after cancelling the equilibrium terms, we are left with

$$\frac{\partial f_{\rm imp}}{\partial t} + \mathbf{v}\cdot\nabla f_{\rm imp} + \nabla\psi_{\rm imp}\frac{\partial f}{\partial \mathbf{v}} + \nabla\psi\frac{\partial f_{\rm imp}}{\partial \mathbf{v}} + \nabla\psi_{\rm imp}\frac{\partial f_{\rm imp}}{\partial \mathbf{v}} = 0. \tag{3.9}$$

We then make the step of *linearising* the Boltzmann equation, that is, neglecting the term of second order in the perturbation, to obtain

$$\frac{\partial f_{\rm imp}}{\partial t} + \mathbf{v}\cdot\nabla f_{\rm imp} + \nabla\psi\frac{\partial f_{\rm imp}}{\partial \mathbf{v}} = -\nabla\psi_{\rm imp}\frac{\partial f}{\partial \mathbf{v}}. \tag{3.10}$$

The left-hand side of this equation is, by definition, the convective derivative $df_{\rm imp}/dt$: the change in $f_{\rm imp}$ witnessed by an observer moving with a star along the unperturbed trajectory. Using this fact, and expanding the right-hand side, we have

$$\frac{df_{\rm imp}}{dt} = \frac{\partial f_{\rm imp}}{\partial t} + \mathbf{v}\cdot\nabla f_{\rm imp} + \nabla\psi\frac{\partial f_{\rm imp}}{\partial \mathbf{v}} = -\frac{\partial \psi_{\rm imp}}{\partial R}\frac{\partial f}{\partial u} - \frac{1}{R}\frac{\partial \psi_{\rm imp}}{\partial \theta}\frac{\partial f}{\partial v}. \tag{3.11}$$

Now the equilibrium distribution function f depends only on E and L_z, so we rewrite the derivatives as

$$\frac{\partial f}{\partial u} = \frac{\partial f}{\partial E}\frac{\partial E}{\partial u} + \frac{\partial f}{\partial L_z}\frac{\partial L_z}{\partial u} = u\frac{\partial f}{\partial E}, \tag{3.12}$$

$$\frac{\partial f}{\partial v} = \frac{\partial f}{\partial E}\frac{\partial E}{\partial v} + \frac{\partial f}{\partial L_z}\frac{\partial L_z}{\partial v} = v\frac{\partial f}{\partial E} + R\frac{\partial f}{\partial L_z}. \tag{3.13}$$

Recasting (3.11), we obtain

$$\frac{df_{\rm imp}}{dt} = -\left(u\frac{\partial \psi_{\rm imp}}{\partial R} + \frac{v}{R}\frac{\partial \psi_{\rm imp}}{\partial \theta}\right)\frac{\partial f}{\partial E} - \frac{\partial \psi_{\rm imp}}{\partial \theta}\frac{\partial f}{\partial L_z}. \tag{3.14}$$

This relates the change in the distribution function to the forces experienced by the star as a result of the density perturbation. The radial and tangential forces on the star are respectively

$$F_R = \frac{\partial \psi_{\rm imp}}{\partial R}, \qquad\qquad F_\theta = \frac{1}{R}\frac{\partial \psi_{\rm imp}}{\partial \theta}. \tag{3.15}$$

Then the term in brackets in (3.14) is $(uF_R + vF_\theta)$ which is just the rate of change of energy of the star dE/dt. Similarly, the term $\partial\psi_{\rm imp}/\partial\theta$ is just the torque acting on the star, i.e. the rate of change of angular momentum dL_z/dt. So we have the result that

$$\frac{df_{\rm imp}}{dt} = -\frac{\partial f}{\partial E}\frac{dE}{dt} - \frac{\partial f}{\partial L_z}\frac{dL_z}{dt}. \tag{3.16}$$

To understand this equation, we recall that the number of stars must stay the same, even though their energy and angular momentum are no longer conserved. Stars are pulled out of their equilibrium orbits by the density perturbation, causing a change $f_{\rm imp}$ in the distribution function. The rate at which stars move to perturbed orbits, $df_{\rm imp}/dt$, must be equal to the rate at which stars leave equilibrium orbits. Now the equilibrium distribution function f does not depend on time, but it does depend on the energy E and angular momentum L_z. Previously these were constants for each star, but now they change as the star moves. We can therefore view f as depending on time through E and L_z. Then the rate at which stars leave equilibrium orbits is $-\frac{df}{dE}\cdot\frac{dE}{dt} - \frac{df}{dL_z}\cdot\frac{dL_z}{dt}$. Thus we arrive at a physical understanding of eq. (3.16): it states that the rate at which stars move to perturbed orbits is equal to the rate at which stars leave equilibrium orbits.

However, we are interested not so much in the rate of change of $f_{\rm imp}$, but in its value at a particular time t. For this, we must integrate (3.16) over the entire time of the perturbation, to obtain

$$f_{\rm imp}(t) = -\frac{\partial f}{\partial E}\Delta E - \frac{\partial f}{\partial L_z}\Delta L_z, \tag{3.17}$$

where ΔE is the change in the star's energy due to the perturbation,

$$\begin{aligned} \Delta E &= \int_{-\infty}^{t} \frac{dE}{dt'}dt', \\ &= \int_{-\infty}^{t} \left(u\left(t'\right)F_R + v\left(t'\right)F_\theta\right)dt' = \int_{-\infty}^{t}\left(u'\frac{\partial\psi_{\rm imp}}{\partial R'} + \frac{v'}{R'}\frac{\partial\psi_{\rm imp}}{\partial\theta'}\right)dt', \end{aligned} \tag{3.18}$$

and ΔL_z is the change in its angular momentum

$$\Delta L_z = \int_{-\infty}^{t}\frac{dL_z}{dt'}dt' = \int_{-\infty}^{t} R\left(t'\right)F_\theta dt' = \int_{-\infty}^{t}\frac{\partial\psi_{\rm imp}}{\partial\theta'}dt'. \tag{3.19}$$

These definitions assume that the perturbation vanished in the distant past.

3.3 The logarithmic spirals

A general imposed density perturbation $\Sigma_{\rm imp}$ can be expanded in a Fourier series of azimuthal harmonics, each with a different order m of rotational symmetry:

$$\Sigma_{\rm imp} = \sum_{m=0}^{\infty} e^{im\theta}g_m(R,t). \tag{3.20}$$

In the linear regime, the response of the disk to the change in potential has the same order m of rotational symmetry as the initial density perturbation imposed on the disk. This means that we can confine our investigations to a single value of m at a time. We therefore impose a density perturbation of the form $\Sigma_{\mathrm{imp}} = e^{im\theta} g(R,t)$. This causes a change in potential which moves the stars so as to give a response density $\Sigma_{\mathrm{res}} = e^{im\theta} h(R,t)$. We look for *self-consistent* density perturbations – that is, disturbances of the density which cause changes in the potential such as to move the stars into precisely the density pattern assumed by the initial disturbance, so that $g(R,t) = h(R,t)$.

We next expand the function $g(R,t)$ in terms of some basic density components. It makes sense for our density components to be as realistic as possible. Also, of course, they must be analytically convenient. Fortunately both requirements are met by the *logarithmic spirals.* These were made famous by Kalnajs (1965), and have surface density

$$\Sigma_{\mathrm{ls}} = \Sigma_{\mathrm{p}} e^{im\theta} \left(\frac{R_0}{R}\right)^{3/2-i\alpha}, \tag{3.21}$$

where Σ_{p} is a constant amplitude. We work with a complex perturbation density; as usual, it is to be understood that the real part of this quantity corresponds to the physical density. The power of $R^{-3/2}$ usefully cancels with the $R^{3/2}$ which appears in (3.5). Further, real galaxies are frequently observed to have nearly constant pitch angle (the angle between the spiral arm and a reference circle) (Bertin & Lin 1996). This property is exhibited by the log-spirals, which have constant pitch angle $\cot(\alpha/m)$. A general self-similar quantity in a disk must be of the form $q = R^b f(\theta - a \ln R)$ (Lynden-Bell & Lemos 1996). The log-spirals are of this form, and are thus suited to exploit the self-similarity of the power-law disks. To understand what a log-spiral density looks like, let us consider the real part of (3.21):

$$\mathrm{Re}\left(\Sigma_{\mathrm{ls}}\right) = \Sigma_{\mathrm{p}} \left(\frac{R_0}{R}\right)^{3/2} \cos\left(m\theta + \alpha \ln \frac{R}{R_0}\right). \tag{3.22}$$

As we have seen, m is the order of rotational symmetry of the density disturbance: i.e. the density is unchanged under $\theta \to \theta + 2\pi/m$. The pattern thus has m lobes. The logarithmic wavenumber α controls how tightly the spiral is wrapped. The pattern is more tightly wound for high values of α. We can see this by noting that, as we move out along a radius, successive maxima of Σ_{imp} will be separated by ΔR, where

$$\ln\left(\frac{R+\Delta R}{R_0}\right) = \ln\left(\frac{R}{R_0}\right) + \frac{2\pi}{\alpha}. \tag{3.23}$$

We see that

$$\frac{\Delta R}{R} \approx \frac{2\pi}{\alpha}. \tag{3.24}$$

For larger α, ΔR will be smaller, i.e. the maxima will be closer together.

Figs. 3.1 and 3.2 show log-spirals with different values of m and α; they are in units $R_0 = 1$. The first three plots show log-spirals with $m = 1$.

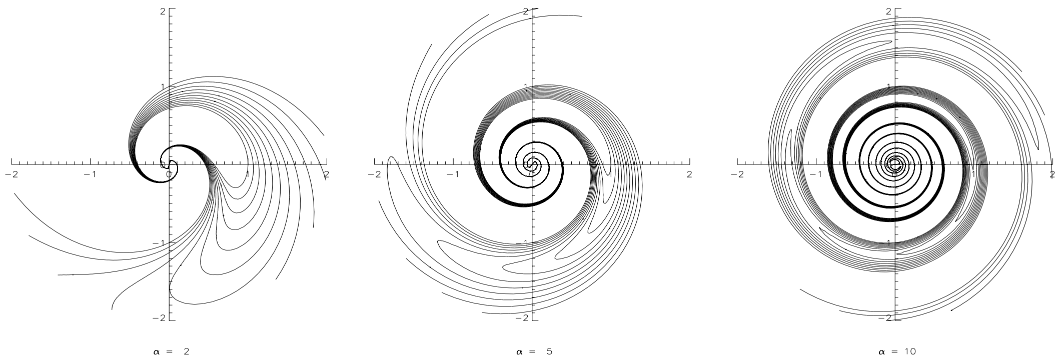


Figure 3.1: A log-spiral density perturbation with $m = 1$, and (left to right) $\alpha = 2$, 5, 10. Eleven contours are drawn at 0.1 intervals from 0 to 1, with $\Sigma_{\mathrm{p}} = 1$. For clarity, the under-density contours are not shown.

Fig. 3.2 shows log-spirals with $m = 2$.

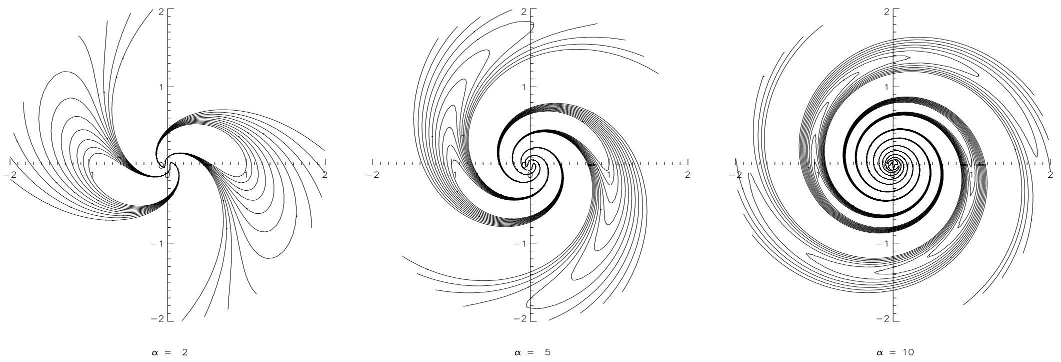


Figure 3.2: A log-spiral density perturbation with $m = 2$, and (left to right) $\alpha = 2$, 5, 10. Eleven contours are drawn at 0.1 intervals from 0 to 1, with $\Sigma_{\mathrm{p}} = 1$. For clarity, the under-density contours are not shown.

We allow our imposed density distribution to grow or decay by including a factor e^{st}, where s is the *growth rate*. We shall also allow it to rotate at a constant speed, the *pattern speed* Ω_{p}. We achieve this by replacing θ with $\theta - \Omega_{\mathrm{p}} t$. The density of a rotating logarithmic spiral is

$$\Sigma_{\mathrm{imp}}^{\alpha m} = \Sigma_{\mathrm{p}} e^{st} e^{im(\theta - \Omega_{\mathrm{p}} t)} \left(\frac{R_0}{R}\right)^{3/2 - i\alpha} = \Sigma_{\mathrm{p}} e^{i(m\theta - \omega t)} \left(\frac{R_0}{R}\right)^{3/2 - i\alpha} . \tag{3.25}$$

where the growth rate and pattern speed have been collected into the complex frequency $\omega = m\Omega_{\mathrm{p}} + is$. This decomposition rules out rotation of $m = 0$ modes. Purely on physical grounds, this can involve no loss of generality – nothing can be changed by rotating an axisymmetric pattern. From a mathematical point of view, Goldreich & Lynden-Bell (1965b) have proved for gas, and Kalnajs (1971) for stellar disks, that any exponentially growing or decaying axisymmetric mode with time-dependence $\exp(i\omega t)$ must have $\mathrm{Re}(\omega) = 0$.

By adding log-spirals with different α, we can build up a density distribution with rotational symmetry m:

$$\Sigma_{\text{imp}}\left(R,\theta\right) = \int_{-\infty}^{+\infty} d\alpha A_{\text{imp}}\left(\alpha\right) \Sigma_{\text{imp}}^{\alpha m}. \tag{3.26}$$

The expansion in log-spirals is equivalent to taking the Fourier transform in the variable $\tilde{x} = \ln(R/R_0)$. Defining

$$\Sigma_m = \Sigma_{\text{p}} e^{i(m\theta-\omega t)} e^{-3\tilde{x}/2} \tag{3.27}$$

we have the conventional transform pair:

$$\frac{\Sigma_{\text{imp}}}{\Sigma_m} = \int_{-\infty}^{+\infty} d\alpha\, A_{\text{imp}}\left(\alpha\right) e^{i\alpha\tilde{x}}, \qquad A_{\text{imp}}\left(\alpha\right) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\tilde{x}\, \frac{\Sigma_{\text{imp}}}{\Sigma_m} e^{-i\alpha\tilde{x}}. \tag{3.28}$$

A_{imp} is the density transform of the imposed perturbation. α is the logarithmic wavenumber corresponding to the logarithmic radial coordinate $\tilde{x}$. In future α will be referred to simply as the wavenumber.

In terms of R, the density transform is

$$A_{\text{imp}}\left(\alpha\right) = \frac{e^{-i(m\theta-\omega t)}}{2\pi\Sigma_{\text{p}} R_0^{3/2-i\alpha}} \int_0^{+\infty} \frac{dR}{R}\, \Sigma_{\text{imp}}\left(R,\theta\right) R^{3/2-i\alpha}. \tag{3.29}$$

For the transform to exist, $\Sigma_{\text{imp}}\left(R,\theta\right) R^{3/2}$ must tend to zero as $R \to 0$ and $R \to \infty$. $\Sigma_{\text{imp}}\left(R,\theta\right)$ can diverge at the centre no faster than $R^{-3/2}$, and must fall off at large radii more quickly than $R^{-3/2}$.

What effect does such an imposed density have on the potential experienced by the stars? All our equations are now linear, so we can consider the effect of each imposed density component separately. For a single log-spiral component, the reduced density (3.5) is:

$$S^{\alpha m}(X,\theta) = \Sigma_{\text{p}} R_0^{3/2-i\alpha} e^{i(m\theta-\omega t)} e^{i\alpha X}, \tag{3.30}$$

where $X = \ln R$. Substituting $S = S^{\alpha m}$ into (3.6), we obtain the reduced potential

$$\begin{aligned} Y^{\alpha m}\left(X,\theta\right) = {} & G\Sigma_{\text{p}} R_0^{3/2-i\alpha} e^{-i\omega t} \\ & \times \int_{\theta'=0}^{\theta'=2\pi} \int_{X'=-\infty}^{X'=+\infty} K'\left(X-X',\theta-\theta'\right) e^{i\left(\alpha X'+m\theta'\right)} dX'\, d\theta'. \end{aligned} \tag{3.31}$$

Substituting $X'' = X - X'$ and $\theta'' = \theta - \theta'$, this becomes

$$\begin{aligned} Y^{\alpha m}\left(X,\theta\right) = {} & G\Sigma_{\text{p}} R_0^{3/2-i\alpha} e^{i(\alpha X+m\theta-\omega t)} \\ & \times \int_{\theta'=0}^{\theta'=2\pi} \int_{X'=-\infty}^{X'=+\infty} K'\left(X'',\theta''\right) e^{-i\left(\alpha X''+m\theta''\right)} dX''\, d\theta''. \end{aligned} \tag{3.32}$$

It can be shown (Kalnajs 1971; Binney & Tremaine 1987, ch. 2, sec. 6) that

$$\int_{\theta'=0}^{\theta'=2\pi} \int_{X'=-\infty}^{X'=+\infty} K'\left(X'', \theta''\right) e^{-i\left(\alpha X''+m\theta''\right)} dX''\, d\theta'' = 2\pi K(\alpha, m), \tag{3.33}$$

where $K(\alpha, m)$ is the *Kalnajs gravity function*

$$K(\alpha, m) = \frac{1}{2} \frac{\Gamma\left[\frac{1}{2}\left(\frac{1}{2}+m+i\alpha\right)\right] \Gamma\left[\frac{1}{2}\left(\frac{1}{2}+m-i\alpha\right)\right]}{\Gamma\left[\frac{1}{2}\left(\frac{3}{2}+m+i\alpha\right)\right] \Gamma\left[\frac{1}{2}\left(\frac{3}{2}+m-i\alpha\right)\right]}. \tag{3.34}$$

This function is real and positive for real α (Kalnajs 1971). We note the symmetry $K(\alpha, m) = K(-\alpha, m)$.

Then we have

$$Y^{\alpha m}\left(X, \theta\right) = 2\pi G \Sigma_{\mathrm{p}} K(\alpha, m) R_0^{3/2-i\alpha} e^{i(\alpha X+m\theta-\omega t)}. \tag{3.35}$$

We have now recovered the results of Kalnajs (1965; 1971), giving the potential $\psi_{\mathrm{imp}}^{\alpha m}$ of each log-spiral component of the density perturbation $\Sigma_{\mathrm{imp}}^{\alpha m}$:

$$\psi_{\mathrm{imp}}^{\alpha m} = 2\pi G \Sigma_{\mathrm{p}} K(\alpha, m) R_0 e^{i(m\theta-\omega t)} \left(\frac{R}{R_0}\right)^{i\alpha-\frac{1}{2}}, \tag{3.36}$$

$$\Sigma_{\mathrm{imp}}^{\alpha m} = \Sigma_{\mathrm{p}} e^{i(m\theta-\omega t)} \left(\frac{R}{R_0}\right)^{i\alpha-3/2}. \tag{3.37}$$

Using the potential due to a single log-spiral component (3.36) in (3.15), we find that the radial force contributed by each component is

$$F_R\left(R, \theta\right) = 2\pi G \Sigma_{\mathrm{p}} K(\alpha, m) e^{i(m\theta-\omega t)} \left(i\alpha-\frac{1}{2}\right) \left(\frac{R}{R_0}\right)^{i\alpha-\frac{3}{2}}, \tag{3.38}$$

while the tangential force is

$$F_\theta\left(R, \theta\right) = 2\pi G \Sigma_{\mathrm{p}} K(\alpha, m) e^{i(m\theta-\omega t)} i m R_0 \left(\frac{R}{R_0}\right)^{i\alpha-\frac{1}{2}}. \tag{3.39}$$

For an axisymmetric perturbation ($m = 0$), the stars experience no tangential component of force.

3.4 The integral equation

To find the change in surface density $\Sigma_{\mathrm{res}}^{\alpha m}$ caused by a single log-spiral component $\Sigma_{\mathrm{imp}}^{\alpha m}$, we must integrate $f_{\mathrm{imp}}^{\alpha m}$ over all velocities u and v

$$\Sigma_{\mathrm{res}}^{\alpha m} = \iint f_{\mathrm{imp}}^{\alpha m}\, du\, dv. \tag{3.40}$$

To find the *total* change in density caused by the whole disturbance $\Sigma_{\rm imp} = \int d\alpha\, A_{\rm imp}\,(\alpha)\, \Sigma^{\alpha m}_{\rm imp}$, we integrate over all the log-spiral components

$$\Sigma_{\rm res} = \int_{-\infty}^{+\infty} d\alpha\, A_{\rm imp}\,(\alpha)\, \Sigma^{\alpha m}_{\rm res}. \tag{3.41}$$

We could now equate $\Sigma_{\rm res}$ and $\Sigma_{\rm imp}$, and seek to solve the resulting equation for self-consistent solutions. However, a more profitable approach is to relate the density transforms. We define the response transform $A_{\rm res}\,(\alpha)$ analogously to $A_{\rm imp}\,(\alpha)$ in (3.28), so that

$$\frac{\Sigma_{\rm res}}{\Sigma_m} = \int_{-\infty}^{+\infty} d\alpha\, A_{\rm res}\,(\alpha)\, e^{i\alpha\tilde{x}}. \tag{3.42}$$

Then, taking the Fourier transform, we obtain an analogue of (3.28)

$$A_{\rm res}\,(\alpha) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\tilde{x}\, \frac{\Sigma_{\rm res}}{\Sigma_m} e^{-i\alpha\tilde{x}}. \tag{3.43}$$

Substituting for $\Sigma_{\rm res}$ from (3.41), this becomes

$$A_{\rm res}\,(\alpha) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\tilde{x}\, \frac{e^{-i\alpha\tilde{x}}}{\Sigma_m} \int_{-\infty}^{+\infty} d\alpha' A_{\rm imp}\,(\alpha')\, \Sigma^{\alpha' m}_{\rm res}. \tag{3.44}$$

If we exchange the order of integration, we obtain (cf. Z3.42):

$$A_{\rm res}\,(\alpha) = \int_{-\infty}^{+\infty} d\alpha' A_{\rm imp}\,(\alpha')\, \mathcal{S}_m\,(\alpha, \alpha')\,, \tag{3.45}$$

where $$\mathcal{S}_m\,(\alpha, \alpha') = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\tilde{x}\, \frac{\Sigma^{\alpha' m}_{\rm res}}{\Sigma_m} e^{-i\alpha\tilde{x}}. \tag{3.46}$$

Self-consistency requires $A_{\rm imp}\,(\alpha) = A_{\rm res}\,(\alpha)$. $\mathcal{S}_m$ is the *transfer function.* Suppose we disturb the disk by applying a pure logarithmic spiral disturbance with a particular wavenumber α' (i.e. $A_{\rm imp}(\alpha'') = \delta(\alpha'' - \alpha')$). The disk will respond with a pattern made up of components with many different wavenumbers: just as a flute, with an almost sinusoidal waveform, might excite a church bell to resonate at many different frequencies. $\mathcal{S}_m(\alpha, \alpha')$ tells us how much of the disk's response to the disturbance with wavenumber α' occurs at the particular wavenumber α.

Kalnajs (1971) derived a completely general integral equation in action-angle coordinates. Palmer & Papaloizou (1990) restricted their attention to disks built from epicyclic orbits. Our integral equation has the same physical content as those of the earlier workers, but is adapted to exploit fully the self-similarity of the equilibrium disk.

3.5 The transfer function

The next step is to find a way of evaluating the transfer function $\mathcal{S}_m$ (3.46). In this section we introduce several more variables. These are all collected for reference in Appendix A. First of all, we rewrite the expression for ΔE (3.18) as

$$\Delta E = \int_{-\infty}^{t} \left(\frac{d\psi_{\rm imp}}{dt'} - \frac{\partial \psi_{\rm imp}}{\partial t'} \right) dt' = \psi_{\rm imp}(t) - \int_{-\infty}^{t} \frac{\partial \psi_{\rm imp}}{\partial t'} dt'. \tag{3.47}$$

ΔE is the difference between the potential here and now, and an averaged potential sampled by the orbit over its history. From eq. (3.36), we obtain

$$\frac{\partial \psi_{\rm imp}^{\alpha m}}{\partial t} = -i\omega \psi_{\rm imp}^{\alpha m}(t), \qquad \frac{\partial \psi_{\rm imp}^{\alpha m}}{\partial \theta} = im\psi_{\rm imp}^{\alpha m}(t). \tag{3.48}$$

Then the changes in energy and angular momentum due to a single log-spiral component are

$$\Delta E = \psi_{\rm imp}^{\alpha m}(t) + i\omega \int_{-\infty}^{t} \psi_{\rm imp}^{\alpha m}(t')\, dt', \qquad \Delta L_z = im \int_{-\infty}^{t} \psi_{\rm imp}^{\alpha m}(t')\, dt'. \tag{3.49}$$

The problem with evaluating $f_{\rm imp}$ for a non-axisymmetric perturbation arises from the integrands in these equations. In the axisymmetric case, $\psi_{\rm imp}$ has no dependence on θ, and therefore is periodic with frequency κ. But for $m \neq 0$ the integrand involves both radial and azimuthal frequencies κ and Ω. In general the orbit is not closed, i.e. κ and Ω are not commensurable, and the integrands are not periodic. Improper integrals are slow to converge numerically, especially in a case like this when the integrand does not fall off rapidly as $t' \to -\infty$. We can avoid such unpleasantness by shifting to the frame which rotates at the star's average velocity Ω, in which its orbit closes.

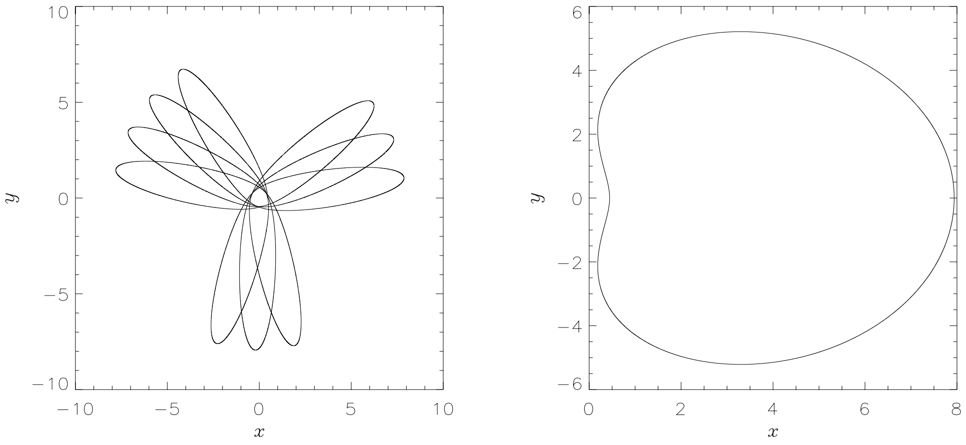


Figure 3.3: The left-hand plot shows the orbit of a star with $U = 1.5$ and $R_H = 1$, in the $\beta = 0.25$ disk. The right-hand plot shows the same orbit as viewed by an observer rotating with the star's average angular velocity $\Omega = 0.167$ (units are $R_0 = v_\beta = 1$).

We define two dimensionless coordinates describing the periodic excursions of the stellar orbits:

$$\tilde{X} = \ln \frac{R}{R_H} = \ln \tilde{R}, \qquad \tilde{Y} = \theta - \Omega t = \theta - \tilde{\Omega}\tilde{t}. \tag{3.50}$$

The following plot shows how $\tilde{X}$ and $\tilde{Y}$ vary over four radial periods for the orbit shown in fig. 3.3.

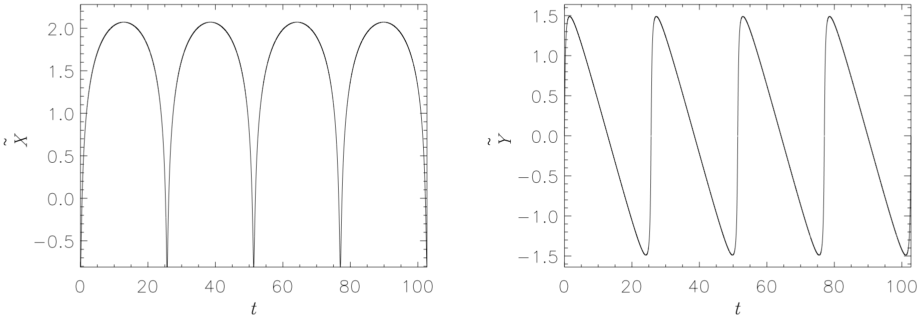


Figure 3.4: Excursions of a stellar orbit. The left-hand plot shows the scaled logarithmic radius $\tilde{X}$, and the right-hand the deviation from mean angular motion $\tilde{Y}$, against time. [$U = 1.5$, $R_H = 1$, $\beta = 0.25$, $R_0 = v_\beta = 1$.]

The potential perturbation ψ_{imp} (3.36) can be rewritten in terms of $\tilde{X}$ and $\tilde{Y}$

$$\psi_{\text{imp}}^{\alpha m}(t) = 2\pi G\Sigma_{\text{p}} K(\alpha, m) R_0 \tilde{R}_H^{i\alpha-\frac{1}{2}} \exp\left\{i(m\Omega - \omega)t + im\tilde{Y} + (i\alpha - \tfrac{1}{2})\tilde{X}\right\}. \tag{3.51}$$

This is the potential at time t when the star is at (R, θ). In order to find the value of f_{imp} at our present time t, we shall need to evaluate the potential at previous times t', when the star's position was (R', θ'), and $\tilde{X}' = \ln \tilde{R}'$, $\tilde{Y}' = \theta' - \Omega t'$. This is

$$\psi_{\text{imp}}^{\alpha m}(t') = 2\pi G\Sigma_{\text{p}} K(\alpha, m) R_0 \tilde{R}_H^{i\alpha-\frac{1}{2}} \exp\left\{i(m\Omega - \omega)t' + im\tilde{Y}' + (i\alpha - \tfrac{1}{2})\tilde{X}'\right\}. \tag{3.52}$$

Now $\tilde{X}$ and $\tilde{Y}$ both have period T, so the term $\exp\left\{im\tilde{Y}' + (i\alpha - \frac{1}{2})\tilde{X}'\right\}$ is similarly periodic as t' varies. It can therefore be expanded in a Fourier series (a method used by Kalnajs (1972) and Zang (1976)):

$$\exp\left\{im\tilde{Y}' + (i\alpha - \tfrac{1}{2})\tilde{X}'\right\} = \sum_{l=-\infty}^{+\infty} Q_{lm}(\alpha) \exp\left\{\frac{2i\pi l}{T} t'\right\}, \tag{3.53}$$

where the Fourier coefficient $Q_{lm}(\alpha)$ is given by

$$Q_{lm}(\alpha) = \frac{1}{T} \int_0^T \exp\left\{im\tilde{Y}' + (i\alpha - \tfrac{1}{2})\tilde{X}' - \frac{2i\pi l}{T} t'\right\} dt'. \tag{3.54}$$

If we change variables to the orbital phase $\chi = \kappa t$, the Fourier coefficient becomes (Z3.27)

$$Q_{lm}(\alpha) = \frac{1}{2\pi} \int_0^{2\pi} \exp\left\{im\tilde{Y}' + (i\alpha - \tfrac{1}{2})\tilde{X}' - il\chi'\right\} d\chi'. \tag{3.55}$$

Substituting this into our expression for $\psi_{\text{imp}}^{\alpha m}$, we obtain

$$\psi_{\text{imp}}^{\alpha m}(t') = 2\pi G\Sigma_{\text{p}} K(\alpha, m) R_0 \tilde{R}_H^{i\alpha-\frac{1}{2}} e^{i(m\Omega-\omega)t'} \sum_{l=-\infty}^{+\infty} Q_{lm}(\alpha) e^{il\kappa t'}. \tag{3.56}$$

We can easily integrate this to give

$$\int_{-\infty}^{t} \psi_{\mathrm{imp}}^{\alpha m}(t')dt' = -i2\pi G\Sigma_{\mathrm{p}} K(\alpha,m) R_0 \tilde{R}_H^{i\alpha-\frac{1}{2}} e^{i(m\Omega-\omega)} \sum_{l=-\infty}^{+\infty} \frac{Q_{lm}(\alpha)e^{il\chi}}{l\kappa+m\Omega-\omega}. \tag{3.57}$$

We now substitute for $\psi_{\mathrm{imp}}^{\alpha m}$ in (3.49) to obtain:

$$\Delta E = 2\pi G\Sigma_{\mathrm{p}} K(\alpha,m) R_0 \tilde{R}_H^{i\alpha-\frac{1}{2}} e^{i(m\Omega-\omega)t} \sum_{l=-\infty}^{+\infty} Q_{lm}(\alpha)e^{il\chi} \frac{l\kappa+m\Omega}{l\kappa+m\Omega-\omega}, \tag{3.58}$$

$$\Delta L_z = 2\pi m G\Sigma_{\mathrm{p}} K(\alpha,m) R_0 \tilde{R}_H^{i\alpha-\frac{1}{2}} e^{i(m\Omega-\omega)t} \sum_{l=-\infty}^{+\infty} \frac{Q_{lm}(\alpha)e^{il\chi}}{l\kappa+m\Omega-\omega}. \tag{3.59}$$

These correspond to (Z3.29) and (Z3.30), although Zang there uses a slightly different form of the Fourier coefficient.

Substituting these expressions into (3.17) gives:

$$\begin{aligned} f_{\mathrm{imp}}^{\alpha m}(t) &= -2\pi G\Sigma_{\mathrm{p}} K(\alpha,m) R_0 \tilde{R}_H^{i\alpha-\frac{1}{2}} e^{i(m\Omega-\omega)t} \\ &\quad \times \sum_{l=-\infty}^{+\infty} \frac{Q_{lm}(\alpha)e^{il\chi}}{l\kappa+m\Omega-\omega} \left\{ (l\kappa+m\Omega)\frac{\partial f}{\partial E} + m\frac{\partial f}{\partial L_z} \right\}. \end{aligned} \tag{3.60}$$

So far, our results have been quite general. We now specialise to the power-law disks by substituting for the derivatives of the distribution function (2.81). Bearing in mind that the energy E has the opposite sign to β, we can combine the derivatives obtained for positive and negative β.

$$\frac{\partial f}{\partial E} = -\tilde{C}\left|\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}\right| H(L_z) L_z^{\gamma} |E|^{1/\beta+\gamma/\beta-\gamma/2-1}, \tag{3.61}$$

$$\frac{\partial f}{\partial L_z} = \tilde{C} L_z^{\gamma} |E|^{1/\beta+\gamma/\beta-\gamma/2} \left(\gamma\frac{H(L_z)}{L_z} + \frac{dH}{dL_z}\right). \tag{3.62}$$

We now obtain

$$\begin{aligned} f_{\mathrm{imp}}^{\alpha m}(t) &= 2\pi G\Sigma_{\mathrm{p}} K(\alpha,m) \tilde{C} R_0 \tilde{R}_H^{i\alpha-\frac{1}{2}} e^{i(m\Omega-\omega)t} L_z^{\gamma} |E|^{1/\beta+\gamma/\beta-\gamma/2} \\ &\quad \times \sum_{l=-\infty}^{+\infty} \frac{Q_{lm}(\alpha)e^{il\chi}}{l\kappa+m\Omega-\omega} \left[\left\{ (l\kappa+m\Omega)\left|\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}\right| \frac{1}{|E|} - \frac{\gamma m}{L_z} \right\} H(L_z) - m\frac{dH}{dL_z} \right]. \end{aligned} \tag{3.63}$$

Replacing the dimensional quantities with their dimensionless analogues (eqs. (2.19), (2.33), (2.37)) [2]

$$L_z = v_\beta R_0 \tilde{L}_z, \qquad \frac{dH}{dL_z} = \frac{1}{v_\beta R_0}\frac{d\tilde{H}}{d\tilde{L}_z}, \qquad E = \tilde{E} v_\beta^2 \tilde{L}_z^{\frac{2\beta}{\beta-2}}, \tag{3.64}$$

$$\kappa = \tilde{\kappa}\frac{v_\beta}{R_0}\tilde{L}_z^{\frac{2+\beta}{\beta-2}}, \qquad \Omega = \tilde{\Omega}\frac{v_\beta}{R_0}\tilde{L}_z^{\frac{2+\beta}{\beta-2}}, \qquad \omega = \tilde{\omega}\frac{v_\beta}{R_0}, \tag{3.65}$$

---

[2] At first sight it seems odd to define the dimensionless analogue of ω as $\tilde{\omega} = \omega R_0/v_\beta$ when the other frequencies in the problem, $\tilde{\kappa}$ and $\tilde{\Omega}$, are defined as $\tilde{\kappa} = \kappa R_0 \tilde{R}_H^{1+\beta/2}/v_\beta$. This choice of $\tilde{\omega}$ results in the presence of $\tilde{L}_z^{(2+\beta)/(2-\beta)} = \tilde{R}_H^{1+\beta/2}$ in the expression for $f_{\mathrm{imp}}^{\alpha m}$, which would otherwise be absorbed into $\tilde{\omega}$. It is helpful to keep this factor visible. The point is that the dimensionless quantity $\omega R_0 \tilde{R}_H^{1+\beta/2}/v_\beta$ really does depend on $\tilde{L}_z$, whereas the factors of $\tilde{R}_H$ in $\tilde{\kappa}$ and $\tilde{\Omega}$ cancel, so that $\tilde{\kappa}$ and $\tilde{\Omega}$ in fact depend only on $\tilde{U}$, and so can be calculated outside the integral over $\tilde{L}_z$. Defining $\tilde{\omega}$ similarly might lead to the erroneous assumption that it too was independent of $\tilde{L}_z$.

we obtain

$$
\begin{aligned}
f_{\mathrm{imp}}^{\alpha m}(t) = {} & 2\pi G \Sigma_{\mathrm{p}} K(\alpha, m) \tilde{C} R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)-2} e^{i(m\Omega-\omega)t} e^{i\alpha \frac{2}{2-\beta} \ln \tilde{L}_z} \\
& \times \tilde{L}_z^{\frac{2\beta-3}{2-\beta}} |\tilde{E}|^{1/\beta+\gamma/\beta-\gamma/2} \sum_{l=-\infty}^{\infty} \frac{Q_{lm}(\alpha) e^{il\chi}}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} \tilde{L}_z^{\frac{2+\beta}{2-\beta}}} \\
& \times \left[\left\{ (l\tilde{\kappa} + m\tilde{\Omega}) \left| \frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2} \right| \frac{1}{|\tilde{E}|} - \gamma m \right\} \tilde{H}(\tilde{L}_z) - m\tilde{L}_z \frac{d\tilde{H}}{d\tilde{L}_z} \right].
\end{aligned}
\tag{3.66}
$$

This is the change in the distribution function brought about by a single log-spiral component.

We can now proceed to find the transfer function $\mathcal{S}_m$. Substituting for $\Sigma_{\mathrm{res}}^{\alpha' m}$ in (3.46), we have:

$$
\mathcal{S}_m(\alpha, \alpha') = \frac{1}{2\pi} \int_{-\infty}^{+\infty} d\tilde{x} \, \frac{e^{-i\alpha\tilde{x}}}{\Sigma_m} \iint f_{\mathrm{imp}}^{\alpha' m} \, du \, dv.
$$

We follow Zang in transforming this integral to one over the eccentric velocity $\tilde{U}$, the orbital phase χ and the dimensionless angular momentum $\tilde{L}_z$. We write

$$
\iiint d\tilde{x} \, du \, dv = \iiint d\tilde{U} \, d\tilde{L}_z \, d\chi \frac{\partial(\tilde{x}, u, v)}{\partial(\chi, \tilde{U}, \tilde{L}_z)}. \tag{3.67}
$$

We work out the Jacobian in two stages. First we transform to variables $\tilde{U}$, $\tilde{L}_z$, $\tilde{R}$. We can express $\tilde{x}$, u and v as functions of $\tilde{U}$, $\tilde{L}_z$ and $\tilde{R}$:

$$
\tilde{x} = \ln \tilde{R} + \frac{2}{2-\beta} \ln \tilde{L}_z, \quad u^2 = v_\beta^2 \tilde{L}_z^{\frac{-2\beta}{2-\beta}} \left[\tilde{U}^2 + 1 - \tilde{R}^{-2} + \frac{2}{\beta} \left(\tilde{R}^{-\beta} - 1 \right) \right], \quad v = v_\beta \tilde{L}_z^{\frac{-\beta}{2-\beta}} \tilde{R}^{-1}. \tag{3.68}
$$

Then

$$
du \, dv \, d\tilde{x} = v_\beta^2 \frac{\tilde{U}}{\tilde{u}\tilde{R}^2} \tilde{L}_z^{-\frac{2+\beta}{2-\beta}} d\tilde{U} \, d\tilde{L}_z \, d\tilde{R}. \tag{3.69}
$$

The integrals of motion $\tilde{U}$ and $\tilde{L}_z$ define the orbit of the star. The dimensionless variable $\tilde{R}$ then measures whereabouts the star is in that orbit, i.e. the orbital phase. Using the dimensionless equations of motion, we have

$$
\frac{d\tilde{R}}{d\tilde{t}} = \tilde{u}, \qquad \frac{d\tilde{R}}{d\chi} = \frac{\tilde{u}}{\tilde{\kappa}} = \frac{\tilde{u}\mathcal{J}_0(\tilde{U})}{2\pi}. \tag{3.70}
$$

Thus we obtain the desired Jacobian, which we shall write as

$$
\frac{\partial(\tilde{x}, u, v)}{\partial(\chi, \tilde{U}, \tilde{L}_z)} = \frac{v_\beta^2}{2\pi} \frac{\mathcal{J}_0(\tilde{U})\tilde{U}}{\tilde{R}} e^{-\tilde{x}} \tilde{L}_z^{\frac{\beta}{\beta-2}}. \tag{3.71}
$$

We then have

$$\mathcal{S}_m(\alpha,\alpha') = \frac{v_\beta^2}{2\pi} \iiint d\tilde{U}\, d\chi\, d\tilde{L}_z \frac{e^{-(i\alpha+1)\tilde{x}}}{\Sigma_m} \frac{\mathcal{J}_0(\tilde{U})}{2\pi} \frac{\tilde{U}}{\tilde{R}} \tilde{L}_z^{\frac{\beta}{\beta-2}} f_{\text{imp}}^{\alpha' m}(\tilde{L}_z,\tilde{U},\chi). \tag{3.72}$$

The angular momentum is integrated from zero to infinity, and the orbital phase from 0 to 2π. The eccentric velocity is integrated from zero to infinity for negative β; for positive β the upper limit is $(2/\beta-1)^{1/2}$. For brevity, these limits of the integration are not shown explicitly.

We shall work in terms of $\tilde{X} = \ln\tilde{R} = \ln(R/R_H) = \tilde{x} - \frac{2}{2-\beta}\ln\tilde{L}_z$. Then $\tilde{R} = e^{\tilde{X}}$, and when we substitute for Σ_m from (3.27) we obtain

$$\begin{aligned}\mathcal{S}_m(\alpha,\alpha') = \frac{v_\beta^2}{2\pi\Sigma_{\text{p}}}\frac{1}{2\pi} \iiint d\tilde{U}\, d\chi\, d\tilde{L}_z e^{-i(m\theta-\omega t)} e^{-(i\alpha+\frac{1}{2})\tilde{X}} \\ \times\, e^{\frac{1-\beta-2i\alpha}{2-\beta}\ln\tilde{L}_z} \mathcal{J}_0(\tilde{U})\tilde{U} f_{\text{imp}}^{\alpha' m}(\tilde{L}_z,\tilde{U},\chi).\end{aligned} \tag{3.73}$$

Note that t, $\tilde{X}$ and θ describe where the star is in its orbit, and therefore depend on the orbital phase χ, as well as on $\tilde{U}$ which describes the shape of the orbit. Substituting for $f_{\text{imp}}^{\alpha' m}$ from (3.66):

$$\begin{aligned}\mathcal{S}_m(\alpha,\alpha') = \frac{1}{2\pi} R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} G K(\alpha',m)\tilde{C} \iiint d\tilde{U}\, d\chi\, \frac{d\tilde{L}_z}{\tilde{L}_z} e^{-im(\theta-\Omega t)} e^{-(i\alpha+\frac{1}{2})\tilde{X}} \\ \times\, e^{\frac{-2i}{2-\beta}(\alpha-\alpha')\ln\tilde{L}_z} \tilde{U}\mathcal{J}_0(\tilde{U}) |\tilde{E}|^{1/\beta+\gamma/\beta-\gamma/2} \sum_{l=-\infty}^{+\infty} \frac{Q_{lm}(\alpha')e^{il\chi}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}\tilde{L}_z^{\frac{2+\beta}{2-\beta}}} \\ \times \left[\left\{(l\tilde{\kappa}+m\tilde{\Omega})\left|\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}\right|\frac{1}{|\tilde{E}|} - \gamma m\right\}\tilde{H}(\tilde{L}_z) - m\tilde{L}_z\frac{d\tilde{H}}{d\tilde{L}_z}\right].\end{aligned} \tag{3.74}$$

Substituting $\tilde{E} = (\beta\tilde{U}^2+\beta-2)/(2\beta)$ and $\theta - \Omega t = \tilde{Y}(\chi)$,

$$\begin{aligned}\mathcal{S}_m(\alpha,\alpha') = R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} 2\pi G K(\alpha',m)\tilde{C} \int d\tilde{U}\mathcal{J}_0(\tilde{U})\tilde{U} \left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}} \\ \times \sum_{l=-\infty}^{+\infty} Q_{lm}(\alpha')\frac{1}{2\pi}\int d\chi e^{il\chi-im\tilde{Y}-(i\alpha+\frac{1}{2})\tilde{X}} \frac{1}{2\pi}\int \frac{d\tilde{L}_z}{\tilde{L}_z} \frac{e^{\frac{-2i}{2-\beta}(\alpha-\alpha')\ln\tilde{L}_z}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}\tilde{L}_z^{\frac{2+\beta}{2-\beta}}} \\ \times \left[\left\{(l\tilde{\kappa}+m\tilde{\Omega})\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \gamma m\right\}\tilde{H}(\tilde{L}_z) - m\tilde{L}_z\frac{d\tilde{H}}{d\tilde{L}_z}\right].\end{aligned} \tag{3.75}$$

From the definition of the Fourier coefficient (3.55), the integral over χ is just the complex conjugate of $Q_{lm}(\alpha)$. Then

$$\begin{aligned}\mathcal{S}_m(\alpha,\alpha') = R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} 2\pi G K(\alpha',m)\tilde{C} \int d\tilde{U}\mathcal{J}_0(\tilde{U})\tilde{U} \left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}} \\ \times \sum_{l=-\infty}^{+\infty} Q_{lm}(\alpha')Q_{lm}^*(\alpha)\frac{1}{2\pi}\int_0^\infty \frac{d\tilde{L}_z}{\tilde{L}_z} \frac{e^{\frac{-2i}{2-\beta}(\alpha-\alpha')\ln\tilde{L}_z}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}\tilde{L}_z^{\frac{2+\beta}{2-\beta}}} \\ \times \left[\left\{(l\tilde{\kappa}+m\tilde{\Omega})\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \gamma m\right\}\tilde{H}(\tilde{L}_z) - m\tilde{L}_z\frac{d\tilde{H}}{d\tilde{L}_z}\right].\end{aligned} \tag{3.76}$$

To make this expression a little more manageable, let us define the integral over angular momentum to be the *angular momentum function* $F_{lm}(\eta)$, where $\eta = \alpha - \alpha'$:

$$\begin{aligned} F_{lm}(\eta) = {} & \frac{1}{2\pi} \int_0^\infty \frac{e^{-i\eta \frac{2}{2-\beta} \ln \tilde{L}_z}}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega}\tilde{L}_z^{\frac{2+\beta}{2-\beta}}} \\ & \times \left[\left\{ (l\tilde{\kappa} + m\tilde{\Omega}) \left| \frac{2 + 2\gamma - \beta\gamma}{\beta\tilde{U}^2 + \beta - 2} \right| - \gamma m \right\} \tilde{H}(\tilde{L}_z) - m\tilde{L}_z \frac{d\tilde{H}}{d\tilde{L}_z} \right] \frac{d\tilde{L}_z}{\tilde{L}_z}. \end{aligned} \tag{3.77}$$

Note that F_{lm} depends on $\tilde{U}$ through the frequencies $\tilde{\kappa}$ and $\tilde{\Omega}$.

For $\beta = 0$ F_{lm} becomes (Zang 1976, p. 73)

$$F_{lm} = \frac{1}{2\pi} \int_0^\infty \frac{e^{-i\eta \ln \tilde{L}_z}}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega}\tilde{L}_z} \left[\left\{ (l\tilde{\kappa} + m\tilde{\Omega})(1+\gamma) - \gamma m \right\} \tilde{H}(\tilde{L}_z) - m\tilde{L}_z \frac{d\tilde{H}}{d\tilde{L}_z} \right] \frac{d\tilde{L}_z}{\tilde{L}_z}. \tag{3.78}$$

Then the transfer function is

$$\boxed{\begin{aligned} \mathcal{S}_m(\alpha, \alpha') = {} & 2\pi G K(\alpha', m) \tilde{C} R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} \int d\tilde{U} \, \mathcal{J}_0(\tilde{U}) \tilde{U} \\ & \times \left| \frac{\beta\tilde{U}^2 + \beta - 2}{2\beta} \right|^{\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}} \sum_{l=-\infty}^{+\infty} Q_{lm}(\alpha') Q_{lm}^*(\alpha) F_{lm}(\alpha - \alpha'). \end{aligned}} \tag{3.79}$$

where $\tilde{C}$ is given in eq. (2.55). For convenience Q_{lm} and F_{lm} are shown as depending only upon the wavenumber α, although in reality they also depend upon the eccentric velocity $\tilde{U}$ as well as on the disk parameter β. The corresponding result for the Toomre-Zang disk is (Z3.40):

$$\begin{aligned} \mathcal{S}_m(\alpha, \alpha') = {} & \frac{(\gamma+1)^{1+\gamma/2}}{2^{\gamma/2} e^{(\gamma+1)/2} \sqrt{\pi} \Gamma\left[\frac{\gamma+1}{2}\right]} \left(\frac{2}{\gamma+1} \right)^{\gamma/2} K(\alpha', m) \\ & \times \int d\tilde{U} \, \mathcal{J}_0(\tilde{U}) \tilde{U} e^{-\frac{\tilde{U}^2}{2\tilde{\sigma}_u^2}} \sum_{l=-\infty}^{+\infty} Q_{lm}(\alpha') Q_{lm}^*(\alpha) F_{lm}(\alpha - \alpha'). \end{aligned} \tag{3.80}$$

where the velocity dispersion $\tilde{\sigma}_u$ is given by $\tilde{\sigma}_u^{-2} = \gamma + 1$

Can we gain some rather more physical understanding of this expression for $\mathcal{S}_m$? The transfer function $\mathcal{S}_m(\alpha, \alpha')$ describes the contribution of the imposed log-spiral component with wavenumber α' to the response component with wavenumber α. To see how this is calculated, we consider a star orbiting in the disk. The shape of its orbit is characterised by its eccentric velocity $\tilde{U}$. The Fourier coefficient Q_{lm} describes the "match" between a particular log-spiral component and this orbit. Specifically, the changes in the star's energy and angular momentum caused by the log-spiral perturbation are expanded in harmonics of the orbit's radial period. The Fourier coefficient Q_{lm} gives the contribution to the lth component of the perturbation experienced by the star. Thus, in the expression for the transfer function (3.79), the first Fourier coefficient $Q_{lm}(\alpha')$ describes how far the star is forced out of its unperturbed orbit by the imposed perturbation, and hence the star's tendency to stop contributing to the imposed log-spiral. The second Fourier coefficient $Q_{lm}^*(\alpha)$ describes how well matched the perturbed star is to the response log-spiral component with wavenumber α.

Both these depend only on the shape of the star's orbit, not on its size; nor has any consideration yet been made of the "interaction" between the log-spirals. This is accounted for by the angular momentum function F_{lm}. How easy it is for a star to move from the imposed to the response log-spiral depends on the difference between the response and imposed wavenumbers, $\eta = \alpha - \alpha'$, as well as the charactistics of the perturbation (its azimuthal symmetry, growth rate and pattern speed, given by m and $\tilde{\omega}$), and also on the size of the orbit, $\tilde{L}_z$. The integral in the angular momentum adds up similar-shaped orbits of all different sizes. F_{lm} then describes how feasible it is for density to move from wavenumber α' to α in a perturbation with this m and $\tilde{\omega}$.

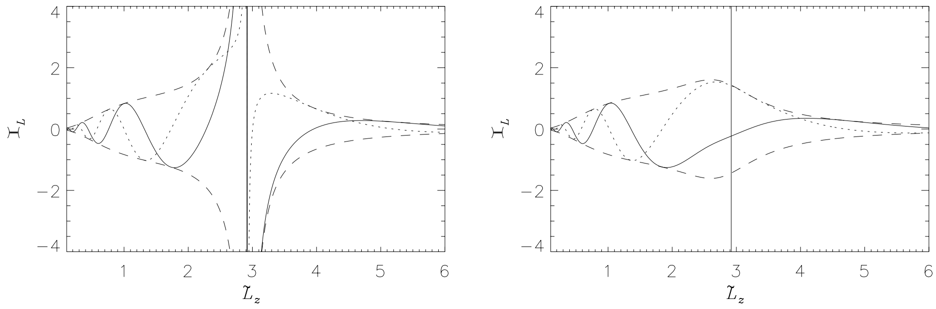
Due to the scale-free nature of the disk, we have been able to deal simultaneously with all orbits of a given shape, irrespective of their size. The $\tilde{U}$-dependent parts of the integrand in $\mathcal{S}_m$ measure how many stars there are with each shape of orbit. These are then added up to determine how much density ultimately moves from the log-spiral component with wavenumber α' to that with wavenumber α.

3.6 The angular momentum function

Remarkably, all the dependence on angular momentum, growth rate and pattern speed is contained within a single integral, the angular momentum function F_{lm} (3.77). Orbits with the same $\tilde{U}$ but different $\tilde{L}_z$ are scale models of each other, so this integral adds up orbits of the same shape at all possible scales. To gain an initial understanding of F_{lm}, we consider the integrand $\mathcal{I}_L(\tilde{L}_z)$, defined as

$$\mathcal{I}_L = \frac{1}{2\pi\tilde{L}_z} \frac{e^{-i\eta\frac{2}{2-\beta}\ln\tilde{L}_z}}{\left(l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega}\tilde{L}_z^{\frac{2+\beta}{2-\beta}}\right)} \left[\left\{(l\tilde{\kappa} + m\tilde{\Omega})\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \gamma m\right\}\tilde{H} - m\tilde{L}_z\frac{d\tilde{H}}{d\tilde{L}_z}\right] \tag{3.81}$$

If the growth rate s is zero and $l\tilde{\kappa} + m\tilde{\Omega}$ has the same sign as $\tilde{\Omega}_\mathrm{p}$, the integrand has a singularity at $\tilde{L}_z = \tilde{L}_\mathrm{sing} \equiv [(l\tilde{\kappa} + m\tilde{\Omega})/(m\tilde{\Omega}_p)]^{\frac{2-\beta}{2+\beta}}$. Even if the growth rate is non-zero, when $(l\tilde{\kappa} + m\tilde{\Omega})/(m\tilde{\Omega}_\mathrm{p})$ is positive the maximum value of $|\mathcal{I}_L|$ is usually close to $\tilde{L}_\mathrm{sing}$, as illustrated in these plots:



s = 0.0 s = 0.3

Figure 3.5: Graph of the integrand $\mathcal{I}_L$ plotted against $\tilde{L}_z$ for two different growth rates: In the left-hand plot, $s = 10^{-7}$; on the right, $s = 0.3$. The solid line is the real part of $\mathcal{I}_L$, and the dotted line the imaginary, for $\eta = -5$. The dashed line shows the envelope $\pm|\mathcal{I}_L|$. The solid vertical line indicates the position of $\tilde{L}_z = \tilde{L}_\mathrm{sing}$. [$\beta = 0.25$, $N = 2$, $M = 6$, $\tilde{L}_\mathrm{c} = 31.6$; $\gamma = 13.7$, $\Omega_\mathrm{p} = 0.5$, $m = 2$; $l = 2$, $\tilde{U} = 0.5$.]

The magnitude of the integrand, $|\mathcal{I}_L|$, is independent of η. Within this envelope, marked with a solid line in fig. 3.5, the real and imaginary parts of $\mathcal{I}_L$ oscillate as a function of $\tilde{L}_z$. The frequency of this oscillation increases with η. For non-zero growth rate, the integrand is everywhere finite. It can then easily be seen from the right-hand plot above that, as $\eta \to \pm\infty$, the integration will add equal positive and negative contributions. We therefore expect the angular momentum function to tend to zero as $\eta \to \pm\infty$. This is what we would expect physically. F_{lm} describes how easy it is to shift density between log-spirals of different wavenumber. Intuitively we would expect this task to become harder – and F_{lm} to decrease – when the two wavenumbers are far apart. Note however that when the growth rate vanishes, the integrand has a singularity. Nearly adjacent values of η will then give vastly different values of $\mathcal{I}_L$, so we cannot be sure that cancellation will occur.

Somewhat surprisingly, the integration over $\tilde{L}_z$ can be performed analytically for the cut-out functions $\tilde{H}(\tilde{L}_z)$ given in section 2.4. The calculation and results are given in detail in Appendix C. In the next few pages, we consider the behaviour of this analytic expression for F_{lm}.

Figs. 3.6 and 3.7 compare F_{lm} for two different growth rates s. In each case, the left-hand plot shows $F_{lm}(\eta)$ for vanishing growth rate, while the right-hand one has $s = 0.3$. Fig. 3.6 is for $\beta = +0.25$, and fig. 3.7 for $\beta = -0.25$.

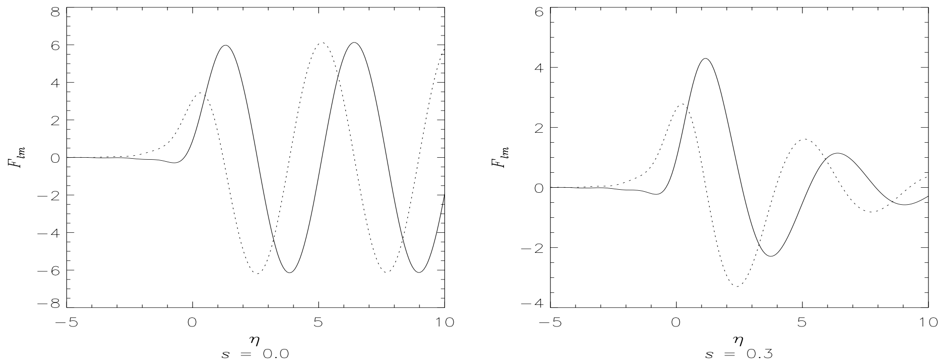


Figure 3.6: Graph of F_{22} plotted against η for two different growth rates in a $\beta = 0.25$ disk. In the left-hand plot, $s = 10^{-7}$; on the right, $s = 0.3$. In each case, the solid line is the real part, and the dotted line the imaginary. [$\beta = 0.25$, $N = 2$, $M = 6$, $\tilde{L}_{\rm c} = 31.6$; $\gamma = 13.7$, $\Omega_{\rm p} = 0.5$, $m = 2$; $l = 2$, $\tilde{U} = 0.5$.]

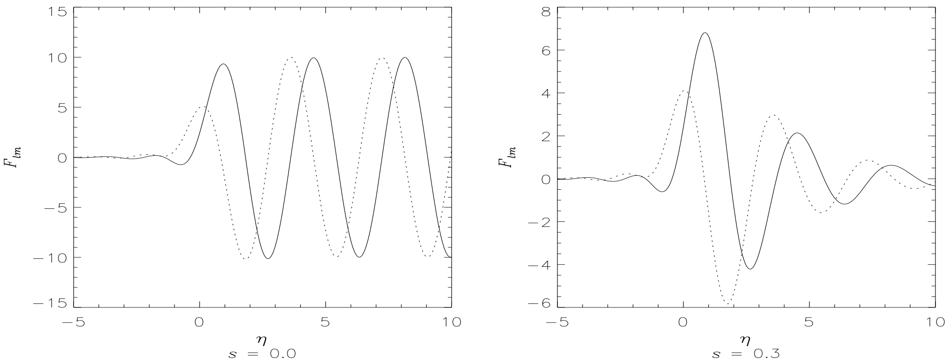


Figure 3.7: Graph of F_{22} plotted against η for two different growth rates in a $\beta = -0.25$ disk. In the left-hand plot, $s = 10^{-7}$; on the right, $s = 0.3$. In each case, the solid line is the real part, and the dotted line the imaginary. [$\beta = -0.25$, $N = 2$, $M = 6$, $\tilde{L}_{\rm c} = 31.6$; $\gamma = 13.7$, $\Omega_{\rm p} = 0.5$, $m = 2$; $l = 2$, $\tilde{U} = 0.5$.]

As expected, for finite growth rate $F_{lm} \to 0$ as $\eta \to \pm\infty$. However, F_{lm} also exhibits a striking asymmetry about η. For positive η, F_{lm} oscillates and its amplitude decays only slowly; for negative β, F_{lm} quickly decays to zero. The asymmetry is most pronounced at zero growth rate; then for positive η F_{lm} is almost purely oscillatory, whereas its amplitude decays rapidly for $\eta < 0$. This turns out to be a characteristic of F_{lm}. The symmetry of F_{lm} varies, but where asymmetry exists, the decay is faster for negative η. The asymmetry of F_{lm} about η means that our disks have a *trailing bias*; this is discussed in more detail in section 3.7. The asymmetry occurs only where $l\tilde{\kappa} + m\tilde{\Omega} > 0$, as we see from fig. 3.8, which compares F_{lm} for positive and negative radial harmonics l. The plot on the left shows $F_{2,2}$, while that on the right shows $F_{-2,2}$.

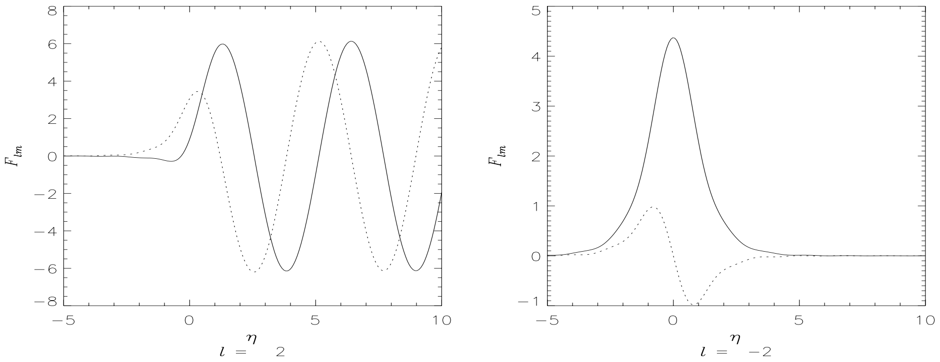


Figure 3.8: Graph of F_{lm} plotted against η for two different radial harmonics l. In the left-hand plot, $l = +2$ and $l\tilde{\kappa} + m\tilde{\Omega} = +4.0$; on the right, $l = -2$ and $l\tilde{\kappa} + m\tilde{\Omega} = -0.6$. In each case, the solid line is the real part, and the dotted line the imaginary. [$\beta = 0.25$, $N = 2$, $M = 6$, $\tilde{L}_{\rm c} = 31.6$; $\gamma = 13.7$, $\Omega_{\rm p} = 0.5$, $s = 10^{-7}$, $m = 2$; $\tilde{U} = 0.5$.]

The symmetry of F_{lm} also depends on the eccentric velocity $\tilde{U}$, as demonstrated in the set of plots in fig. 3.9.

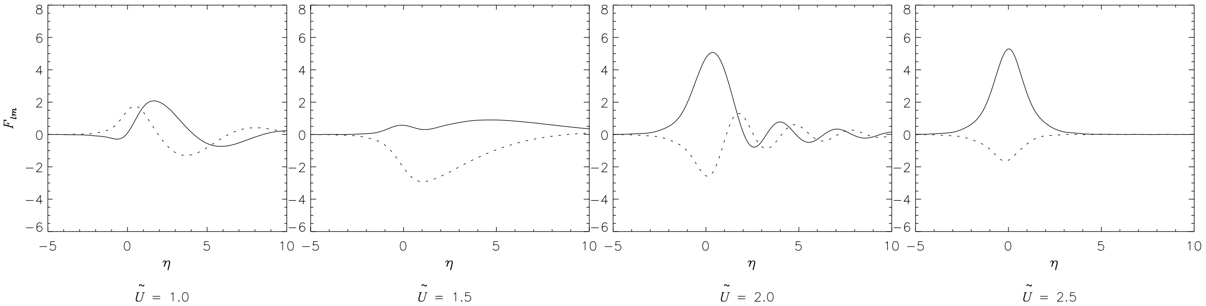
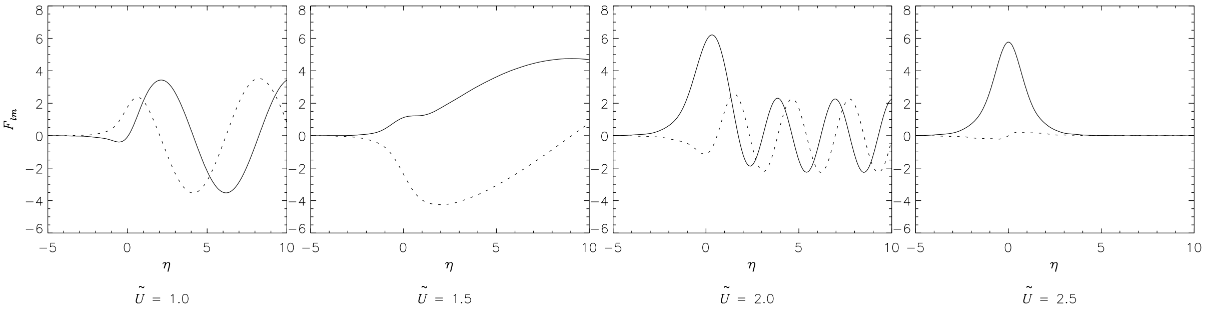


Figure 3.9: Graph of F_{22} plotted against η for four different eccentric velocities $\tilde{U}$, when the growth rate $s = 0.3$. From left to right: $\tilde{U} = 1.0, 1.5, 2.0, 2.5$. In each case, the solid line is the real part, and the dotted line the imaginary. [$\beta = 0.25$, $N = 2$, $M = 6$, $\tilde{L}_{\rm c} = 31.6$; $\gamma = 13.7$, $\Omega_{\rm p} = 0.5$, $s = 0.3$, $m = 2$; $l = 2$.]

We see that, as $\tilde{U}$ increases, the maximum value of F_{lm} increases and moves to $\eta = 0$; F_{lm} becomes more symmetric about $\eta = 0$. As fig. 3.10 shows, this effect persists even at $s \to 0$, when for low values of $\tilde{U}$ the asymmetry is most pronounced.



$\tilde{U} = 1.0$ $\tilde{U} = 1.5$ $\tilde{U} = 2.0$ $\tilde{U} = 2.5$

Figure 3.10: Graph of F_{22} plotted against η for four different eccentric velocities $\tilde{U}$, when the growth rate $s = 10^{-7}$. From left to right: $\tilde{U} = 1.0, 1.5, 2.0, 2.5$. In each case, the solid line is the real part, and the dotted line the imaginary. [$\beta = 0.25$, $N = 2$, $M = 6$, $\tilde{L}_{\rm c} = 31.6$; $\gamma = 13.7$, $\Omega_{\rm p} = 0.5$, $s = 10^{-7}$, $m = 2$; $l = 2$.]

3.7 The trailing bias

For negative values of $l\tilde{\kappa} + m\tilde{\Omega}$ or for high eccentric velocities, F_{lm} decays rapidly as $|\eta|$ moves away from zero; the decay is similar in either direction. However, where $l\tilde{\kappa} + m\tilde{\Omega}$ is positive and the eccentric velocity is low, the decay for $\eta > 0$ is greatly attenuated.

The mathematical origin of this asymmetry is seen most simply in the expression for F_{lm} in the case of a scale-free disk (C.16). This depends on η as

$$F_{lm}(\eta) \propto \frac{e^{-i\hat{\eta}\ln\frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}}}{1 - e^{2\pi\hat{\eta}}}, \tag{3.82}$$

where $\hat{\eta}$ is defined as $2\eta/(2+\beta)$ (see Appendix C, (C.10)).

If the growth rate is zero, then when we take the principal value of the logarithm as in (C.43) we obtain

$$F_{lm}(\eta) \propto e^{-i\hat{\eta}\ln\left|\frac{l\tilde{\kappa}+m\tilde{\Omega}}{m\Omega_{\rm p}}\right|}\frac{e^{2\pi\hat{\eta}}}{1 - e^{2\pi\hat{\eta}}}, \qquad l\tilde{\kappa} + m\tilde{\Omega} > 0; \tag{3.83}$$

$$F_{lm}(\eta) \propto e^{-i\hat{\eta}\ln\left|\frac{l\tilde{\kappa}+m\tilde{\Omega}}{m\Omega_{\rm p}}\right|}\frac{e^{\pi\hat{\eta}}}{1 - e^{2\pi\hat{\eta}}}, \qquad l\tilde{\kappa} + m\tilde{\Omega} < 0. \tag{3.84}$$

Considering extreme values of η on either side of zero, we see immediately that for $l\tilde{\kappa} + m\tilde{\Omega} > 0$ the angular momentum function will be highly asymmetric in η. The magnitude of F_{lm} will tend to a constant value for large positive η, whereas it will decay rapidly as $\exp(-2\pi|\hat{\eta}|)$ for negative η. Conversely, for $l\tilde{\kappa} + m\tilde{\Omega} < 0$ no such asymmetry is apparent; F_{lm} decays as $\exp(-\pi|\hat{\eta}|)$ in either direction.

To gain an insight into how the asymmetry arises when $l\tilde{\kappa} + m\tilde{\Omega} > 0$, we consider the complex part of the integrand of F_{lm} (see Appendix C, eq. (C.7))

$$\mathcal{I}_{\rm c}(h) = \frac{e^{-i\hat{\eta}h}}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega}e^{h}}. \tag{3.85}$$

This function is shown in the following figure.

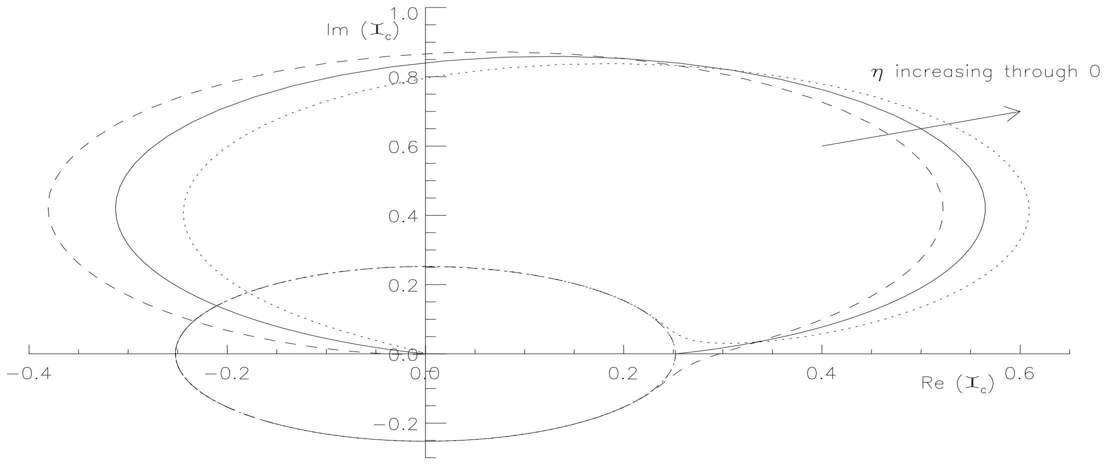


Figure 3.11: The complex part of the integrand of the angular momentum function, $\mathcal{I}_{\rm c}$, is plotted for three different values of $\hat{\eta}$: 0.0 (solid line), +0.1 (dotted line), -0.1 (dashed line). See text for explanation. $\beta = 0.25$, $\Omega_{\rm p} = 0.5$, $s = 0.3$, $l = 2$, $m = 2$, $\tilde{U} = 0.5$.]

We can understand the shape of these curves as follows. For $\eta = 0$ (the solid line in the above figure), the complex nature of the integrand comes purely from $\tilde{\omega}$, the imaginary part of which is the growth rate s. We track the behaviour of $\mathcal{I}_{\rm c}(h)$ from $h = -\infty$ to $+\infty$. At the start, $\mathcal{I}_{\rm c}$ lies on the real axis at $(l\tilde{\kappa} + m\tilde{\Omega})^{-1}$. We take this to be positive, since the trailing bias does not exist when $l\tilde{\kappa} + m\tilde{\Omega} < 0$. As h increases, $\mathcal{I}_{\rm c}$ acquires a positive imaginary component, $s\exp(h)$, and moves round the upper complex plane, ending up at the origin as $h \to \infty$. It is clear that when $\mathcal{I}_{\rm c}$ is integrated to obtain F_{lm}, all the contributions to the imaginary part are positive, so we expect $\mathrm{Im}(F_{lm})$ to be positive at $\eta = 0$. As the previous plots of the behaviour of F_{lm} show, this is indeed the case. The real part of $\mathcal{I}_{\rm c}$ contains both positive and negative contributions, so some cancellation occurs. However, the asymmetry in the position of the curve about the imaginary axis means that we expect the positive contributions to exceed the negative; $\mathrm{Re}(F_{lm})$ is thus typically positive at $\eta = 0$.

The dotted and dashed curves show the behaviour for $\hat{\eta} = \pm 0.1$. Now for extreme negative values of h, $\mathcal{I}_{\rm c} \approx \exp(-i\eta h)$. The curve thus begins with a loop about the origin, traversed in the clockwise direction for positive η, and anticlockwise when η is negative. As h increases, the curve again moves anticlockwise around the upper half-plane, ending up at the origin. However, the presence of the term $\exp(-i\eta h)$ now tilts the curve clockwise, for positive η, and anticlockwise, for negative η, relative to the curve for $\eta = 0$. For positive η, therefore, the dominance of positive over negative contributions to F_{lm} is enhanced, and F_{lm} is increased. Conversely, for negative η, more cancellation occurs, and F_{lm} is reduced. This is the basis of the trailing bias.

We can also use this approach to understand why $F_{lm}(\eta = 0)$ tends to increase with $\tilde{U}$, as seen in fig. 3.9. As $\tilde{U}$ increases, $\tilde{\kappa}$ and $\tilde{\Omega}$ approach zero, so the $\eta = 0$ curve touches the real axis at progressively larger values $(l\tilde{\kappa} + m\tilde{\Omega})^{-1}$. This means that the asymmetry of the $\eta = 0$ curve about the imaginary axis is increased, and so $F_{lm}(\eta = 0)$ is larger.

So for positive values of $l\tilde{\kappa} + m\tilde{\Omega}$ and for low eccentric velocities, F_{lm} decays more rapidly for negative

η; otherwise it is reasonably symmetric about $\eta = 0$. Thus on average, the magnitude of F_{lm} is likely to be greater for positive η ($\alpha > \alpha'$) than for negative. This in turn means that, on average, the magnitude of $\mathcal{S}_m(\alpha, \alpha')$ is greater for $\alpha > \alpha'$, as exemplified in fig. 3.12.

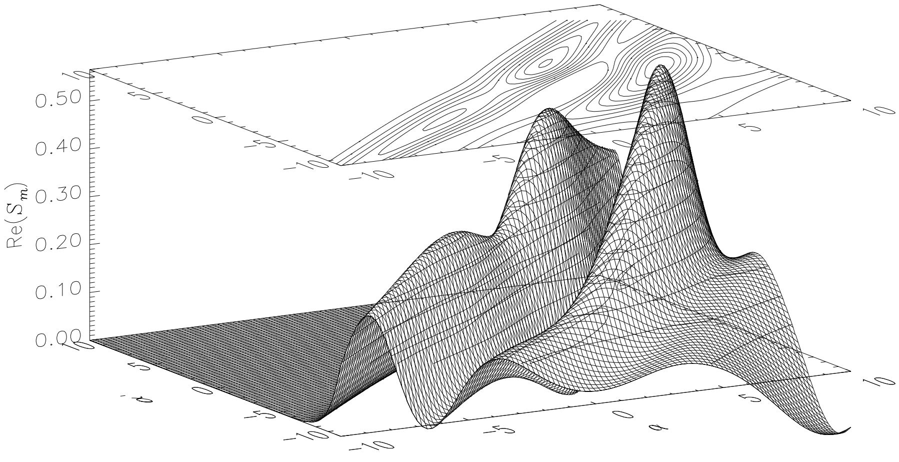


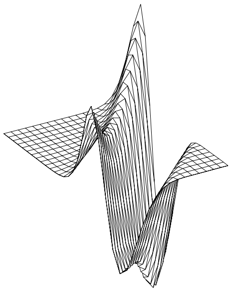
Figure 3.12: Plot of Re($\mathcal{S}_m$) plotted against α and α'. The superposed contour plot shows ten equally-spaced contours. [$\beta = 0.25$, $N = 2$, $M = 6$, $\tilde{L}_c = 31.6$; $\gamma = 13.7$, $\Omega_p = 0.5$, $s = 10^{-7}$, $m = 2$. [Numerical accuracy parameters: $n = 101$, $\Delta\alpha = 0.2$, $l_{\min} = -20$, $l_{\max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2.0$.]

$\mathcal{S}_m(\alpha, \alpha')$ tells us how much the imposed component with wavenumber α' contributes to the response component with wavenumber α. Consider a particular component of the response, which has wavenumber α. This is made up of contributions from the full range of wavenumbers α' in the imposed pattern. But since $\mathcal{S}_m(\alpha, \alpha')$ is systematically larger for $\alpha > \alpha'$ than for $\alpha < \alpha'$, most of the response at α is due to imposed components with wavenumber *less* than α. An alternative way of viewing the situation is that imposed components mostly go to make up response components with *larger* wavenumber than their own. Thus the asymmetry in the transfer function means that our disk tries to make whatever pattern is imposed on it more like a trailing spiral. Tightly-wound leading spirals are unwrapped and made more loosely-wound; loosely-wound leading spirals are transformed into loosely-wound trailing spirals; loosely-wound trailing spirals are wrapped up into tightly-wound trailing spirals.

3.8 Summary

In this Chapter, we have derived the integral equation which will form the basis for our stability analysis. We have expressed the perturbation as an integral over logarithmic spiral components, and have then further expanded each component in a Fourier series. In the course of this work, we have introduced many new quantities which will be used throughout the thesis. These are collected for reference in Appendix A.

Chapter 4
The self-consistent disk



4.1 Introduction

In this chapter, we examine the self-consistent disk, where there are no cut-out functions removing part of the density, and the potential experienced by stars in the disk is entirely due to their own gravity. The self-consistent disk is perfectly scale-free. This leads to some rather strange effects. As originally pointed out by Kalnajs in 1974 (see Zang 1976), all dependence on growth rate and pattern speed can now be factored out of the integral equation. This means that the stability cannot depend on the growth rate and pattern speed. The same detail was noted independently by Lynden-Bell & Lemos (1996).

The behaviour of the scale-free disk must be very different from that of a cut-out disk. In general, we expect disks to be stable when they are sufficiently hot, and to become unstable at some critical temperature as the disk is gradually cooled. In a disk possessing a length-scale, modes have a characteristic growth rate and pattern speed. At the critical temperature, the mode is marginally stable, i.e. its growth rate is zero. As the disk is cooled below the critical temperature, modes with positive growth rates become possible. In contrast, the symmetry of the scale-free disk forbids a preference for one growth rate over another. At a given temperature, the disk must admit modes with all frequencies or none. Thus there are three options: either there is a *continuum of modes* with all possible pattern speeds and growth rates, or there are *neutral modes*, with growth rate and pattern speed both zero, or there are *no modes* at all.

We derive the response function describing whether neutral modes are possible. Solutions of this for the various azimuthal harmonics m will be presented in subsequent chapters. In this chapter, we investigate the possibility that the integral equation admits a continuum of solutions, and present some plausibility arguments as to why this may not be possible for $m \neq 0$. Our final conclusions on this problem are postponed to later in the thesis (Chapter 10).

4.2 The symmetry of the scale-free disk

The distinction between the isolated neutral modes ($\omega = 0$) and the continuum of growing modes (all $\omega \neq 0$) is built into our integral equation. For the scale-free disk, the angular momentum function is (3.77)

$$F_{lm}(\eta) = \left\{ (l\tilde{\kappa} + m\tilde{\Omega}) \left| \frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2} \right| - \gamma m \right\} \frac{1}{2\pi} \int_0^\infty \frac{e^{-i\eta\frac{2}{2-\beta}\ln\tilde{L}_z}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}\tilde{L}_z^{\frac{2+\beta}{2-\beta}}} \frac{d\tilde{L}_z}{\tilde{L}_z}. \tag{4.1}$$

For finite values of ω, the angular momentum function consists of a regular function for $\eta \neq 0$, and a Dirac delta-function about $\eta = 0$ (Appendix C). We write

$$F_{lm}(\eta) = F_{lm}^{(1)}(\eta) + F_{lm}^{(2)}\delta(\eta) \tag{4.2}$$

where (C.29)

$$F_{lm}^{(1)}(\eta) = -\frac{2-\beta}{2+\beta} \left\{ \left| \frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2} \right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}} \right\} \frac{ie^{-i\hat{\eta}\ln\frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}}}{1-e^{2\pi\hat{\eta}}}, \tag{4.3}$$

$$F_{lm}^{(2)} = \frac{2-\beta}{4} \left\{ \left| \frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2} \right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}} \right\}, \tag{4.4}$$

where $\hat{\eta} = 2\eta/(2+\beta)$. (For $m \neq 0$, these expressions hold for all l. For $m = 0$, there is an ambiguity at the $l = 0$ harmonic. We shall see that the correct choices are then $F_{00}^{(1)} = F_{00}^{(2)} = 0$.)

When $\omega = 0$, the expression (4.1) becomes

$$F_{lm}(\eta) = \left\{ \left| \frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2} \right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}} \right\} \frac{1}{2\pi} \int_0^\infty e^{-i\eta\frac{2}{2-\beta}\ln\tilde{L}_z} \frac{d\tilde{L}_z}{\tilde{L}_z}. \tag{4.5}$$

Hence the angular momentum function is simply a delta-function in η:

$$F_{lm}(\eta) = \frac{2-\beta}{2} \left\{ \left| \frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2} \right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}} \right\} \delta(\eta). \tag{4.6}$$

(Again, for $m = 0$, $F_{00}(\eta) = 0$.) For $\omega = 0$, we do *not* pick up the "off-diagonal" terms contained in $F_{lm}^{(1)}(\eta)$ in eq. (4.2). These are terms with $\eta \neq 0$, describing a response at wavelengths different from that of the original perturbation. Thus the response to a neutral log-spiral must be a pure log-spiral at the same wavelength. There is no coupling mechanism for exciting responses at different wavelengths. We note that $F_{lm}^{(1)}(\eta)$ has no unique limiting value as $\omega \to 0$; specifically, it does not tend to zero in this limit. There is a discontinuity: the case $\omega = 0$ cannot be recovered by taking the limit $\omega \to 0$. Physically, this is a consequence of the symmetry of the disk. A pure log-spiral is self-similar (Lynden-Bell & Lemos 1996), and a neutral log-spiral introduces no length- or time-scales through its growth rate and pattern speed.

Thus the response must be purely self-similar, i.e., a pure log-spiral. Once the log-spiral is allowed to grow or rotate – however slowly – a length- and time-scale has been introduced. The response is no longer self-similar, but involves contributions at many wavelengths. We discuss the neutral modes, with $\omega = 0$, in section 4.3. We now proceed to show that any modes with $\omega \neq 0$ must form a *continuum*. In section 4.4 we investigate these growing modes in more detail, and discuss whether such modes are in fact possible.

As we did for the angular momentum function in (4.2), we divide the transfer function into a regular part and a Dirac delta-function:

$$\mathcal{S}_m(\alpha, \alpha') = \mathcal{S}_m^{(1)}(\alpha, \alpha') + \delta(\alpha - \alpha')\mathcal{S}_m^{(2)}(\alpha, \alpha') \tag{4.7}$$

$\mathcal{S}_m^{(1)}(\alpha, \alpha')$ and $\mathcal{S}_m^{(2)}(\alpha, \alpha')$ may be obtained from (3.79), using $F_{lm}^{(1)}$ and $F_{lm}^{(2)}$ respectively in place of F_{lm}. From eq. (4.3), we see that $F_{lm}^{(1)}(\eta)$ admits the factorisation

$$F_{lm}^{(1)}(\eta) = e^{i\hat{\eta}\ln\tilde{\omega}}\hat{F}_{lm}^{(1)}(\eta), \tag{4.8}$$

where $\hat{F}_{lm}^{(1)}(\eta)$ is *independent* of the growth rate and pattern speed. Since the transfer function depends on the growth rate only through the angular momentum function, $\mathcal{S}_m^{(2)}(\alpha, \alpha')$ is independent of $\tilde{\omega}$, while the dependence of $\mathcal{S}_m^{(1)}(\alpha, \alpha')$ on $\tilde{\omega}$ can be factorised out as

$$\mathcal{S}_m^{(1)}(\alpha, \alpha') = e^{i(\hat{\alpha}-\hat{\alpha}')\ln\tilde{\omega}}\hat{\mathcal{S}}_m^{(1)}(\alpha, \alpha'), \tag{4.9}$$

where $\hat{\alpha}$ is defined analogously to $\hat{\eta}$, i.e. $\hat{\alpha} = 2\alpha/(2+\beta)$. $\hat{\mathcal{S}}_m^{(1)}(\alpha, \alpha')$ is independent of $\tilde{\omega}$. We write the response transform $A_{\text{res}}(\alpha)$ as a complex multiple of the imposed transform $A(\alpha)$, so that $A_{\text{res}}(\alpha) = \lambda A(\alpha)$ (a step explained in more detail in section 5.5). We then define

$$\hat{A}(\alpha) = e^{-i\hat{\alpha}\ln\tilde{\omega}}A(\alpha). \tag{4.10}$$

Then the integral equation (3.45) becomes

$$\boxed{\lambda\hat{A}(\alpha) = \hat{A}(\alpha)\mathcal{S}_m^{(2)}(\alpha, \alpha) + \int_{-\infty}^{+\infty} d\alpha'\hat{A}(\alpha')\hat{\mathcal{S}}_m^{(1)}(\alpha, \alpha'), \qquad (4.11)}$$

where none of the quantities depend on $\tilde{\omega}$. This means that if a non-trivial solution $\hat{A}(\alpha)$ exists at all, we can build self-consistent modes $A(\alpha)$ with *any* growth rate and pattern speed. This is not a consequence of the particular distribution function employed; rather, it is a necessary property of a self-similar disk, which possesses no characteristic length- or time-scale, and hence cannot distinguish between different growth rates and pattern speeds.

Why does the reference radius R_0 not provide a length-scale? We have introduced the reference radius R_0, reference density Σ_0 and reference velocity v_β, related by (2.4). It thus appears as if we have two degrees of freedom; for example, we could choose Σ_0 and R_0 (v_β then being defined through (2.4)). If this were the case, R_0 would provide a length-scale, and R_0/v_β a time-scale. However, one of these "degrees of freedom" is illusory. The surface density is $\Sigma = \Sigma_0 R_0^{1+\beta} R^{-1-\beta}$, so we cannot tell the difference between disks with the same $\Sigma_0 R_0^{1+\beta}$. We are therefore free only to choose $\Sigma_0 R_0^{1+\beta}$, which then determines $v_\beta^2 R_0^\beta$ through (2.4). The disk is therefore characterised by only two quantities, one with dimensions $[M][L]^{\beta-1}$, and the other with $[L]^{4+\beta}[T]^{-2}$. It is apparent that these cannot be combined to yield a quantity with dimensions of length or of time. This means that either there is no "special" frequency, or that that frequency is zero (Lynden-Bell & Lemos 1996).

When we carve out the centre of the disk, we break the self-similarity. The radius $R = R_0$ at which the inner cut-out is applied defines a length-scale. A time-scale is then provided by the period of a circular orbit at this radius. The integral equation (3.45) then depends on the frequency $\tilde{\omega}$. We are free to adjust $\tilde{\omega}$ so as to obtain non-trivial solutions. The self-similarity of the scale-free disk removes this freedom. The only parameter we can adjust is the temperature. We expect that when the disk is hot, no modes exist. As we decrease the velocity dispersion below a critical value, we may find that modes become possible. The symmetry of the disk prevents it from selecting modes with a special frequency, unless that frequency is zero. Thus the disk may admit isolated neutral modes; equally it may admit a continuum of modes with all possible frequencies. Both types of modes are in principle possible.

We should stress that the neutral modes are not related to the growing modes. There is no reason to suppose that the two should set in at the same temperature. Further, the isolated modes necessarily have a different morphology to the growing modes. As we have seen, the neutral modes are pure log-spirals. We show in section 4.4 that the growing modes *must* contain contributions at many different wavelengths. Thus the growing mode with $\tilde{s} = \tilde{\Omega}_\mathrm{p} = 10^{-50}$ has a completely different form from the neutral mode with $\tilde{s} = \tilde{\Omega}_\mathrm{p} = 0$. We emphasise this point because the situation is counter-intuitive. It is tempting to try and derive the neutral modes from the continuum of growing modes, or draw conclusions about the growing modes by making small adjustments to the neutral modes. This is not possible.

4.3 Neutral modes

We have seen that for neutral modes, the angular momentum function is simply a delta-function in η (4.6). Thus the transfer function (3.79) is also proportional to a delta-function. We write $\mathcal{S}_m(\alpha, \alpha') = \mathcal{R}\delta(\alpha - \alpha')$,where $\mathcal{R}$ is the *response function*:

$$\begin{aligned} \mathcal{R} = {} & \left(1 - \frac{\beta}{2}\right) 2\pi G K(\alpha, m) \tilde{C} R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} \int d\tilde{U}\, \mathcal{J}_0(\tilde{U})\tilde{U} \\ & \times \left| \frac{\beta\tilde{U}^2 + \beta - 2}{2\beta} \right|^{\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}} \sum_{l=-\infty}^{+\infty} |Q_{lm}(\alpha)|^2 \left\{ \left| \frac{2 + 2\gamma - \beta\gamma}{\beta\tilde{U}^2 + \beta - 2} \right| - \frac{\gamma m}{l\tilde{\kappa} + m\tilde{\Omega}} \right\} \end{aligned} \tag{4.12}$$

We note that $\mathcal{R}(\alpha) = 2\mathcal{S}_m^{(2)}(\alpha,\alpha)$. This is because the angular momentum function $F_{lm}(\eta)$ for neutral modes is twice the delta-function term $F_{lm}(\eta)^{(2)}\delta(\eta)$.

Due to the delta-function in $\mathcal{S}_m$, the response density transform $A_{\rm res}(\alpha)$ at any wavenumber α is simply proportional to the imposed density transform $A_{\rm imp}(\alpha)$ at that wavenumber, so there is no need to build a matrix. The ratio of the response density to the imposed perturbation is given by the response function $\mathcal{R}$. If $\mathcal{R} > 1$, the perturbation grows; if $\mathcal{R} < 1$, the perturbation dies away. Self-consistent modes require $\mathcal{R} = 1$. We note that $\mathcal{R}$ is real and independent of the radial position in the disk, so that the phase of the perturbation is unchanged, and the disk is either stable or unstable everywhere. We recall the symmetry property of the Kalnajs function ($K(\alpha,m) = K(-\alpha,m)$, eq. (3.34)) and of the Fourier coefficients ($Q_{lm}(\alpha) = Q_{lm}^*(-\alpha)$, eq. (3.55)). Thus the response function is independent of the sign of the wavenumber. This is the expression of the anti-spiral theorem (Lynden-Bell & Ostriker 1967) for the scale-free disk. Every spiral of wavenumber α comes with an "anti-spiral" of wavenumber $-\alpha$. Hence any neutral modes will come in pairs, one leading and one trailing.

It is instructive to consider an alternative derivation of the response function (4.12). This is especially useful for the case $m = 0$, since it resolves the ambiguity at the $l = 0$ harmonic. In this derivation, we begin by considering the response to a growing perturbation. We also *assume* that the response to a pure log-spiral is itself a pure log-spiral. This ensures that we recover the neutral modes when we finally let $\omega \to 0$.

Putting together eqs. (3.17), (3.49), (3.61) and (3.62), we obtain

$$\begin{aligned} f_{\rm imp}(t) = {}& \tilde{C} L_z^{\gamma-1} |E|^{1/\beta+\gamma/\beta-\gamma/2} \psi^{\alpha m}_{\rm imp}(t) \\ & \times \left\{ \left|\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}\right| \frac{L_z}{|E|} + i\left(\left|\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}\right| \frac{L_z}{|E|}\omega - m\gamma \right) \int_{-\infty}^{t} \frac{\psi^{\alpha m}_{\rm imp}(t')}{\psi^{\alpha m}_{\rm imp}(t)} dt' \right\} \end{aligned} \tag{4.13}$$

Substituting eq. (3.36) for $\psi^{\alpha m}_{\rm imp}$, we obtain

$$\begin{aligned} f_{\rm imp}(t) = {}& 2\pi G \tilde{C} \Sigma_{\rm p} K(\alpha,m) R_0 L_z^{\gamma-1} |E|^{1/\beta+\gamma/\beta-\gamma/2} e^{i(m\theta-\omega t)} \left(\frac{R}{R_0}\right)^{i\alpha-\frac{1}{2}} \left\{ \left|\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}\right| \frac{L_z}{|E|} \right. \\ & \left. + i\left(\left|\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}\right| \frac{L_z}{|E|}\omega - m\gamma \right) \int_{-\infty}^{t} e^{i\left(m(\theta'-\theta)-\omega(t'-t)\right)} \left(\frac{R'}{R}\right)^{i\alpha-\frac{1}{2}} dt' \right\} \end{aligned} \tag{4.14}$$

The integral in this equation can be split up into an infinite sequence of integrals over the radial period T. Successive integrals differ by $\exp(-iT(m\Omega - \omega))$.

$$\begin{aligned} f_{\rm imp}(t) = {}& 2\pi G \tilde{C} \Sigma_{\rm p} K(\alpha,m) R_0 L_z^{\gamma-1} |E|^{1/\beta+\gamma/\beta-\gamma/2} e^{i(m\theta-\omega t)} \left(\frac{R}{R_0}\right)^{i\alpha-\frac{1}{2}} \left\{ \left|\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}\right| \frac{L_z}{|E|} \right. \\ & \left. + i\left(\left|\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}\right| \frac{L_z}{|E|}\omega - m\gamma \right) \frac{1}{1 - e^{-iT(m\Omega-\omega)}} \int_{t-T}^{t} e^{i\left(m(\theta'-\theta)-\omega(t'-t)\right)} \left(\frac{R'}{R}\right)^{i\alpha-\frac{1}{2}} dt' \right\} \end{aligned} \tag{4.15}$$

For $m \neq 0$, we simply set $\omega = 0$ to obtain

$$\begin{aligned} f_{\rm imp}(t) = {}& 2\pi G \tilde{C} \Sigma_{\rm p} K(\alpha,m) R_0 L_z^{\gamma-1} |E|^{1/\beta+\gamma/\beta-\gamma/2} e^{im\theta} \left(\frac{R}{R_0}\right)^{i\alpha-\frac{1}{2}} \\ & \times \left\{ \left|\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}\right| \frac{L_z}{|E|} - \frac{im\gamma}{1 - e^{-im\Omega T}} \int_{t-T}^{t} e^{im(\theta'-\theta)} \left(\frac{R'}{R}\right)^{i\alpha-\frac{1}{2}} dt' \right\}. \end{aligned} \tag{4.16}$$

This expression can then be integrated over velocity space to obtain the response density Σ_{res}.

$$\begin{aligned} \Sigma_{\text{res}} &= 2\pi G\tilde{C}\Sigma_{\text{p}}K(\alpha,m)R_0 e^{im\theta}\left(\frac{R}{R_0}\right)^{i\alpha-\frac{1}{2}}\iint du\,dv\,L_z^{\gamma-1}|E|^{1/\beta+\gamma/\beta-\gamma/2} \\ &\times\left\{\left|\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}\right|\frac{L_z}{|E|}-\frac{im\gamma}{1-e^{-im\Omega T}}\int_{t-T}^{t}e^{im(\theta'-\theta)}\left(\frac{R'}{R}\right)^{i\alpha-\frac{1}{2}}dt'\right\}. \end{aligned} \tag{4.17}$$

For self-consistent solutions, we require this response density Σ_{res} to be equal to the imposed density Σ_{imp}. Since we are considering the response to a single log-spiral component (3.25) with zero growth rate, the imposed density is (3.25)

$$\Sigma_{\text{imp}} = \Sigma_{\text{p}}e^{im\theta}\left(\frac{R_0}{R}\right)^{3/2-i\alpha}.$$

As in the previous section, we define the response function $\mathcal{R} = \Sigma_{\text{res}}/\Sigma_{\text{imp}}$. This is then

$$\begin{aligned} \mathcal{R} &= 2\pi G\tilde{C}K(\alpha,m)R\iint du\,dv\,L_z^{\gamma-1}|E|^{1/\beta+\gamma/\beta-\gamma/2} \\ &\times\left\{\left|\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}\right|\frac{L_z}{|E|}-\frac{im\gamma}{1-e^{-im\Omega T}}\int_{t-T}^{t}e^{im(\theta'-\theta)}\left(\frac{R'}{R}\right)^{i\alpha-\frac{1}{2}}dt'\right\}. \end{aligned} \tag{4.18}$$

This expression is equivalent to that obtained directly from the integral equation, eq. (4.12). As a preliminary to proving this, we first transform the integration variables in eq. (4.18), performing the integration over $\tilde{U}$ and $\tilde{R}$ instead of u and v. The Jacobian of the transformation is given by combining

$$du\,dv = dU\,dR_H\left(1-\frac{\beta}{2}\right)\frac{v_\beta}{R}\left(\frac{R_0}{R_H}\right)^{\beta/2}\frac{\tilde{U}}{\sqrt{\tilde{U}^2+1-\tilde{R}^{-2}+\frac{2}{\beta}(\tilde{R}^{-\beta}-1)}}, \tag{4.19}$$

and

$$dU\,dR_H = d\tilde{U}\,d\tilde{R}\,v_\beta R\tilde{R}_{\text{H}}^{-\beta/2}\tilde{R}^{-2} = d\tilde{U}\,d\tilde{R}\,v_\beta R_0\tilde{R}_{\text{H}}^{1-\beta/2}\tilde{R}^{-1}. \tag{4.20}$$

We also substitute for energy and angular momentum in terms of eccentric velocity and home radius (eqs. (2.19) and (2.24)). The response function (4.18) then becomes

$$\begin{aligned} \mathcal{R} &= \left(1-\frac{\beta}{2}\right)2\pi GK(\alpha,m)\tilde{C}R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)}\iint d\tilde{U}\,\tilde{U}\frac{d\tilde{R}}{\tilde{R}}\frac{1}{\sqrt{\tilde{U}^2+1-\tilde{R}^{-2}+\frac{2}{\beta}(\tilde{R}^{-\beta}-1)}} \\ &\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\frac{im\gamma v_\beta}{R_0\tilde{R}_{\text{H}}^{1+\beta/2}}\frac{1}{1-e^{-im\Omega T}}\int_{t-T}^{t}e^{im(\theta'-\theta)}\left(\frac{R'}{R}\right)^{i\alpha-\frac{1}{2}}dt'\right\}, \end{aligned} \tag{4.21}$$

where, as in section 3.5, the eccentric velocity is integrated from zero to infinity for negative β; for positive β the upper limit is $(2/\beta-1)^{1/2}$. $\tilde{R}$ is integrated from $\tilde{R}_{\text{min}}$ to $\tilde{R}_{\text{max}}$.

We now use the relation

$$\frac{d\tilde{R}}{d\chi} = \frac{\tilde{u}\mathcal{J}_0(\tilde{U})}{2\pi} = \frac{\mathcal{J}_0(\tilde{U})}{2\pi}\sqrt{\tilde{U}^2+1-\tilde{R}^{-2}+\frac{2}{\beta}\left(\tilde{R}^{-\beta}-1\right)} \tag{4.22}$$

so that

$$\begin{aligned}\mathcal{R} = \left(1 - \frac{\beta}{2}\right) 2\pi G K(\alpha, m) \tilde{C} R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} \int d\tilde{U}\; \mathcal{J}_0(\tilde{U}) \tilde{U} \left|\frac{\beta \tilde{U}^2 + \beta - 2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}} \\ \times \int \frac{1}{2\pi} \frac{d\chi}{\tilde{R}} \left\{ \left|\frac{2 + 2\gamma - \beta\gamma}{\beta\tilde{U}^2 + \beta - 2}\right| - \frac{i m \gamma v_\beta}{R_0 \tilde{R}_{\mathrm{H}}^{1+\beta/2}} \frac{1}{1 - e^{-im\Omega T}} \int_{t-T}^{t} e^{im(\theta' - \theta)} \left(\frac{R'}{R}\right)^{i\alpha - \frac{1}{2}} dt' \right\}.\end{aligned} \tag{4.23}$$

Eq. (4.23) holds for $m \neq 0$. When the imposed perturbation is axisymmetric, eq. (4.15) becomes.

$$\begin{aligned}f_{\mathrm{imp}}(t) = 2\pi G K(\alpha, 0) \tilde{C} \Sigma_{\mathrm{p}} R_0 e^{st} \left(\frac{R}{R_0}\right)^{i\alpha - \frac{1}{2}} L_z^{\gamma-1} |E|^{1/\beta + \gamma/\beta - \gamma/2} \\ \times \left\{ \left|\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}\right| \frac{L_z}{|E|} - \left|\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}\right| \frac{L_z}{|E|} \frac{s}{1 - e^{-sT}} \int_{t-T}^{t} e^{s(t'-t)} \left(\frac{R'}{R}\right)^{i\alpha - \frac{1}{2}} dt' \right\}\end{aligned} \tag{4.24}$$

We must then use l'Hôpital's rule to take the limit as $s \to 0$.

$$\begin{aligned}f_{\mathrm{imp}}(t) = 2\pi G K(\alpha, 0) \tilde{C} \Sigma_{\mathrm{p}} R_0 \left|\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}\right| \left(\frac{R}{R_0}\right)^{i\alpha - \frac{1}{2}} \\ \times L_z^{\gamma} |E|^{1/\beta + \gamma/\beta - \gamma/2 - 1} \left\{ 1 - \frac{1}{T} \int_{t-T}^{t} \left(\frac{R'}{R}\right)^{i\alpha - \frac{1}{2}} dt' \right\}.\end{aligned} \tag{4.25}$$

This can be integrated over velocity to obtain the $m = 0$ response function

$$\begin{aligned}\mathcal{R} = 2\pi G K(\alpha, 0) \tilde{C} \left|\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}\right| R \\ \times \iint du\, dv\, L_z^{\gamma} |E|^{1/\beta + \gamma/\beta - \gamma/2 - 1} \left\{ 1 - \frac{1}{T} \int_{t-T}^{t} \left(\frac{R'}{R}\right)^{i\alpha - \frac{1}{2}} dt' \right\}.\end{aligned} \tag{4.26}$$

For the Toomre-Zang disk, the axisymmetric response function becomes

$$\mathcal{R} = 2\pi G K(\alpha, 0) \tilde{C} \frac{R^{\gamma+1}}{\sigma_u^2} \iint du\, dv\, v^{\gamma}\, e^{-(u^2+v^2)/(2\sigma_u^2)} \left\{ 1 - \frac{1}{T} \int_{t-T}^{t} \left(\frac{R'}{R}\right)^{i\alpha - \frac{1}{2}} dt' \right\}, \tag{4.27}$$

where $\sigma_u^2 = v_0^2/(\gamma + 1)$, and $\tilde{C}$ is a function of γ given by (2.61). This $\beta = 0$ limit was clearly already well known to Toomre and Zang in 1976.

As for the non-axisymmetric case, we can rewrite the integral so that it is carried out over $\tilde{U}$ and χ instead of u and v. We then obtain

$$\begin{aligned}\mathcal{R} = \left(1 - \frac{\beta}{2}\right) 2\pi G K(\alpha, 0) \tilde{C} R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} \left|\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}\right| \int d\tilde{U} \mathcal{J}_0(\tilde{U}) \tilde{U} \\ \times \left|\frac{\beta\tilde{U}^2 + \beta - 2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}-1} \frac{1}{2\pi} \int \frac{d\chi}{\tilde{R}} \left\{ 1 - \frac{1}{T} \int_{t-T}^{t} \left(\frac{R'}{R}\right)^{i\alpha - \frac{1}{2}} dt' \right\}.\end{aligned} \tag{4.28}$$

We now embark on demonstrating the relationship between the expressions derived here, eqs. (4.23) and (4.28), and that obtained directly from the integral equation, eq. (4.12). It is easiest to work backwards from the latter expression. We begin with the case $m \neq 0$. We consider the sum over l

in (4.12). Substituting for $Q^*_{lm}(\alpha)$ from the definition of the Fourier coefficient (3.55), we obtain

$$\begin{aligned}\sum_{l=-\infty}^{\infty} Q_{lm}(\alpha)Q^*_{lm}(\alpha)&\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}\\ &= \frac{1}{2\pi}\int_0^{2\pi} e^{-im(\theta-\Omega t)-(i\alpha+\frac{1}{2})\tilde{X}} d\chi \sum_{l=-\infty}^{\infty} Q_{lm}(\alpha)e^{il\chi}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}\end{aligned} \tag{4.29}$$

But from (3.53) we have

$$\sum_{l=-\infty}^{+\infty} Q_{lm}(\alpha)e^{il\chi} = e^{im(\theta-\Omega t)+(i\alpha-\frac{1}{2})\tilde{X}} \tag{4.30}$$

where $\tilde{X} = \ln\tilde{R} = \ln(R/R_{\mathrm{H}})$. This gives us the first term in curly brackets in (4.29). To obtain the second term, we multiply both sides of (4.30) by $\exp(im\Omega t + i\omega t)$ and integrate between $t' = t - T$ and $t' = t$:

$$-i\left[\sum_{l=-\infty}^{+\infty} \frac{Q_{lm}(\alpha)e^{il\kappa t'+im\Omega t'+i\omega t'}}{l\kappa+m\Omega+\omega}\right]_{t'=t-T}^{t'=t} = \int_{t-T}^{t} e^{im\theta'+i\omega t'}\left(\frac{R'}{R_{\mathrm{H}}}\right)^{i\alpha-\frac{1}{2}} dt'. \tag{4.31}$$

Remembering that $\exp(il\kappa T) = \exp(2i\pi l) = 1$, this is

$$-i\sum_{l=-\infty}^{+\infty} \frac{Q_{lm}(\alpha)e^{il\kappa t+im\Omega t+i\omega t}}{l\kappa+m\Omega+\omega}\left(1-e^{-im\Omega T-i\omega T}\right) = \int_{t-T}^{t} e^{im\theta'+i\omega t'}\left(\frac{R'}{R_{\mathrm{H}}}\right)^{i\alpha-\frac{1}{2}} dt'. \tag{4.32}$$

We now let $\omega \to 0$ to obtain (in terms of $\chi = \kappa t$ and the dimensionless variables $\tilde{\kappa}$ and $\tilde{\Omega}$)

$$\sum_{l=-\infty}^{+\infty} \frac{Q_{lm}(\alpha)e^{il\chi}}{l\tilde{\kappa}+m\tilde{\Omega}} = \frac{iv_\beta}{R_0\tilde{R}_{\mathrm{H}}^{1+\beta/2}}\frac{e^{-im\Omega t}}{1-e^{-im\Omega T}}\int_{t-T}^{t} e^{im\theta'}\left(\frac{R'}{R_{\mathrm{H}}}\right)^{i\alpha-\frac{1}{2}} dt'. \tag{4.33}$$

We can now substitute (4.30) and (4.33) into (4.29):

$$\begin{aligned}\sum_{l=-\infty}^{\infty} Q_{lm}(\alpha)Q^*_{lm}(\alpha)&\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\} = \frac{1}{2\pi}\int_0^{2\pi}\frac{d\chi}{\tilde{R}}\\ &\times\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \frac{im\gamma v_\beta}{R_0\tilde{R}_{\mathrm{H}}^{1+\beta/2}}\frac{1}{1-e^{-im\Omega T}}\int_{t-T}^{t} e^{im(\theta'-\theta)}\left(\frac{R'}{R}\right)^{i\alpha-\frac{1}{2}} dt'.\right\}\end{aligned} \tag{4.34}$$

With this result, the expression for the response function (4.12) obtained from the integral equation is identical to the alternative expression (4.23) obtained by assuming that the response is a pure log-spiral.

Eq. (4.34) fails for $m = 0$. To derive a suitable expression for the axisymmetric case, we consider the sum of the modulus-squared of the Fourier coefficients over all radial harmonics except the zeroth: $\sum_{l=-\infty,l\neq 0}^{+\infty}|Q_{l0}(\alpha)|^2$. Using eq. (3.55) to substitute for $Q^*_{l0}(\alpha)$, we write this as

$$\sum_{l=-\infty,l\neq 0}^{+\infty} Q_{l0}(\alpha)Q^*_{l0}(\alpha) = \frac{1}{2\pi}\int_0^{2\pi} e^{-(i\alpha+\frac{1}{2})\tilde{X}}\sum_{l=-\infty,l\neq 0}^{+\infty} Q_{l0}(\alpha)e^{il\kappa t}d\chi. \tag{4.35}$$

The definition of the Fourier coefficients (3.53) gives us

$$e^{(i\alpha-\frac{1}{2})\tilde{X}} = \sum_{l=-\infty}^{+\infty} Q_{l0}(\alpha)e^{il\kappa t} = \sum_{l=-\infty,l\neq 0}^{+\infty} Q_{l0}(\alpha)e^{il\kappa t} + Q_{00}(\alpha). \tag{4.36}$$

Writing

$$Q_{00}(\alpha) = \frac{1}{T}\int_{t-T}^{t} \tilde{R}'^{i\alpha-\frac{1}{2}}dt', \qquad e^{(i\alpha+\frac{1}{2})\tilde{X}} = \tilde{R}^{(i\alpha+\frac{1}{2})}, \tag{4.37}$$

where $\tilde{R}'$ is used as a shorthand for $\tilde{R}(t')$, we have

$$e^{-(i\alpha+\frac{1}{2})\tilde{X}} \sum_{l=-\infty,l\neq 0}^{+\infty} Q_{l0}(\alpha)e^{il\kappa t} = \frac{1}{\tilde{R}}\left\{1 - \frac{1}{T}\int_{t-T}^{t}\left(\frac{R'}{R}\right)^{i\alpha-\frac{1}{2}}dt'\right\}. \tag{4.38}$$

Substituting this into the expression for the Fourier series (4.35), we obtain

$$\sum_{l=-\infty,l\neq 0}^{+\infty} Q_{l0}(\alpha)Q_{l0}^{*}(\alpha) = \frac{1}{2\pi}\int_{0}^{2\pi}\frac{d\chi}{\tilde{R}}\left\{1 - \frac{1}{T}\int_{t-T}^{t}\left(\frac{R'}{R}\right)^{i\alpha-\frac{1}{2}}dt'\right\}. \tag{4.39}$$

We can now substitute this into our expression for the response function (4.28) to obtain

$$\begin{aligned}\mathcal{R} = {} & 2\pi GK(\alpha,0)\tilde{C}\left|\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}\right|\left(1-\frac{\beta}{2}\right)v_{\beta}^{\frac{2}{\beta}(1+\gamma)}R_0^{\gamma+1} \\ & \times\int \mathcal{J}_0(\tilde{U})\tilde{U}d\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}-1}\sum_{l=-\infty,l\neq 0}^{+\infty}|Q_{l0}(\alpha)|^2.\end{aligned} \tag{4.40}$$

Equating this to our previous expression (4.12) with $m = 0$, we now see how to resolve the troublesome $l = 0$ harmonic. It is simply omitted! In terms of our integral equation approach, this means that we must take $F_{00} = 0$.

In subsequent chapters, we shall investigate the form of the response function for different azimuthal harmonics. We shall find that modes, for which $\mathcal{R} = 1$, are in general possible at sufficiently low temperature.

4.4 Growing modes

As discussed in section 4.2, growing modes are not self-similar, so that the response to a pure log-spiral now contains contributions at many wavenumbers. Thus any investigation of the growing modes necessarily involves the full integral equation approach. We have also shown that the dependence on pattern speed and growth rate could be factored out of the integral equation. We obtained a modified integral equation (4.11):

$$\lambda\hat{A}(\alpha) = \hat{A}(\alpha)\mathcal{S}_m^{(2)}(\alpha,\alpha) + \int_{-\infty}^{+\infty} d\alpha'\hat{A}(\alpha')\hat{\mathcal{S}}_m^{(1)}(\alpha,\alpha').$$

This equation is hard to solve numerically, since the term $\mathcal{S}_m^{(1)}(\alpha,\alpha)$ is singular at $\alpha = \alpha'$. We have therefore concentrated on its analytical properties. It immediately seems unlikely that growing, non-axisymmetric modes could exist in the self-consistent disk. All the quantities in eq. (4.11) are complex, so we effectively have two equations: we must satisfy simultaneously $\mathrm{Re}(\lambda) = 1$ and $\mathrm{Im}(\lambda) = 0$. In the cut-out case, we had two free parameters: the frequency ω and temperature γ. Now, in the self-consistent case, we have only one free parameter: the temperature γ. It seems unlikely that we can satisfy both equations simultaneously without some special circumstances. In the axisymmetric case, this is exactly what happens. We shall show that when $m = 0$, the kernel is Hermitian and the imaginary part of the eigenvalue vanishes identically. We thus have one equation and one free parameter, which we can in principle adjust to obtain growing axisymmetric modes. When $m \neq 0$, we have been unable to prove that there are no solutions. However, we believe it is unlikely that any exist.

In this section, we record our investigations into the growing modes. We begin by defining a modified transfer function and eigenfunction.

$$T_m(\alpha,\alpha') = \sqrt{\frac{K(\alpha,m)}{K(\alpha',m)}}\hat{\mathcal{S}}_m^{(1)}(\alpha,\alpha'), \qquad\qquad B(\alpha) = \sqrt{K(\alpha,m)}\hat{A}(\alpha). \tag{4.41}$$

The integral equation (4.11) now becomes

$$\lambda B(\alpha) = B(\alpha)\mathcal{S}_m^{(2)}(\alpha,\alpha) + \int_{-\infty}^{+\infty} d\alpha' B(\alpha') T_m(\alpha,\alpha'). \tag{4.42}$$

We have, from eqs. (4.41), (4.3) and (3.79)

$$\begin{aligned} T_m(\alpha,\alpha') = -i\frac{2\pi G\tilde{C}}{1-e^{2\pi\hat{\eta}}}.\frac{2-\beta}{2+\beta}R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)}\sqrt{K(\alpha,m)K(\alpha',m)}\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}} \\ \times\sum_{l=-\infty}^{+\infty}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\} Q_{lm}(\alpha')Q_{lm}^*(\alpha)e^{-i\hat{\eta}\ln(l\tilde{\kappa}+m\tilde{\Omega})}. \end{aligned} \tag{4.43}$$

For axisymmetric perturbations, the angular momentum function is zero at $l = 0$: $F_{00} = 0$. Thus the $l = 0$ term must be omitted from the sum in this case, although for notational brevity this is not explicitly indicated. Care is required with the logarithm. $l\tilde{\kappa}+m\tilde{\Omega}$ can be positive or negative depending on the sign of l. As discussed in Appendix C, we choose the phase of the quantity $(l\tilde{\kappa}+m\tilde{\Omega})/\tilde{\omega}$ to be such that the imaginary part of the logarithm $\ln[(l\tilde{\kappa}+m\tilde{\Omega})/\tilde{\omega}]$ is in the range 0 to 2π. This requires the choice (C.43)

$$\ln(l\tilde{\kappa}+m\tilde{\Omega}) = \ln|l\tilde{\kappa}+m\tilde{\Omega}| + 2i\pi, \qquad\qquad l\tilde{\kappa}+m\tilde{\Omega} > 0, \tag{4.44}$$

$$\ln(l\tilde{\kappa}+m\tilde{\Omega}) = \ln|l\tilde{\kappa}+m\tilde{\Omega}| + i\pi, \qquad\qquad l\tilde{\kappa}+m\tilde{\Omega} < 0. \tag{4.45}$$

We therefore have to split the sum in eq. (4.43) into prograde and retrograde parts. We define the critical harmonic l_c to be the integer less than or equal to $-m\tilde{\Omega}/\tilde{\kappa}$. Then the prograde terms are those for which $l \geq l_\mathrm{c}$, and the retrograde those for which $l < l_\mathrm{c}$.

Then (4.43) becomes

$$T_m(\alpha,\alpha') = -i\frac{2\pi G\tilde{C}e^{\pi\hat{\eta}}}{1-e^{2\pi\hat{\eta}}}\frac{2-\beta}{2+\beta}R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)}\sqrt{K(\alpha,m)K(\alpha',m)}\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}}$$
$$\times\left\{\sum_{l=-\infty}^{l_c-1}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}Q_{lm}(\alpha')Q^*_{lm}(\alpha)e^{-i\hat{\eta}\ln|l\tilde{\kappa}+m\tilde{\Omega}|}\right.$$
$$\left.+\sum_{l=l_c}^{+\infty}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}Q_{lm}(\alpha')Q^*_{lm}(\alpha)e^{-i\hat{\eta}\ln|l\tilde{\kappa}+m\tilde{\Omega}|}e^{\pi\hat{\eta}}\right\}. \tag{4.46}$$

We now make the further definitions

$$T_m(\alpha,\alpha') = G_m(\alpha,\alpha')e^{\pi\hat{\eta}/2}, \qquad B(\alpha) = C(\alpha)e^{\pi\hat{\alpha}/2}. \tag{4.47}$$

The integral equation (4.42) now becomes

$$\lambda C(\alpha) = C(\alpha)\mathcal{S}_m^{(2)}(\alpha,\alpha) + \int_{-\infty}^{+\infty}d\alpha' C(\alpha')G_m(\alpha,\alpha'). \tag{4.48}$$

The new transfer function is

$$G_m(\alpha,\alpha') = -i\frac{2\pi G\tilde{C}e^{\pi\hat{\eta}/2}}{1-e^{2\pi\hat{\eta}}}\frac{2-\beta}{2+\beta}R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)}\sqrt{K(\alpha,m)K(\alpha',m)}\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}}$$
$$\times\left\{\sum_{l=-\infty}^{l_c-1}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}Q_{lm}(\alpha')Q^*_{lm}(\alpha)e^{-i\hat{\eta}\ln|l\tilde{\kappa}+m\tilde{\Omega}|}\right.$$
$$\left.+\sum_{l=l_c}^{+\infty}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}Q_{lm}(\alpha')Q^*_{lm}(\alpha)e^{-i\hat{\eta}\ln|l\tilde{\kappa}+m\tilde{\Omega}|}e^{\pi\hat{\eta}}\right\}. \tag{4.49}$$

We now divide this into Hermitian and anti-Hermitian parts. We write

$$G_m(\alpha,\alpha') = G_m^{(h)}(\alpha,\alpha') + G_m^{(a)}(\alpha,\alpha') \tag{4.50}$$

where

$$G_m^{(h)*}(\alpha',\alpha) = G_m^{(h)}(\alpha,\alpha'), \qquad G_m^{(a)*}(\alpha',\alpha) = -G_m^{(a)}(\alpha,\alpha'). \tag{4.51}$$

We find that

$$G_m^{(h)}(\alpha,\alpha') = i\frac{2\pi G\tilde{C}}{4\sinh\frac{\hat{\eta}\pi}{2}}\frac{2-\beta}{2+\beta}R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)}\sqrt{K(\alpha,m)K(\alpha',m)}\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}}$$
$$\times\sum_{l=-\infty}^{+\infty}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}Q_{lm}(\alpha')Q^*_{lm}(\alpha)e^{-i\hat{\eta}\ln|l\tilde{\kappa}+m\tilde{\Omega}|}. \tag{4.52}$$

$$
\begin{aligned}
G_m^{(a)}(\alpha,\alpha') = i\frac{2\pi G\tilde{C}}{4\cosh\frac{\hat{\eta}\pi}{2}}\frac{2-\beta}{2+\beta}R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)}\sqrt{K(\alpha,m)K(\alpha',m)}\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}} \\
\times\left\{\sum_{l=l_c}^{+\infty}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}Q_{lm}(\alpha')Q_{lm}^*(\alpha)e^{-i\hat{\eta}\ln|l\tilde{\kappa}+m\tilde{\Omega}|}\right. \\
\left.-\sum_{l=-\infty}^{l_c-1}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}Q_{lm}(\alpha')Q_{lm}^*(\alpha)e^{-i\hat{\eta}\ln|l\tilde{\kappa}+m\tilde{\Omega}|}\right\}.
\end{aligned}
\tag{4.53}
$$

Note that for axisymmetric disturbances, the critical harmonic l_c is zero, and we have the additional symmetry $Q_{l0}(\alpha) = Q_{-l,0}(\alpha)$ (see eq. (3.55)). Furthermore, we have seen that the $l = 0$ term must be omitted in the axisymmetric case. Thus the prograde and retrograde sums cancel in eq. (4.53), so that $G_0^{(a)} = 0$. Axisymmetric modes have Hermitian kernels. (We shall see in Chapter 6 that this is true for cut-out disks as well.)

We record a further symmetry property of $G_m^{(h)}$ and $G_m^{(a)}$:

$$
G_m^{(h)*}(-\alpha,-\alpha') = G_m^{(h)}(\alpha,\alpha'), \qquad G_m^{(a)*}(-\alpha,-\alpha') = -G_m^{(a)}(\alpha,\alpha'). \tag{4.54}
$$

From eqs. (3.79) and (4.4), $\mathcal{S}_m^{(2)}(\alpha,\alpha')$ is

$$
\begin{aligned}
\mathcal{S}_m^{(2)}(\alpha,\alpha') = \frac{2-\beta}{2}\pi G\tilde{C}R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)}K(\alpha',m)\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}} \\
\times\sum_{l=-\infty}^{+\infty}Q_{lm}(\alpha')Q_{lm}^*(\alpha)\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}
\end{aligned}
\tag{4.55}
$$

We note that $\mathcal{S}_m^{(2)}(\alpha,\alpha)$ is purely real. From the symmetry property of the Fourier coefficients, we also have the symmetry $\mathcal{S}_m^{(2)}(\alpha,\alpha) = \mathcal{S}_m^{(2)}(-\alpha,-\alpha)$.

The integral equation (4.48) is

$$
\lambda C(\alpha) = C(\alpha)\mathcal{S}_m^{(2)}(\alpha,\alpha) + \int_{-\infty}^{+\infty}d\alpha' C(\alpha')G_m^{(h)}(\alpha,\alpha') + \int_{-\infty}^{+\infty}d\alpha' C(\alpha')G_m^{(a)}(\alpha,\alpha'). \tag{4.56}
$$

In principle, this must be solved to obtain the eigenvalues λ and eigenfunctions $C(\alpha)$. We are interested in proving definite results about the eigenvalue – for instance, if we could show that its real part is necessarily negative, we could conclude that a continuum of modes is not possible. We have derived separate equations for $\mathrm{Re}(\lambda)$ and $\mathrm{Im}(\lambda)$ in terms of $C(\alpha)$. Even though $C(\alpha)$ is unknown, our hope was that we should be able to derive constraints on λ which hold for *arbitrary* functions $C(\alpha)$. In fact, we have been unable to prove any such results for non-axisymmetric modes. Nevertheless, in the following paragraphs we present our investigations into the problem.

Multiplying (4.56) by $C^*(\alpha)$ and integrating over wavenumber α, we obtain

$$
\begin{aligned}
\lambda = \int_{-\infty}^{+\infty}d\alpha|C(\alpha)|^2\mathcal{S}_m^{(2)}(\alpha,\alpha) + \int_{-\infty}^{+\infty}d\alpha\int_{-\infty}^{+\infty}d\alpha' C(\alpha')C^*(\alpha)G_m^{(h)}(\alpha,\alpha') \\
+\int_{-\infty}^{+\infty}d\alpha\int_{-\infty}^{+\infty}d\alpha' C(\alpha')C^*(\alpha)G_m^{(a)}(\alpha,\alpha'),
\end{aligned}
\tag{4.57}
$$

where for convenience we have assumed that the eigenfunction is normalised such that $\int |C(\alpha)|^2 d\alpha = 1$. We take the complex conjugate and interchange α and α' in the double integral:

$$\begin{aligned} \lambda^* = \int_{-\infty}^{+\infty} d\alpha |C(\alpha)|^2 \mathcal{S}_m^{(2)}(\alpha,\alpha) + \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} d\alpha' C^*(\alpha) C(\alpha') G_m^{(h)*}(\alpha',\alpha) \\ + \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} d\alpha' C^*(\alpha) C(\alpha') G_m^{(a)*}(\alpha',\alpha), \end{aligned} \tag{4.58}$$

where we have used the fact that $\mathcal{S}_m^{(2)}(\alpha,\alpha)$ is real. We now use the symmetry properties of $G_m^{(h)}$ and $G_m^{(a)}$:

$$\begin{aligned} \lambda^* = \int_{-\infty}^{+\infty} d\alpha |C(\alpha)|^2 \mathcal{S}_m^{(2)}(\alpha,\alpha) + \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} d\alpha' C(\alpha') C^*(\alpha) G_m^{(h)}(\alpha,\alpha') \\ - \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} d\alpha' C(\alpha') C^*(\alpha) G_m^{(a)}(\alpha,\alpha'), \end{aligned} \tag{4.59}$$

Adding eq. (4.59) to eq. (4.57), we obtain

$$\mathrm{Re}(\lambda) = \int_{-\infty}^{+\infty} d\alpha |C(\alpha)|^2 \mathcal{S}_m^{(2)}(\alpha,\alpha) + \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} d\alpha' C(\alpha') C^*(\alpha) G_m^{(h)}(\alpha,\alpha') \tag{4.60}$$

and subtracting

$$i\mathrm{Im}(\lambda) = \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} d\alpha' C(\alpha') C^*(\alpha) G_m^{(a)}(\alpha,\alpha'). \tag{4.61}$$

We introduce a function

$$D_{lm}(\alpha) = \sqrt{K(\alpha,m)}\, C(\alpha) Q_{lm}(\alpha) e^{i\hat{\alpha} \ln |l\tilde{\kappa} + m\tilde{\Omega}|} \tag{4.62}$$

In terms of this function, the real equation (4.60) can be written (substituting in the expression for the Hermitian kernel $G_m^{(h)}$ from (4.52)).

$$\begin{aligned} \mathrm{Re}(\lambda) = \frac{2-\beta}{2} \pi G \tilde{C} R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} \int d\tilde{U}\, \mathcal{J}_0(\tilde{U}) \tilde{U} \left| \frac{\beta \tilde{U}^2 + \beta - 2}{2\beta} \right|^{\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}} \\ \times \sum_{l=-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\alpha \left\{ \left| \frac{2 + 2\gamma - \beta\gamma}{\beta \tilde{U}^2 + \beta - 2} \right| - \frac{\gamma m}{l\tilde{\kappa} + m\tilde{\Omega}} \right\} |D_{lm}(\alpha)|^2 \\ + \frac{i}{2+\beta} \frac{2-\beta}{2} \pi G \tilde{C} R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} \int d\tilde{U}\, \mathcal{J}_0(\tilde{U}) \tilde{U} \left| \frac{\beta \tilde{U}^2 + \beta - 2}{2\beta} \right|^{\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}} \\ \times \sum_{l=-\infty}^{+\infty} \int_{-\infty}^{+\infty} d\alpha D_{lm}^*(\alpha) \int_{-\infty}^{+\infty} d\alpha' D_{lm}(\alpha') \left\{ \left| \frac{2 + 2\gamma - \beta\gamma}{\beta \tilde{U}^2 + \beta - 2} \right| - \frac{\gamma m}{l\tilde{\kappa} + m\tilde{\Omega}} \right\} \mathrm{cosech} \frac{\pi(\hat{\alpha} - \hat{\alpha}')}{2} \end{aligned} \tag{4.63}$$

The imaginary equation (4.61) can be written (substituting in the expression for the anti-Hermitian kernel

$G_m^{(a)}$ from (4.53))

$$\begin{aligned}\mathrm{Im}(\lambda) = {} & \frac{\pi G\tilde{C}}{2}\frac{2-\beta}{2+\beta}R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)}\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}} \\ & \times\left\{\sum_{l=l_c}^{+\infty}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}\int_{-\infty}^{+\infty}d\alpha D_{lm}^*(\alpha)\int_{-\infty}^{+\infty}d\alpha' D_{lm}(\alpha')\mathrm{sech}\frac{\pi(\hat{\alpha}-\hat{\alpha}')}{2}\right. \\ & \left.-\sum_{l=-\infty}^{l_c-1}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}\int_{-\infty}^{+\infty}d\alpha D_{lm}^*(\alpha)\int_{-\infty}^{+\infty}d\alpha' D_{lm}(\alpha')\mathrm{sech}\frac{\pi(\hat{\alpha}-\hat{\alpha}')}{2}\right\}.\end{aligned} \tag{4.64}$$

We note that the double integrals over wavenumber contain a convolution. We define the Fourier transform $\tilde{D}_{lm}(x)$ as

$$D_{lm}(\alpha) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}dx\,e^{i\alpha x}\tilde{D}_{lm}(x), \qquad \tilde{D}_{lm}(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}d\alpha\,e^{-i\alpha x}D_{lm}(\alpha). \tag{4.65}$$

From the definition of a convolution (e.g. Arfken 1966),

$$D_{lm} * f = \int_{-\infty}^{+\infty}d\alpha' D_{lm}(\alpha')f(\alpha-\alpha'). \tag{4.66}$$

Substituting for f in terms of its Fourier transform $\tilde{f}(x)$:

$$D_{lm} * f = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}dx\tilde{f}(x)e^{i\alpha x}\int_{-\infty}^{+\infty}d\alpha' D_{lm}(\alpha')e^{-i\alpha' x} = \int_{-\infty}^{+\infty}dx\,e^{i\alpha x}\tilde{f}(x)\tilde{D}_{lm}(x). \tag{4.67}$$

This is the convolution theorem, which states that the convolution of two functions is equal to the inverse Fourier transform of the product of the transforms of the individual functions. Conversely, the Fourier transform of a convolution is equal to the product of the individual transforms. We therefore have

$$\int_{-\infty}^{+\infty}d\alpha\,e^{-i\alpha x}\int_{-\infty}^{+\infty}d\alpha' D_{lm}(\alpha')f(\alpha-\alpha') = 2\pi\tilde{D}_{lm}(x)\tilde{f}(x). \tag{4.68}$$

This result is useful in our expressions for the real and imaginary parts of the eigenvalue. We begin by substituting the Fourier transform of $D_{lm}^*(\alpha)$:

$$\begin{aligned}&\int_{-\infty}^{+\infty}d\alpha D_{lm}^*(\alpha)\int_{-\infty}^{+\infty}d\alpha' D_{lm}(\alpha')f(\alpha-\alpha') \\ &\qquad = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}dx\,\tilde{D}_{lm}^*(x)\int_{-\infty}^{+\infty}d\alpha e^{-i\alpha x}\int_{-\infty}^{+\infty}d\alpha' D_{lm}(\alpha')f(\alpha-\alpha')\end{aligned} \tag{4.69}$$

Now we substitute our result (4.68) and obtain

$$\int_{-\infty}^{+\infty}d\alpha D_{lm}^*(\alpha)\int_{-\infty}^{+\infty}d\alpha' D_{lm}(\alpha')f(\alpha-\alpha') = \sqrt{2\pi}\int_{-\infty}^{+\infty}dx\,|\tilde{D}_{lm}(x)|^2\tilde{f}(x). \tag{4.70}$$

Thus we have converted our double integral over wavenumber to a single integral over radius. The function $\tilde{D}_{lm}(x)$ is unknown, but appears only as $|\tilde{D}_{lm}(x)|^2$, which is positive definite.

In the real equation (4.63), the function $f(\alpha)$ and its transform are (see e.g. Gradshteyn & Ryzhik 1978, eq. (3.981.1))

$$f(\alpha) = \operatorname{cosech}\frac{\pi\alpha}{2+\beta}, \qquad \tilde{f}(x) = -i\frac{2+\beta}{\sqrt{2\pi}}\tanh\left(\frac{2+\beta}{2}x\right). \tag{4.71}$$

In the imaginary equation (4.64), they are (see e.g. Gradshteyn & Ryzhik 1978, eq. (3.981.3))

$$f(\alpha) = \operatorname{sech}\frac{\pi\alpha}{2+\beta}, \qquad \tilde{f}(x) = \frac{2+\beta}{\sqrt{2\pi}}\operatorname{sech}\left(\frac{2+\beta}{2}x\right). \tag{4.72}$$

Then eq. (4.63) becomes

$$\begin{aligned}\mathrm{Re}(\lambda) = {} & \frac{2-\beta}{2}\pi G\tilde{C}R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)}\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}} \\ & \times\sum_{l=-\infty}^{+\infty}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\} \\ & \times\left[\int_{-\infty}^{+\infty}d\alpha|D_{lm}(\alpha)|^2 + \int_{-\infty}^{+\infty}dx|D_{lm}(x)|^2\tanh\left(\frac{2+\beta}{2}x\right).\right]\end{aligned} \tag{4.73}$$

Now we use Parseval's theorem (e.g. Arfken (1966)):

$$\int_{-\infty}^{+\infty}d\alpha|D_{lm}(\alpha)|^2 = \int_{-\infty}^{+\infty}dx|\tilde{D}_{lm}(x)|^2. \tag{4.74}$$

$$\begin{aligned}\mathrm{Re}(\lambda) = {} & \pi G\tilde{C}\left(1-\frac{\beta}{2}\right)R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)}\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}} \\ & \times\sum_{l=-\infty}^{+\infty}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}\int_{-\infty}^{+\infty}dx|\tilde{D}_{lm}(x)|^2\left[1+\tanh\left(\frac{2+\beta}{2}x\right)\right]\end{aligned} \tag{4.75}$$

$$\begin{aligned}\mathrm{Im}(\lambda) = {} & \pi G\tilde{C}\left(1-\frac{\beta}{2}\right)R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)}\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}} \\ & \times\left[\sum_{l=l_c}^{+\infty}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}\int_{-\infty}^{+\infty}dx\,|\tilde{D}_{lm}(x)|^2\operatorname{sech}\left(\frac{2+\beta}{2}x\right)\right. \\ & \left.-\sum_{l=-\infty}^{l_c-1}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}\int_{-\infty}^{+\infty}dx\,|\tilde{D}_{lm}(x)|^2\operatorname{sech}\left(\frac{2+\beta}{2}x\right)\right].\end{aligned} \tag{4.76}$$

Our attempt to constrain the allowed range of λ has proved less fruitful than hoped. For general m, we cannot draw any conclusions about the sign of $\mathrm{Im}(\lambda)$, because we do not know whether the prograde or retrograde sum will "win". Similarly, the presence of the term in curly brackets in eq. (4.75) prevents us from drawing any conclusions about the sign of $\mathrm{Re}(\lambda)$. We can, however, draw an additional conclusion about the eigenvalues of modes with $m = 0$. In this case, the integrand is positive definite. For axisymmetric modes, then, we can at least conclude that the eigenvalue is real and positive. A continuum of axisymmetric modes thus appears possible.

For non-axisymmetric modes, it would be surprising if eqs. (4.75) and (4.76) could simultaneously be satisfied at *any* temperature γ. It seems even less likely that they could be satisfied over a *range* of temperatures γ, as must happen if the growing modes affect all disks below a critical temperature. To clarify the remarkable properties which would have to be possessed by any solution, we rewrite eq. (4.76) as

$$\mathrm{Im}(\lambda) = \pi G \tilde{C} \left(1 - \frac{\beta}{2}\right) R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} \int_{-\infty}^{+\infty} dx \, \mathrm{sech}\left(\frac{2+\beta}{2} x\right) \int d\tilde{U} \, \mathcal{J}_0(\tilde{U}) \tilde{U} \left|\frac{\beta \tilde{U}^2 + \beta - 2}{2\beta}\right|^{\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}}$$
$$\times \sum_{l=-\infty}^{+\infty} \left\{ \left| \frac{2 + 2\gamma - \beta\gamma}{\beta \tilde{U}^2 + \beta - 2} \right| - \frac{\gamma m}{l \tilde{\kappa} + m \tilde{\Omega}} \right\} |\tilde{D}_{lm}(x)|^2 \mathrm{sign}(l - l_\mathrm{c}). \tag{4.77}$$

If the growing modes, like other modes, exist at a range of temperatures, the eigenvalue must be real – and so this integral must vanish – for *all* temperatures less than the critical temperature. This is a powerful constraint. We have a problem where in shorthand

$$g(\gamma) = \int_{-\infty}^{+\infty} dx \, f(x, \gamma). \tag{4.78}$$

Under what circumstances can $g(\gamma)$ vanish for a range of γ? Two obvious possibilities are that the integrand $f(x, \gamma)$ is odd or identically zero. However, there are functions $f(x, \gamma)$ which are even and have a finite number of zeros, which yet give zero $g(\gamma)$ for a range of γ. As an example, if $f(x, \gamma)$ has the form

$$f(x, \gamma) = a_0(\gamma) e^{-B_0(\gamma) x^2} + a_1(\gamma) x e^{-B_1(\gamma) x^2} + a_2(\gamma) x^2 e^{-B_2(\gamma) x^2} + a_3(\gamma) x^3 e^{-B_3(\gamma) x^2} \ldots \tag{4.79}$$

then $g(\gamma)$ is

$$g(\gamma) = \sqrt{\pi} \left(\frac{a_0(\gamma)}{(B_0(\gamma))^{1/2}} + \frac{a_2(\gamma)}{2(B_2(\gamma))^{3/2}} + \frac{3 a_4(\gamma)}{4(B_4(\gamma))^{5/2}} \ldots \right) \tag{4.80}$$

It is apparent that, by suitable choice of the $a_i(y)$ and $B_i(y)$ for even i, the integral $g(\gamma)$ can be made to vanish identically. A simple example is the function

$$f(x, \gamma) = (1 - 2\gamma x^2) e^{-\gamma x^2} \tag{4.81}$$

Thus although it seems to us highly unlikely that there exist any eigenfunctions simultaneously satisfying $\mathrm{Re}(\lambda) = 1$ and $\mathrm{Im}(\lambda) = 0$, we have been unable to rule it out. It is in theory possible that a continuum of non-axisymmetric modes exists, either at one particular temperature γ, or even across a range of temperatures.

The relationship between neutral and growing modes

In section 4.2, we stated that the neutral modes were isolated from the continuum of growing modes, and would in general set in at a different temperature. We also claimed that the growing modes would have a different shape from the neutral modes, since (loosely speaking) the growing modes are built of log-spirals at many wavenumbers, whereas the neutral modes are pure log-spirals. In this section, we justify these claims in more detail by investigating the relationship between the neutral modes and any growing modes which might exist. Specifically, we are interested in whether a neutral mode could ever satisfy the integral equation. If this were the case, the neutral modes would not be isolated, but would form "one end" of the continuum – modes would be possible with all growth rates and pattern speeds, down to *and including* $s = 0$ and $\Omega_{\mathrm{p}} = 0$. In this case, the existence of neutral modes at a particular temperature would imply the existence of a continuum of modes. The temperature at which the disk first admits neutral modes – which we can easily calculate – would then be the temperature at which the disk becomes wildly unstable to a continuum of modes. The neutral modes would thus be of crucial importance in determining the stability of the disk, rather than the isolated curiosities we have so far considered them.

However, we shall now prove that no neutral modes can ever satisfy the integral equation. This is an important result, because it proves that the neutral modes are always isolated from the growing modes. They could conceivably set in at the same temperature, but there would be no compelling reason for them to do so. It means that the neutral modes are irrelevant in determining the existence of a continuum.

We begin by considering a neutral mode consisting of a pure log-spiral, $A(\alpha) = A_1\delta(\alpha-\alpha_1)$. The modified eigenfunction $C(\alpha)$ is then

$$C(\alpha) = A_1\sqrt{K(\alpha,m)}e^{-\pi\hat{\alpha}/2}e^{-i\hat{\alpha}\ln\tilde{\omega}}\delta(\alpha-\alpha_1). \tag{4.82}$$

Substituting this into the modified integral equation (4.48):

$$\begin{aligned}\lambda\delta(\alpha-\alpha_1) &= \delta(\alpha-\alpha_1)\mathcal{S}_m^{(2)}(\alpha,\alpha)\\ &+\sqrt{\frac{K(\alpha_1,m)}{K(\alpha,m)}}e^{\pi(\hat{\alpha}-\hat{\alpha_1})/2}e^{i(\hat{\alpha}-\hat{\alpha_1})\ln\tilde{\omega}}G_m(\alpha,\alpha_1).\end{aligned} \tag{4.83}$$

If we simply integrate this over α, it seems entirely reasonable that, at some temperature, solutions could exist with unit mathematical eigenvalue. For the case $m = 0$, we would simply require

$$\sqrt{K(\alpha_1,0)}e^{-i\hat{\alpha_1}\ln|s|}\int_{-\infty}^{\infty}d\alpha\, e^{i\hat{\alpha}\ln|s|}\frac{G_0(\alpha,\alpha_1)}{\sqrt{K(\alpha,0)}} = \frac{1}{2}. \tag{4.84}$$

However, although this is a necessary condition for eq. (4.83) to be satisfied, it is not sufficient. Eq. (4.83) must hold for all α independently. Away from $\alpha = \alpha_1$, this requires

$$\sqrt{\frac{K(\alpha_1,m)}{K(\alpha,m)}}e^{\pi(\hat{\alpha}-\hat{\alpha_1})/2}e^{i(\hat{\alpha}-\hat{\alpha_1})\ln\tilde{\omega}}G_m(\alpha,\alpha_1) = 0, \tag{4.85}$$

for all $\alpha \neq \alpha_1$. It immediately seems highly implausible that such a condition could be satisfied for *all* wavenumbers, and the reader may be happy to conclude that neutral modes cannot satisfy the integral equation. However, it is worth going through a slightly more detailed treatment for the axisymmetric case. This is the easiest case to deal with analytically, and also the most interesting, since it is only for $m = 0$ that we believe a continuum could exist.

For $m = 0$, eq. (4.85) becomes (substituting for G_m from eq. (4.52))

$$\begin{aligned} &K(\alpha_1, 0)e^{\pi(\hat{\alpha}-\hat{\alpha_1})/2}e^{i(\hat{\alpha}-\hat{\alpha_1})\ln\tilde{\omega}}i\frac{2\pi G\tilde{C}}{4\sinh\frac{(\hat{\alpha}-\hat{\alpha}_1)\pi}{2}}\frac{2-\beta}{2+\beta}R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)}\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U} \\ &\times\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}}\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|\sum_{l=-\infty,l\neq0}^{+\infty}Q_{l0}(\alpha_1)Q_{l0}^*(\alpha)e^{-i(\hat{\alpha}-\hat{\alpha}_1)\ln|l\tilde{\kappa}+m\tilde{\Omega}|} = 0. \end{aligned} \tag{4.86}$$

We can simplify this by multiplying throughout by

$$e^{-\pi(\hat{\alpha}-\hat{\alpha_1})/2}e^{-i(\hat{\alpha}-\hat{\alpha_1})\ln\tilde{\omega}}\sinh\frac{(\hat{\alpha}-\hat{\alpha}_1)\pi}{2}$$

since this expression is guaranteed to be non-zero for $\alpha \neq \alpha_1$. Then eq. (4.86) becomes

$$\begin{aligned} f(\alpha) &= 2\pi G\tilde{C}K(\alpha_1, m)\frac{2-\beta}{2+\beta}\left|\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}\right|R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)} \\ &\times\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}-1}\sum_{l=-\infty,l\neq0}^{+\infty}Q_{l0}(\alpha_1)Q_{l0}^*(\alpha)e^{-i(\hat{\alpha}-\hat{\alpha}_1)\ln|l\tilde{\kappa}+m\tilde{\Omega}|} = 0, \end{aligned} \tag{4.87}$$

where for brevity we have called the left-hand side of this equation $f(\alpha)$. We require $f(\alpha) = 0$ for all $\alpha \neq \alpha_1$. Now, by hypothesis, α_1 is the wavenumber of a neutral mode at this temperature. Thus $\mathcal{R}(\alpha_1) = 1$, and from (4.40)

$$\begin{aligned} &2\pi GK(\alpha_1, 0)\tilde{C}\frac{2-\beta}{2+\beta}\left|\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}\right|R_0^{\gamma+1}v_\beta^{\frac{2}{\beta}(1+\gamma)} \\ &\times\int d\tilde{U}\,\mathcal{J}_0(\tilde{U})\tilde{U}\left|\frac{\beta\tilde{U}^2+\beta-2}{2\beta}\right|^{\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}-1}\sum_{l=-\infty,l\neq0}^{+\infty}Q_{l0}(\alpha_1)Q_{l0}^*(\alpha_1) = \frac{2}{2+\beta}. \end{aligned} \tag{4.88}$$

Thus we have $f(\alpha_1) = 2/(2+\beta)$, which is clearly not close to zero. To satisfy our integral equation (4.83), we require $f(\alpha)$ to jump from $2/(2+\beta)$ at $\alpha = \alpha_1$ to zero at $\alpha = \alpha_1 + \epsilon$. That is, $f(\alpha)$ must have a discontinuity at $\alpha = \alpha_1$. However, eq. (4.87) shows that $f(\alpha)$ is in fact a smooth function of α. The Fourier coefficients and exponentials on which it depends change smoothly as the wavenumber is varied. We have therefore shown (4.87) cannot be satisfied for *all* $\alpha \neq \alpha_1$. This means that a pure log-spiral mode is not capable of satisfying the integral equation.

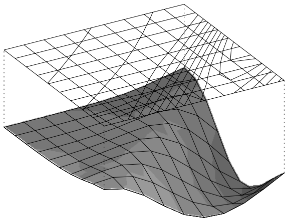
At a given temperature, it is possible that more than one wavenumber could simultaneously satisfy $\mathcal{R}(\alpha) = 1$. If this occurred, the most general neutral mode in that disk would be not a pure log-spiral, but a superposition of several pure log-spirals. The above proof may readily be generalised to show that no superposition of a finite number of log-spirals – and hence no neutral mode at all – is capable of satisfying the integral equation. Equally, we can view this result as proving that any possible solution of the integral equation must be made up of a *continuum* of log-spiral components. (We note that this applies also to the cut-out disks. We have thus found that modes in the cut-out disk must be made up of a continuum of log-spiral components. In subsequent chapters we shall see that this is true, although the density transforms are often fairly peaked at a dominant wavenumber.) Thus, as claimed in section 4.2, the growing modes necessarily have a different morphology from the neutral modes. The neutral modes therefore are isolated from the continuum of growing modes; they can never form one end of the continuum.

4.5 Summary

In this Chapter, we have considered the particular problems associated with a self-similar disk. We have seen that any growing or rotating modes must form a continuum of modes. Thus at any temperature, there are three possibilities. Either there is a continuum of modes with all possible pattern speeds and growth rates, or there are neutral modes, with growth rate and pattern speed both zero, or there are no modes at all. We have found an analytical expression for the response function $\mathcal{R}$, which describes whether or not neutral modes are possible. In subsequent chapters, we shall evaluate this response function for different azimuthal harmonics in turn. We shall find that neutral modes set in for every m at sufficiently low temperature. We have also derived the integral equation which must be satisfied by any growing or rotating modes. We have not been able to find any solutions of this integral equation; equally, we have not be able to prove that none exist. Our treatment is therefore somewhat inconclusive. In this Chapter, we have presented various arguments as to why we believe that there is no continuum of non-axisymmetric modes. We return to this question again in Chapter 10, where we discuss additional evidence gained from our study of the cut-out disks.

Chapter 5

Numerical Methods



5.1 Introduction

This chapter discusses the numerical algorithms required to evaluate the transfer function and to solve the integral equation. Numerical quadrature is needed to find the Fourier coefficients for the stellar orbits and to perform the integration over orbit shape to obtain the transfer function. The integral equation is converted to a standard form: the so-called Fredholm integral equation of the second kind. The kernel is evaluated on a finite grid in wavenumber space. This reduces our problem to a search for the algebraic eigenvalues of a finite matrix.

Results of research that depend largely on numerical calculations have no value unless the computer code has been extensively checked. Part of the aim of this chapter is to convince the reader that the necessary checks and tests have indeed been carried out.

5.2 The Fourier coefficient

The Fourier coefficients are (3.55)

$$Q_{lm}(\alpha) = \frac{1}{2\pi}\int_0^{2\pi} \exp\left\{im\tilde{Y} + \left(i\alpha - \tfrac{1}{2}\right)\tilde{X} - il\chi\right\} d\chi,$$

where $\tilde{Y}$ is the deviation from the mean angular motion, $\tilde{X}$ is the scaled logarithmic radius and χ is the orbital phase. $\tilde{X}$ and $\tilde{Y}$ are both periodic with period T. This means that the integration gives the same result when carried out over any 2π region in χ, as sketched in fig. 5.2. Since the integrand has period 2π, the region on the right marked with lines sloping upwards gives the same contribution as the region on the left marked with lines sloping down. We can replace the integral from 0 to 2π with one centred on the star's pericentre, as indicated by the dashed line. We are therefore free to define $\chi = 0$ to correspond

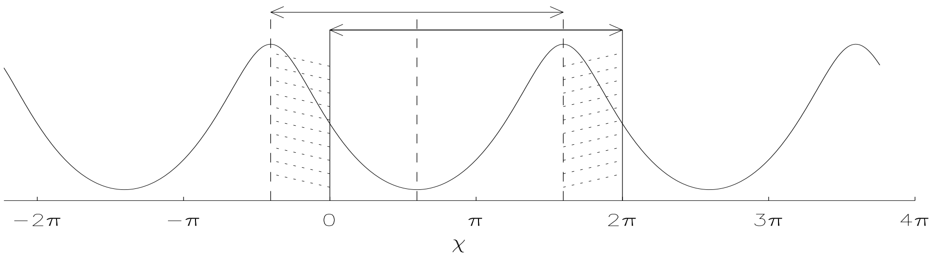


Figure 5.1: A periodic function

to pericentre. This means that at the time $t = 0$, the star has radial coordinate $R = R_{\text{min}}$. Then $\tilde{X}$ will be even, and $\tilde{Y}$ odd, about $\chi = 0$. We write

$$Q_{lm}(\alpha) = \frac{1}{\pi} \int_0^{\pi} \exp\left\{ \left(i\alpha - \tfrac{1}{2}\right)\tilde{X} \right\} \cos(m\tilde{Y} - l\chi)\, d\chi. \tag{5.1}$$

The equations of motion (2.22) are solved by fourth-order Runge-Kutta integration (Press *et al.* 1989, ch. 15) to obtain the stellar position as a function of time, and thus $\tilde{X}$ and $\tilde{Y}$ as a function of χ. The integration over χ is carried out by the mid-point method (Press *et al.* 1989, ch. 4). We used n_ψ points in the mid-point integration, and $2n_\psi$ in the Runge-Kutta. The problem is to choose n_ψ large enough to obtain excellent accuracy, while keeping it as small as possible in order to save time. Eccentric orbits need more work to obtain good accuracy, so, as suggested by Zang (1976), n_ψ is made to depend exponentially on $\tilde{U}$:

$$n_\psi = a_{\text{acc}} \exp(b_{\text{acc}}\tilde{U}). \tag{5.2}$$

We found that the values $a_{\text{acc}} = 10$ and $b_{\text{acc}} = 1.5$ usually worked well.

The following tables compare the effect of different a_{acc} and b_{acc} at two different $\tilde{U}$:

	$Q_{22}(\alpha = 5, \tilde{U} = 0.5)$		
b_{acc}	$a_{\text{acc}} = 5$	$a_{\text{acc}} = 10$	$a_{\text{acc}} = 50$
0.0	$0.196644 - 0.312107i$	$0.197119 - 0.309895i$	$0.197124 - 0.309834i$
1.5	$0.197119 - 0.309895i$	$0.197124 - 0.309837i$	$0.197124 - 0.309834i$
3.0	$0.197124 - 0.309837i$	$0.197124 - 0.309834i$	$0.197124 - 0.309834i$

	$Q_{22}(\alpha = 5, \tilde{U} = 1.5)$		
b_{acc}	$a_{\text{acc}} = 5$	$a_{\text{acc}} = 10$	$a_{\text{acc}} = 50$
0.0	$-0.000510 + 0.031573i$	$-0.044545 - 0.004593i$	$-0.036007 - 0.035313i$
1.5	$-0.022664 - 0.044425i$	$-0.018978 + 0.050528i$	$-0.023356 + 0.039984i$
3.0	$-0.023382 + 0.039989i$	$-0.023318 + 0.039932i$	$-0.023320 + 0.039930i$

The suggested values $a_{\text{acc}} = 10$ and $b_{\text{acc}} = 1.5$ are accurate to 6 significant figures or so for $\tilde{U} = 0.5$, but only 1 s.f. for $\tilde{U} = 1.5$, which might imply that a higher value of a_{acc} or b_{acc} would be better. However,

Q_{22} is also an order of magnitude less at $\tilde{U} = 1.5$ than $\tilde{U} = 0.5$. It turns out that the Fourier coefficients are generally smaller for higher values of the eccentric velocity. In the integrand of the expression for the transfer function, two Fourier coefficients are multiplied together at every eccentric velocity, so the integrand is much smaller at high eccentric velocities. The accuracy with which the transfer function is obtained is thus dominated by the accuracy of the Fourier coefficients at low eccentric velocities. This is useful, since the Fourier coefficients are most costly to evaluate to a given accuracy when the eccentric velocity is high. In section 5.6, we consider the effect of $a_{\rm acc}$ and $b_{\rm acc}$ on the mathematical eigenvalue, and show that the inaccuracy in Q_{lm} at high $\tilde{U}$ does not affect the eigenvalue to 6 s.f.

The following two graphs show the Fourier coefficient as a function of wavenumber, for two different eccentric velocities. The Fourier coefficient is shown only for positive wavenumber. Its behaviour for negative wavenumber is readily deduced from the symmetry property of the Fourier coefficient, $Q_{lm}(\alpha) = Q^*_{lm}(-\alpha)$. We see that $\tilde{U}$ controls the amplitude and the frequency of oscillation of Q_{lm}.

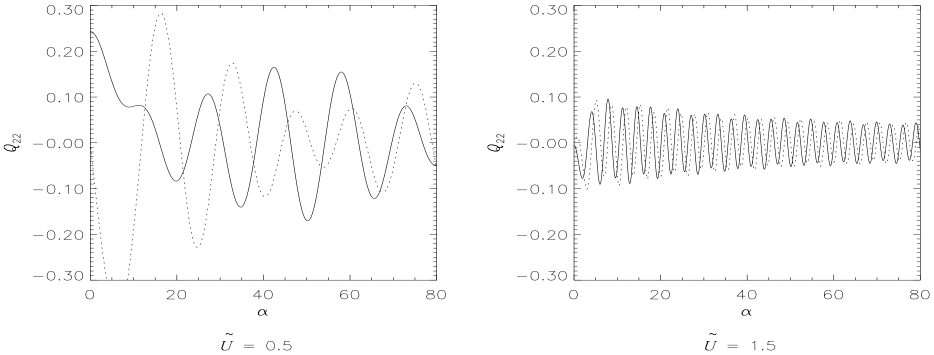


Figure 5.2: Graph of Q_{22} plotted against α. In the left-hand plot, $\tilde{U} = 0.5$; on the right, $\tilde{U} = 1.5$. In each case, the solid line is the real, and the dotted line the imaginary, part of Q_{22}. [$\beta = 0.25$, $m = 2$, $l = 2$; $a_{\rm acc} = 100$, $b_{\rm acc} = 2.5$.]

The next set of plots compares Q_{lm} at different l. For high values of l, Q_{lm} remains close to zero until α is large; thereafter it oscillates with lower frequency. Log-spiral components must be tightly-wound in order to excite responses at high radial harmonics.

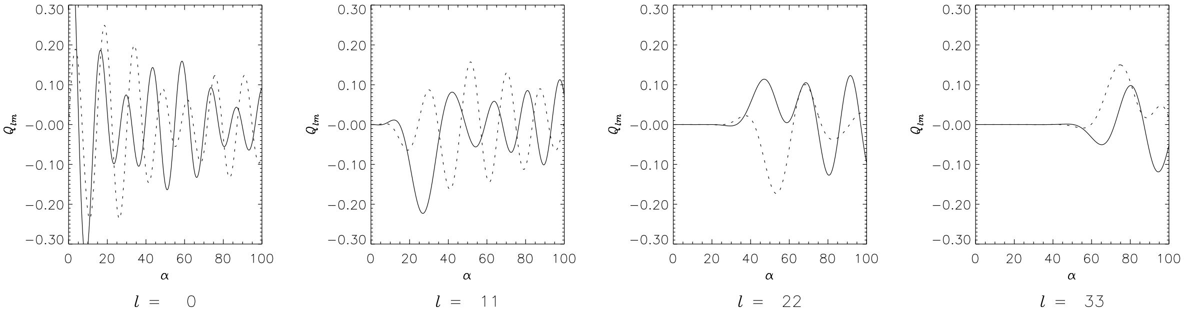


Figure 5.3: Graph of Q_{lm} plotted against α, for four different radial harmonics l. In each case, the solid line is the real, and the dotted line the imaginary, part of Q_{lm}. [$\beta = 0.25$, $m = 2$, $\tilde{U} = 0.5$; $a_{\rm acc} = 100$, $b_{\rm acc} = 2.5$.]

5.3 The transfer function

The summation over Fourier harmonics

The expression for the transfer function (3.79) involves a sum over l from $-\infty$ to $+\infty$, where l is the radial harmonic number. Let us write the summand as S, where $S = Q_{lm}(\alpha')\, Q^*_{lm}(\alpha)\, F_{lm}(\alpha - \alpha')$.

Fortunately, the magnitude of the terms in this sum decreases sharply with $|l|$, as the following figure demonstrates.

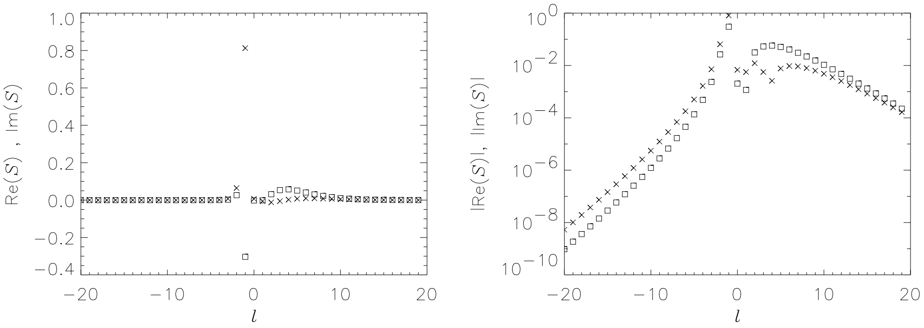


Figure 5.4: The figure shows the summand S plotted against the radial harmonic number l. The crosses show the real part of the function, and the squares the imaginary part (see text). [$\beta = 0.25$, $N = 2$, $M = 6$, $\tilde{L}_z = 100$; $m = 2$, $\Omega_\mathrm{p} = 0.5$, $s = 0.3$, $\gamma = 11.0$; $\tilde{U} = 1.0$, $\alpha' = 1.5$, $\alpha = 1$.]

The left-hand plot shows the real and imaginary parts of S on linear axes. The right-hand plot shows the magnitudes of $\mathrm{Re}(S)$ and $\mathrm{Im}(S)$ plotted with a logarithmic vertical axis, so as to demonstrate the order of magnitude. Clearly only the first few values of $|l|$ contribute to the sum. Here the summand falls by an order of magnitude when $|l|$ exceeds 3 or so. We also see that negative values of l decrease their contribution faster than positive l. This turns out to be true in general. A good approximation to the sum is given by summing from $l = l_\mathrm{min}$ to $l = l_\mathrm{max}$, where l_min is negative and $|l_\mathrm{min}|$ is less than l_max. Values of $l_\mathrm{min} = -20$, $l_\mathrm{max} = 30$ usually worked well (see below, section 5.6).

The integration over orbit shape

$\mathcal{S}_m$ is an integral over $\tilde{U}$, which adds up orbits of all possible shapes. We have (3.79):

$$\mathcal{S}_m(\alpha, \alpha') \propto \int d\tilde{U}\, \mathcal{J}_0(\tilde{U})\tilde{U} \left| \frac{\beta\tilde{U}^2 + \beta - 2}{2\beta} \right|^{1/\beta + \gamma/\beta - \gamma/2} \sum_{l=-\infty}^{+\infty} Q_{lm}(\alpha')\, Q^*_{lm}(\alpha)\, F_{lm}(\alpha - \alpha') \tag{5.3}$$

The equivalent expression for the Toomre-Zang disk (3.80) contains a factor $\exp\big(-\tilde{U}^2/(2\tilde{\sigma}_u^2)\big)$, in place of the term in modulus bars above. This prompted Zang (1976) to use Gauss-Laguerre quadrature to evaluate $\mathcal{S}_m$. The fundamental formula is

$$\int_0^\infty f(x) e^{-x} dx \approx \sum_{i=1}^{n_{GL}} w_i f(x_i), \tag{5.4}$$

where $f(x)$ is a smooth function, well approximated by a polynomial. The weights w_i and abscissae x_i are well-known (see Abramowitz & Stegun 1989; Press *et al.* 1989, ch. 4, sec. 5). Gauss-Laguerre quadrature with n_{GL} points is exact when $f(x)$ is a polynomial of order less than or equal to $2n_{\mathrm{GL}}-1$. This method suggests the substitution $V=\frac{1}{2}\tilde{U}^2/\tilde{\sigma}_u^2$. However, Zang (1976) found it preferable to reduce $\tilde{\sigma}_u$ in this expression, replacing it with $\tilde{\sigma}_n = f_\sigma \tilde{\sigma}_u$, where the fraction f_σ is around 80%. This concentrates attention on the lower values of $\tilde{U}$, where the integrand is larger. For the power-law disks, the distribution of radial velocities at any spot is similar – but not exactly equal to – a Gaussian with dispersion $\tilde{\sigma}_u$. This suggests that Gauss-Laguerre quadrature may still work well. Although our integrand is not directly of form (5.4), it is readily made so. The limits on our integral are zero and infinity for negative β. For positive β, the upper limit is $\tilde{U}=(2/\beta-1)^{1/2}$. At least for the models of most interest in galactic astronomy ($|\beta| \lesssim 0.5$), this distinction is in practice unimportant, since the integrand has fallen to negligibly small values well before this limit on the eccentric velocity.

Let $\mathcal{I}_U$ be the integrand in $\mathcal{S}_m$ when the integration is carried out over $\tilde{U}$, so $\mathcal{S}_m(\alpha,\alpha') \propto \int_0^\infty \mathcal{I}_U d\tilde{U}$. We then have

$$\mathcal{I}_V = \mathcal{I}_U \frac{\tilde{\sigma}_n^2}{\tilde{U}} \exp\left(\frac{\tilde{U}^2}{2\tilde{\sigma}_n^2}\right), \tag{5.5}$$

where $\mathcal{S}_m(\alpha,\alpha') \propto \int_0^\infty \mathcal{I}_V e^{-V} dV$ and $V=\frac{1}{2}\tilde{U}^2/\tilde{\sigma}_n^2$. Fig. 5.5 compares the integrands $\mathcal{I}_U$ and $\mathcal{I}_V$, for different values of f_σ. The vertical dashed lines indicate the positions of the abscissae. In section 5.6, we consider the effect on the mathematical eigenvalues of different values of f_σ.

The reliability of Gauss-Laguerre quadrature can be tested by comparison with the extended midpoint method (Press *et al.* 1989, eq. (4.1.19)). The following results are for the example illustrated in fig. 5.5, with the same perturbation and numerical accuracy parameters. In the Gauss-Laguerre integration, f_σ was 0.8; in the midpoint integration, the lower and upper limits on $\tilde{U}$ were 10^{-5} and 2.5 respectively.

Integration method	Number of evaluations	$\mathcal{S}_m(1.0,1.5)$
Gauss-Laguerre	2	$0.0415657-0.0133073i$
	3	$0.0439962-0.0125315i$
	4	$0.0439962-0.0124946i$
	5	$0.0439941-0.0124965i$
	6	$0.0439942-0.0124965i$
	7	$0.0439942-0.0124965i$
	8	$0.0439942-0.0124965i$
	9	$0.0439942-0.0124965i$
Extended midpoint	10	$0.0418484-0.0116520i$
	20	$0.0435345-0.0123236i$
	50	$0.0439235-0.0124702i$
	100	$0.0439766-0.0124899i$
	200	$0.0439898-0.0124948i$
	500	$0.0439935-0.0124962i$
	1000	$0.0439940-0.0124964i$
	5000	$0.0439942-0.0124965i$

Gauss-Laguerre quadrature is a remarkably efficient way of performing the integral. 6 s.f. accuracy is obtained with only six function evaluations, whereas over a thousand evaluations are needed for equivalent accuracy with the midpoint method!

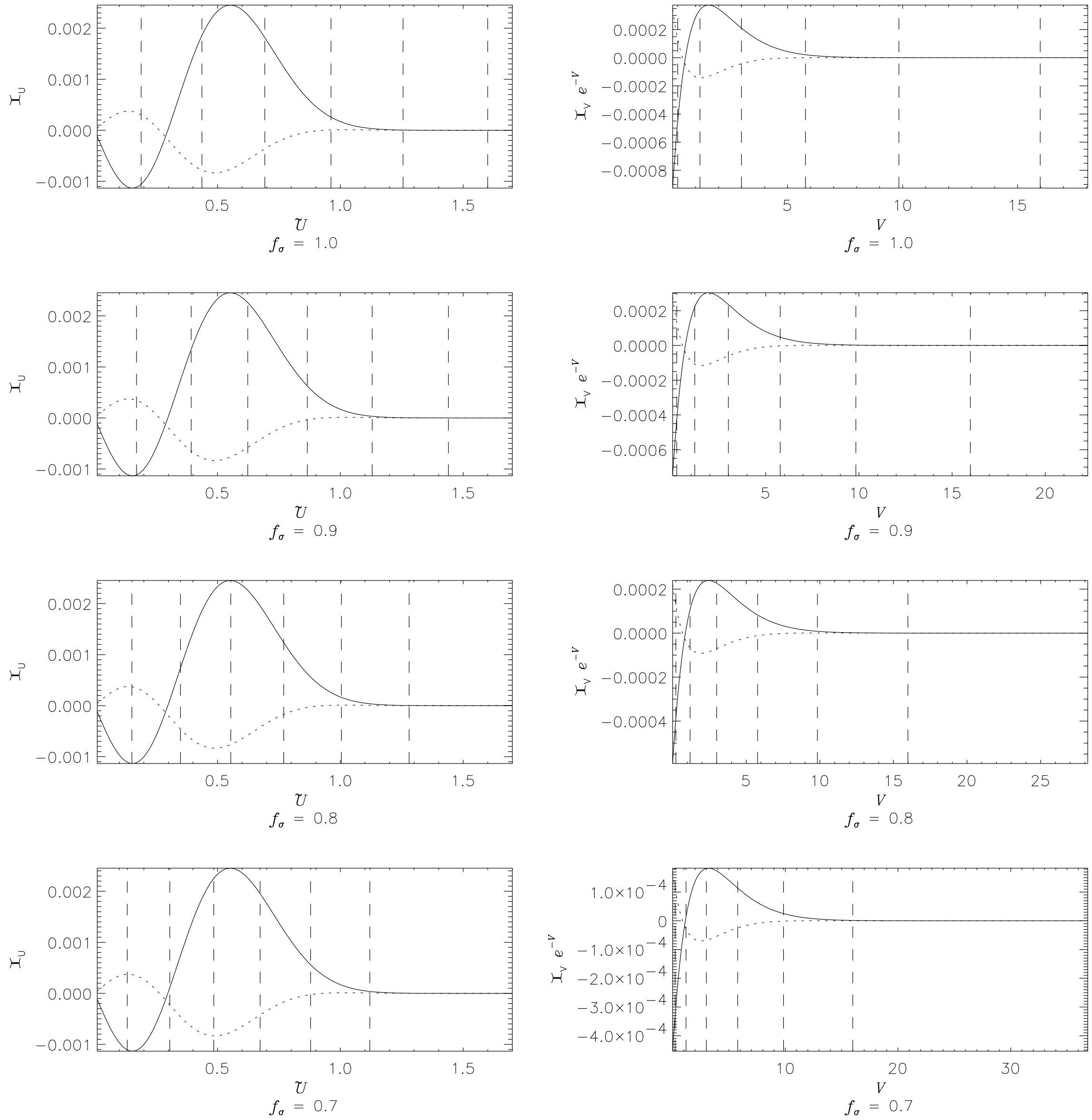


Figure 5.5: The effect of the transformation from $\tilde{U}$ to V on the integrand of the transfer function. The four different rows show different values of f_σ: $f_\sigma = 1.0$, 0.9, 0.8 and 0.7. In each case $n_{\rm GL} = 6$. The four plots in the left-hand column show $\mathcal{I}_U$ plotted against $\tilde{U}$, with the six Gauss-Laguerre abscissae marked by vertical dashed lines. The four plots in the right-hand column show the corresponding $\mathcal{I}_V e^{-V}$ plotted against V. The abscissae are marked with vertical dotted lines (note that they occur at the same value of V in each plot). In each case, the solid line represents the real part, and the dotted line the imaginary part. [$\beta = 0.25$, $N = 2$, $M = 6$, $\tilde{L}_{\rm c} = 100$; $m = 2$, $\Omega_{\rm p} = 0.50$, $s = 0.30$, $\gamma = 11.0$; $\alpha' = 1.5$, $\alpha = 1.0$. Numerical accuracy parameters: $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 6$, $f_\sigma = 0.8$, $a_{\rm acc} = 100$, $b_{\rm acc} = 2$.]

5.4 The response function

In section 4.3, we derived two alternative expressions for the function $\mathcal{R}$ describing the response of the self-consistent disk to neutral disturbances. First, we used the integral equation to obtain expressions (eq. (4.12) for non-axisymmetric modes, and eq. (4.40) for $m = 0$) involving integration over eccentric velocity and an expansion in Fourier harmonics. We also derived alternative expressions (4.18), (4.26) containing a double integral over radial and tangential velocity. We demonstrated analytically that these expressions are equivalent. Checking this numerically provides an additional test of our methods. The following table compares the results for a disk with $\beta = 0.25$. In each case, the upper result in a cell is that obtained by using the midpoint method to perform the double integration over u and v in (4.26). The lower result is that obtained by Gauss-Laguerre quadrature over $\tilde{U}$ as in (4.28). [Numerical accuracy parameters: With the midpoint method, the integrals over u and v were both evaluated to an accuracy of 10^{-8} for $\gamma > 0$. For $\gamma = 0$, the integrals were slow to converge, and were evaluated to an accuracy of 10^{-6}. The integral over t was evaluated using 50 function evaluations. With the Gaussian quadrature, $n_{\mathrm{GL}} = 16$, $l_{\min} = 70$, $l_{\max} = 80$, $\sigma_f = 0.8$, $a_{\mathrm{acc}} = 50$, $b_{\mathrm{acc}} = 2.5$.]

	Wavenumber α				
Temperature	0	10	20	50	100
$\tilde{\sigma}_u = 0.10$, $\gamma = 98.5$	0.437706 0.437706	2.72163 2.72163	2.49766 2.49766	1.24795 1.24795	0.661467 0.687114
$\tilde{\sigma}_u = 0.20$, $\gamma = 23.5$	0.437340 0.437340	1.23916 1.23916	0.750007 0.750007	0.328163 0.339935	0.168698 0.171431
$\tilde{\sigma}_u = 0.30$, $\gamma = 9.61$	0.436690 0.436690	0.622006 0.622006	0.346208 0.346172	0.146671 0.146245	0.074693 0.073160
$\tilde{\sigma}_u = 0.40$, $\gamma = 4.75$	0.435690 0.435690	0.363012 0.362970	0.195862 0.196583	0.081753 0.081337	0.041445 0.037916
$\tilde{\sigma}_u = 0.50$, $\gamma = 2.50$	0.434244 0.434212	0.233902 0.233494	0.124121 0.125953	0.051367 0.048451	0.025971 0.021383
$\tilde{\sigma}_u = 0.82$, $\gamma = 0.00$	0.425064 0.414554	0.081580 0.076716	0.042261 0.035246	0.017262 0.012428	0.008694 0.005122

The agreement is in general excellent. The results start to diverge only for large α and $\tilde{\sigma}_u$. For large $\tilde{\sigma}_u$, the range of eccentric velocities covered by the Gauss-Laguerre integration is much greater. This means that fewer values are sampled in the region of low eccentric velocity where the integrand is largest. This is probably responsible for the discrepancy at high $\tilde{\sigma}_u$.

We now investigate the convergence of the response function as these accuracy parameters are varied. We consider an $m = 2$ perturbation in a disk with $\beta = 0.25$. To begin with, we consider low temperature, $Q_{\mathrm{s}} = 0.3$, and low wavenumber $\alpha = 10$. $\sigma_f = 0.8$.

$m = 2$, $\beta = 0.25$, $Q_{\mathrm{s}} = 0.3$, $\alpha = 10$				
	a_{acc},	$l_{\min}$, $l_{\max}$		
n_{GL}	b_{acc}	-10,20	-20,30	-50,70
6	10,1.5	1.28185	10.62131	42.9951
	20,2.5	-1.45862	-1.45862	14.0694
	50,2.5	-1.45862	-1.45862	-1.45862
8	10,1.5	0.715152	9.97420	41.4265
	20,2.5	-1.45862	-1.45862	13.2438
	50,2.5	-1.45862	-1.45862	-1.45862
10	10,1.5	0.342021	9.51510	39.9970
	20,2.5	-1.45862	-1.45862	12.8208
	50,2.5	-1.45862	-1.45862	-1.45862

For this low temperature, many radial harmonics are not needed. If n_ψ is too low, the terms with higher eccentric velocities and higher radial harmonics are not evaluated accurately. Thus when $a_{\mathrm{acc}} = 10$, $b_{\mathrm{acc}} = 1.5$, different answers are obtained for different orders of Gauss-Laguerre quadrature and different ranges in l. However, when a_{acc} and b_{acc} are large enough, increasing the order of Gauss-Laguerre quadrature and the range in l has no effect, indicating that the eigenvalue has converged.

At higher temperatures, the convergence is worse. We now consider $Q_{\mathrm{s}} = 1.0$, at the same wavenumber $\alpha = 10$.

$m = 2$, $\beta = 0.25$, $Q_{\mathrm{s}} = 1.0$, $\alpha = 10$				
	a_{acc},	l_{min}, l_{max}		
n_{GL}	b_{acc}	-10,20	-20,30	-50,70
6	10,1.5	0.513928	0.853610	2.99424
	20,2.5	0.513925	0.515022	0.766008
	50,2.5	0.513925	0.515022	0.515059
8	10,1.5	0.513548	0.857427	2.81961
	20,2.5	0.513549	0.514623	0.711952
	50,2.5	0.513549	0.514623	0.514665
10	10,1.5	0.515026	0.810299	2.86831
	20,2.5	0.513549	0.514617	0.679567
	50,2.5	0.513549	0.514617	0.514660

At these higher temperatures, more stars are on highly eccentric orbits. Thus more points are required in the orbit integration, more Gauss-Laguerre integration points and a higher range of radial harmonics are needed.

When both the temperature and the wavenumber are high, the convergence is even worse. For $Q_{\mathrm{s}} = 1.0$, and $\alpha = 30$, the accuracy is essentially 1 d.p. The following table includes results from high-order Gauss-Laguerre quadrature, going up to high eccentric velocities. For high n_{GL} and high a_{acc} and b_{acc}, n_ψ (usually determined by (5.2)) is truncated at a particular value, n_{max}. If n_{max} is set too low, then the contribution of the last few terms in the Gauss-Laguerre sum, which should be tiny, tend to be overestimated. We found that it was better to set the contribution from these points to zero, rather than risk overestimating them by using too few points in the orbit integration. This method was employed in the results shown here, for which n_{max} was 10^6.

$m = 1$, $\beta = 0.25$, $Q_{\mathrm{s}} = 1.0$, $\alpha = 30$					
	a_{acc},	l_{min}, l_{max}			
n_{GL}	b_{acc}	-10,20	-20,30	-50,70	-100,200
6	20,2.5	0.287788	0.310521	0.399574	1.04994
	50,2.5	0.287788	0.310521	0.314542	0.400806
	100,3.0	0.287788	0.310521	0.314542	0.314576
8	20,2.5	0.288362	0.312204	0.384100	0.984106
	50,2.5	0.288362	0.312204	0.315533	0.383466
	100,3.0	0.288362	0.312204	0.315533	0.315586
10	20,2.5	0.273070	0.293682	0.370840	0.919548
	50,2.5	0.273070	0.293682	0.297401	0.414498
	100,3.0	0.273070	0.293682	0.297401	0.297455
12	20,2.5	0.224548	0.248162	0.337413	0.841474
	50,2.5	0.224548	0.248162	0.251657	0.382475
	100,3.0	0.224548	0.248162	0.251657	0.251701

The lesson of these investigations is that the response function is calculated reliably at low temperatures, but that convergence is much worse at high temperatures (and, to a lesser extent, high wavenumbers). In this thesis, we shall mainly be interested in the temperature $\tilde{\sigma}_{u,\mathrm{min}}$ at which neutral modes with wavenumber α_{u} become possible. Fortunately, both $\tilde{\sigma}_{u,\mathrm{min}}$ and α_{u} turn out to be low enough that $\mathcal{R}$ can be accurately evaluated. The canonical accuracy parameters adopted were $n_{\mathrm{GL}} = 12$, $l_{\mathrm{min}} = -20$, $l_{\mathrm{max}} = -30$, $a_{\mathrm{acc}} = 50$, $b_{\mathrm{acc}} = 2.5$. The response curves obtained for different m are investigated in later chapters.

5.5 Computational solution of the integral equation

Given that the kernel of the integral equation can be evaluated to high accuracy, we now need to devise a method for its numerical solution. We ultimately seek self-consistent solutions, for which $A_{\mathrm{res}}(\alpha) = A_{\mathrm{imp}}(\alpha)$. However, the first step is to consider the more general problem for which the response density is a complex multiple of the imposed density, i.e., $A_{\mathrm{res}}(\alpha) = \lambda A_{\mathrm{imp}}(\alpha)$. This casts the integral equation (3.45) into the form

$$\lambda A\left(\alpha\right) = \int_{-\infty}^{+\infty} d\alpha' A\left(\alpha'\right) \mathcal{S}_m(\alpha, \alpha'), \tag{5.6}$$

where we have dropped the subscripts on A. We shall refer to λ as the mathematical eigenvalue. Of course, only the instance when λ is unity carries the physical significance of a mode. The advantage of this mathematical artifice is that the integral equation (5.6) is in the standard form of a homogeneous, linear, Fredholm equation of the second kind (see e.g. Courant & Hilbert 1953; Delves & Mohamed 1985). For a given value of λ, such an equation normally admits only the trivial solution, $A(\alpha) \equiv 0$. The values of λ for which non-trivial solutions exist are the eigenvalues of the equation. We seek an iterative scheme which drives the eigenvalue to unity, thus providing a self-consistent mode.

There is a close analogy between linear algebraic equations and linear integral equations. The former define relations between vectors in a finite-dimensional vector space, the latter define relations between functions in an infinite-dimensional vector space (technically a Banach space). This analogy can be made explicit by applying a quadrature rule to the integration in (5.6) to obtain

$$\lambda A\left(\alpha_j\right) = \sum_i w_i A\left(\alpha_i\right) \mathcal{S}_m(\alpha_j, \alpha_i), \tag{5.7}$$

where w_i are some appropriate weights. This general approach is called the Nystrom method (Delves & Mohamed 1985). It evidently reduces the solution of an integral equation to the solution of an algebraic eigenvalue problem. The latter is a classic and well-studied area of numerical analysis, for which many tried and tested techniques are available.

Much of the skill in the numerical solution of integral equations comes from the choice of quadrature rules and weights. Fortunately, we can rely here on the work of Zang (1976), who devised an elegant method based on locally approximating the kernel and the response by Lagrangian interpolating polynomials. This is naturally adapted to the instance when the kernel varies on a much smaller scale than the solution. First, to obtain smoother functions, the Kalnajs gravity factor is extracted by defining

$$\mathcal{S}_m\left(\alpha,\alpha'\right)=K\left(\alpha',m\right)\tilde{\mathcal{S}}_m\left(\alpha,\alpha'\right), \qquad\qquad A\left(\alpha\right)=\frac{\tilde{A}(\alpha)}{K\left(\alpha,m\right)}. \tag{5.8}$$

The eigenvalue equation now becomes

$$\lambda\tilde{A}\left(\alpha\right)=K\left(\alpha,m\right)\int_{-\infty}^{+\infty}d\alpha'\tilde{A}\left(\alpha'\right)\tilde{\mathcal{S}}_m\left(\alpha,\alpha'\right). \tag{5.9}$$

Let us introduce a finite grid of points in wavenumber space, α_r. If we need to know the value of $\tilde{A}$ or $\tilde{\mathcal{S}}_m$ at a value of α intermediate between the gridpoints α_r and α_{r+1}, we interpolate over the eight gridpoints from α_{r-3} to α_{r+4}. This is illustrated in fig. 5.6.

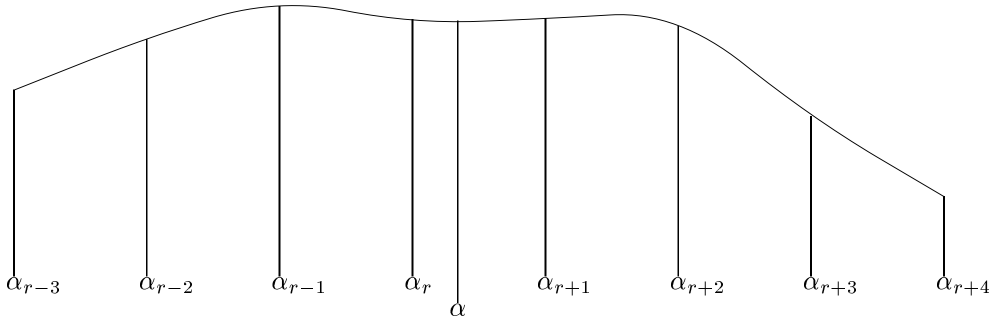


Figure 5.6: The interpolation over eight points in wavenumber space.

Lagrange's classic formula for the interpolating polynomial $P(\alpha)$ through N points $f(\alpha_k)$ is (Press *et al.* 1989, ch. 3):

$$P(\alpha)=\sum_{k=1}^{N}\frac{(\alpha_k-\alpha_k)\prod_{j=1}^{N}(\alpha-\alpha_j)}{(\alpha-\alpha_k)\prod_{i=1}^{N}(\alpha_k-\alpha_i)}f(\alpha_k). \tag{5.10}$$

For 8 equally-spaced points $\Delta\alpha$ apart, going from α_{r-3} to α_{r+4}, this becomes

$$P(\alpha)=\prod_{i=1}^{8}\left(\alpha-\alpha_{r+i-4}\right)\sum_{k=-3}^{4}\frac{(-1)^k}{(3+k)!(4-k)!}\frac{f(\alpha_{r+k})}{(\Delta\alpha)^7(\alpha-\alpha_r-k\Delta\alpha)}. \tag{5.11}$$

Defining $x=(\alpha-\alpha_r)/\Delta\alpha$, this becomes

$$P(\alpha)=\prod_{i=1}^{8}\left(x+4-i\right)\sum_{k=-3}^{4}(-1)^k\frac{f(\alpha_{r+k})}{(3+k)!(4-k)!(x-k)}=\sum_{k=-3}^{4}L_k[x]f(\alpha_{r+k}), \tag{5.12}$$

where

$$L_k[x]=(-1)^k\frac{\prod_{i=1}^{8}\left(x+4-i\right)}{(3+k)!(4-k)!(x-k)}. \tag{5.13}$$

Then the interpolated approximations to the response and the transfer function are (cf. Zang 1976, app. D)

$$\tilde{A}(\alpha') \approx \sum_{k=-3}^{4} L_k[x'] \tilde{A}(\alpha_{r+k}), \qquad \tilde{\mathcal{S}}_m(\alpha, \alpha') \approx \sum_{k=-3}^{4} L_k[x'] \tilde{\mathcal{S}}_m(\alpha, \alpha_{r+k}), \tag{5.14}$$

where $x' = (\alpha' - \alpha_r)/\Delta\alpha$.

The infinite range in the integral equation 5.6 is a kind of singularity. Truncation of the wavenumber range at large finite values is the simplest but most brutal way of handling the singularity. In practice, we found this to be surprisingly effective. Wavenumber space is approximated by a finite grid with n points along each side. We choose n, and the grid-point spacing $\Delta\alpha$, so as to obtain a sufficiently accurate solution of the integral equation (3.45). The finite size of the grid means that we encounter problems when evaluating $\tilde{A}_{\text{imp}}$ and $\tilde{\mathcal{S}}_m$ for values of α near the edges of the grid. We are interpolating over eight points, so we need data from α_{-2} to α_{n+3}. We shall deal with this problem by simply assuming that our function is zero outside the grid. This is illustrated in the following figure. Function evaluations at grid-points lying in the shaded region are set to zero.

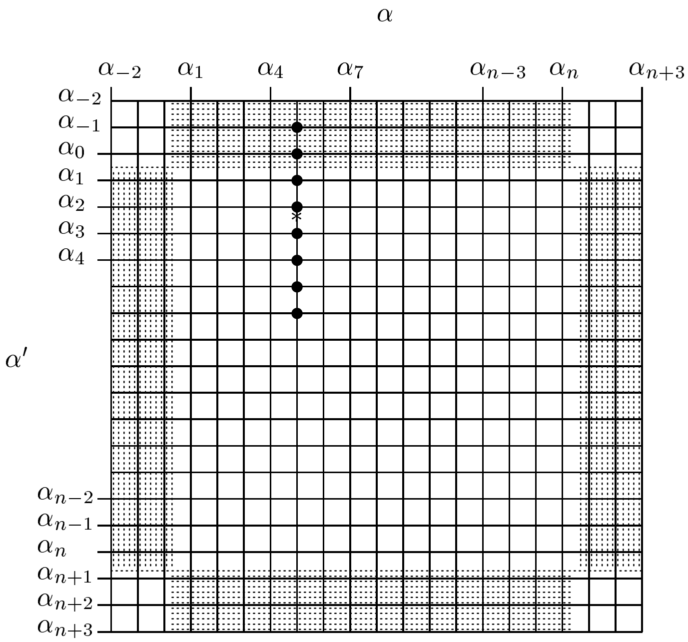


Figure 5.7: Interpolation in a finite grid. Suppose we wish to find the value of the function at the point indicated by an asterisk. Then $r = 2$, so we interpolate over the eight values from α_{-1} to α_6. The values of the function at α_{-1} and α_0 are set to be zero.

This is acceptable provided the grid is large enough. Then the values of the kernel at the missing interpolation points are negligibly small. This procedure is justified empirically by the demonstration that the mathematical eigenvalue converges to a value independent of grid size (see section 5.6).

We are now in a position to carry out the integration. Substituting the Lagrange-interpolated approximations for $\tilde{A}_{\mathrm{imp}}$ and $\tilde{\mathcal{S}}_m$ (5.14) into the modified integral equation (5.9), we obtain

$$\lambda \tilde{A}\left(\alpha\right) = K\left(\alpha, m\right) \sum_{i=-3}^{4} \sum_{k=-3}^{4} \int_{\alpha_1}^{\alpha_n} d\alpha' L_i \left[\frac{\alpha' - \alpha_r}{\Delta\alpha}\right] L_k \left[\frac{\alpha' - \alpha_r}{\Delta\alpha}\right] \tilde{A}(\alpha'_{r+i}) \tilde{\mathcal{S}}_m(\alpha, \alpha'_{r+k}). \tag{5.15}$$

Note that the integration over wavenumber now runs from $\alpha = \alpha_1$ to α_n, instead of from $\alpha = -\infty$ to ∞. We break this integration into n portions of $\Delta\alpha$.

$$\begin{aligned} \lambda \tilde{A}\left(\alpha\right) = K\left(\alpha, m\right) &\sum_{r=1}^{n} \sum_{i=-3}^{4} \sum_{k=-3}^{4} \\ \times &\int_{\alpha_r}^{\alpha_r + \Delta\alpha} d\alpha' L_i \left[\frac{\alpha' - \alpha_r}{\Delta\alpha}\right] L_k \left[\frac{\alpha' - \alpha_r}{\Delta\alpha}\right] \tilde{A}(\alpha'_{r+i}) \tilde{\mathcal{S}}_m(\alpha, \alpha'_{r+k}). \end{aligned} \tag{5.16}$$

Then, changing variables to $x' = (\alpha' - \alpha_r)/\Delta\alpha$, we obtain

$$\lambda \tilde{A}\left(\alpha\right) = K\left(\alpha, m\right) \sum_{r=1}^{n} \sum_{i=-3}^{4} \sum_{k=-3}^{4} \tilde{A}(\alpha_{r+i}) \tilde{\mathcal{S}}_m(\alpha, \alpha_{r+k}) \Delta\alpha \int_0^1 dx' L_i[x'] L_k[x']. \tag{5.17}$$

Let us define the weighting coefficients C_{ik} by (cf. Zang (1976), app. D)

$$C_{ik} = \Delta\alpha \int_0^1 dx L_i[x] L_k[x]. \tag{5.18}$$

Note the symmetry properties of C_{ik}: $C_{ik} \equiv C_{ki}$, and $C_{ik} \equiv C_{(1-i)(1-k)}$. Thus although there appear to be 64 C_{ik}, only 20 of them are independent. The weighting coefficients are easily evaluated using the midpoint method given in Press *et al.* (1989, ch. 4). The following table shows values of $C_{ik}/\Delta\alpha$:

i	$k = -3$	$k = -2$	$k = -1$	$k = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
-3	0.000003	-0.000030	0.000152	-0.000938	-0.000890	0.000150	-0.000030	0.000003
-2	-0.000030	0.000296	-0.001499	0.009323	0.008653	-0.001472	0.000293	-0.000030
-1	0.000152	-0.001499	0.007619	-0.048444	-0.042688	0.007386	-0.001472	0.000150
0	-0.000938	0.009323	-0.048444	0.416790	0.223032	-0.042688	0.008653	-0.000890
1	-0.000890	0.008653	-0.042688	0.223032	0.416790	-0.048444	0.009323	-0.000938
2	0.000150	-0.001472	0.007386	-0.042688	-0.048444	0.007619	-0.001499	0.000152
3	-0.000030	0.000293	-0.001472	0.008653	0.009323	-0.001499	0.000296	-0.000030
4	0.000003	-0.000030	0.000150	-0.000890	-0.000938	0.000152	-0.000030	0.000003

Equation (5.15) can now be written as:

$$\lambda \tilde{A}\left(\alpha_j\right) = K\left(\alpha_j, m\right) \sum_{r=1}^{n} \sum_{i=-3}^{4} \sum_{k=-3}^{4} C_{ik} \tilde{\mathcal{S}}_m(\alpha_j, \alpha_{r+k}) \tilde{A}(\alpha_{r+i}) \tag{5.19}$$

Terms in this multiple sum for which $r + i$ is less than 1 or greater than n contribute nothing. We can thus write the above equation in terms of a single sum from α_1 to α_n, and collect the weighting coefficient C_{ik}, smoothed transfer function $\tilde{\mathcal{S}}_m$ and density profile $\tilde{A}$ together into a single quantity $\mathbf{S}$.

$$\lambda \tilde{A}(\alpha_j) = \sum_{s=1}^{n} S_{js} \tilde{A}(\alpha_s). \tag{5.20}$$

This is of course just a matrix equation for the mathematical eigenvalue λ. If there is an eigenvalue of the matrix $\mathbf{S}$ equal to unity, then the corresponding eigenvector $\mathbf{A}$ gives us a self-consistent mode, i.e. a self-sustaining density perturbation.

S_{js} cannot be simply expressed in terms of C_{ik}, $\tilde{\mathcal{S}}_m$ and $\tilde{A}$. But we see that it represents the total coefficient of $\tilde{A}(\alpha_s)$ in eq. (5.19). So we can calculate its elements by going through the multiple sum in (5.19), and, for each value of j, adding up all the terms for which $r+i=s$. The algorithm is as follows:

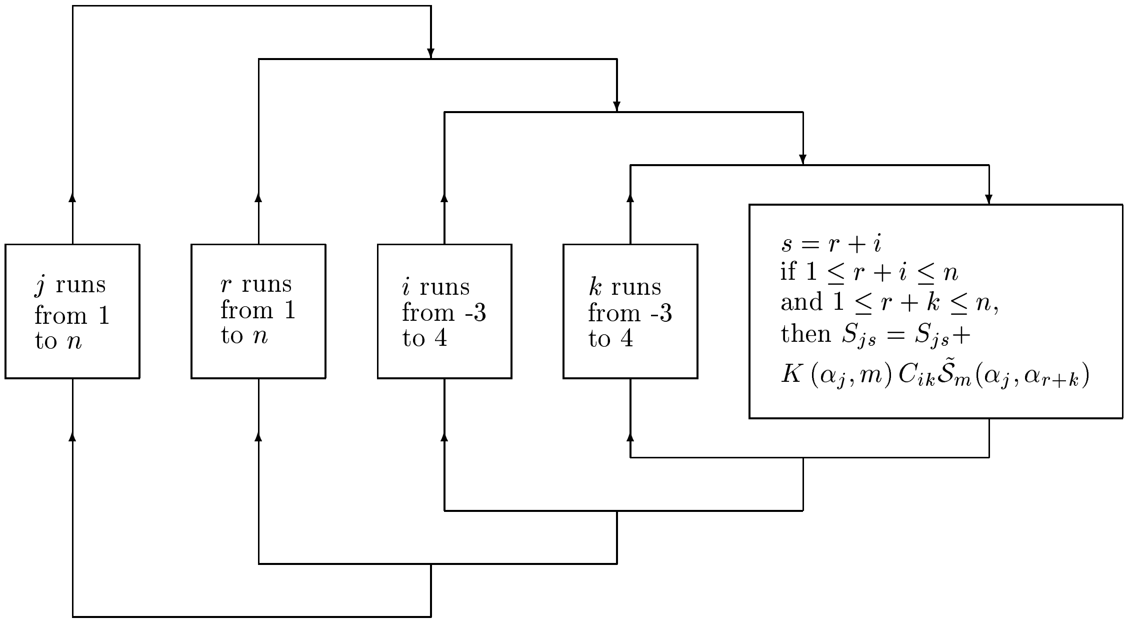


Figure 5.8: Diagram of the algorithm used to calculate the matrix elements

Having obtained the matrix S_{js}, we now wish to find its eigenvalues. We used the eigensystem package (EISPACK) developed by Smith *et al.* (1972), which returns the eigenvalues of a matrix, and, optionally, the corresponding eigenvectors.

EISPACK finds the eigenvalues by converting the matrix to triangular form. The eigenvalues are then given by the diagonal entries. Once the eigenvalues have been obtained, the eigenvectors can be calculated by substitution. The triangularisation process uses a theorem due to Schur, which states that any matrix may be triangularised by a unitary similarity transformation. This is a transformation of the form $\mathbf{A} \rightarrow \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$, in which the matrix $\mathbf{S}$ is unitary, i.e. $\mathbf{S}^{\dagger} = \mathbf{S}^{-1}$. Schur's theorem is implemented by applying a series of similarity transforms:

$$\mathbf{A}_{k+1} = \mathbf{S}_k^{-1}\mathbf{A}_k\mathbf{S}_k \tag{5.21}$$

The iteration is continued until all the subdiagonal elements of $\mathbf{A}_{k+1}$ are less than the round-off errors.

We also used the power method to provide an independent check on the largest eigenvalue. This relies on repeated application of the matrix $\mathbf{S}$ to a test vector $\mathbf{t}$ (Berezin & Zhidkov 1965, ch. 8, sec. 7). Suppose we apply the $n \times n$ matrix $\mathbf{S}$ to some vector $\mathbf{t}$. In general, the result can be expressed as a sum of the n eigenvectors $\mathbf{e}_i$

$$\mathbf{St} = \sum_{i=1}^{n} c_i \mathbf{e}_i. \tag{5.22}$$

Repeated application of the matrix $\mathbf{S}$ to this equation yields

$$\mathbf{S}^j\mathbf{t} = \sum_{i=1}^{n} c_i \lambda_i^{j-1} \mathbf{e}_i. \tag{5.23}$$

As $j \to \infty$, the sum will be dominated by the largest eigenvalue λ_1. Then for each of the components k, we have

$$\left(\mathbf{S}^j\mathbf{t}\right)_k \approx c_1 \lambda_1^{j-1} \left(e_1\right)_k . \tag{5.24}$$

We see that

$$\lambda_1 = \lim_{j\to\infty} \frac{\left(\mathbf{S}^j\mathbf{t}\right)_k}{\left(\mathbf{S}^{j-1}\mathbf{t}\right)_k}. \tag{5.25}$$

When implementing this numerically, we chose k to be the largest component of $\mathbf{S}^j\mathbf{t}$. The power method is a fast and accurate way of obtaining the largest eigenvalue. It typically agreed with EISPACK to 8 s.f. or so.

5.6 The accuracy parameters

In this chapter, various accuracy parameters have been introduced which control the result obtained for the eigenvalue. We shall now spend some time considering the effect of each of these. The accuracy parameters involved in the calculation of the mathematical eigenvalue are

Quantity affected	Accuracy parameter	What it controls
Mathematical eigenvalue λ	n	Number of points along each side of the grid
Mathematical eigenvalue λ	$\Delta\alpha$	Grid-spacing
Fourier components Q_{lm}	$a_{\rm acc}$, $b_{\rm acc}$	Number of steps in mid-point integration
Transfer function $\mathcal{S}_m$	$l_{\rm min}$, $l_{\rm max}$	$\sum_{l=l_{\rm min}}^{l_{\rm max}}$ approximates $\sum_{l=-\infty}^{+\infty}$
Transfer function $\mathcal{S}_m$	$n_{\rm GL}$	Number of Gauss-Laguerre abscissae
Transfer function $\mathcal{S}_m$	f_σ	Position of Gauss-Laguerre abscissae

These quantities are tabulated for reference in Appendix A. Further accuracy parameters are involved in the midpoint routine to find the frequencies $\tilde{\kappa}$, $\tilde{\Omega}$ and the weighting coefficients C_{ik}, and in the Newton-Raphson routine to find the extrema $\tilde{R}_{\min}$, $\tilde{R}_{\max}$ (see Chapter 2). However, it is easy to find these quantities to high accuracy, and so these integrations have virtually no effect on the eigenvalue.

To investigate the effect of these parameters, consider an $m = 2$ perturbation with $\Omega_{\rm p} = 0.5$ and $s = 0.3$ applied to a $\beta = 0.25$ disk with $N = 2$, $M = 6$, $\tilde{L}_{\rm c} = 100$ and $\gamma = 10.99$ ($\tilde{\sigma}_u = 0.283$). For this situation, we believe the correct answer is $0.363837 - 0.118126i$ to 6 s.f. We choose the canonical set of accuracy parameters to be: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = 20$, $l_{\max} = 30$, $n_{\rm GL} = 6$, $f_\sigma = 0.8$, $a_{\rm acc} = 10$, $b_{\rm acc} = 1.5$. We shall vary each of these in turn, while keeping the others constant.

Effect of grid-spacing (constant grid size)

First let us investigate the effect of grid-spacing on the mathematical eigenvalues. We shall do this while holding the range of the grid in α constant. As a first guess, we set the range of the grid $(n-1)\Delta\alpha$ equal to 25.

Grid-spacing $\Delta\alpha$	Number of grid points n	Largest mathematical eigenvalue
1.042	25	$0.364674 - 0.117240i$
0.500	51	$0.363873 - 0.118012i$
0.403	63	$0.363877 - 0.118004i$
0.298	85	$0.363877 - 0.118007i$
0.202	125	$0.363879 - 0.118003i$

The answers have converged to 5 s.f. or so. But is this the right answer, or is the range of α covered by the grid too small? To answer this, we increase the range of the grid to 50 in α.

Grid-spacing $\Delta\alpha$	Number of grid points n	Largest mathematical eigenvalue
1.000	51	$0.363590 - 0.117720i$
0.500	101	$0.363836 - 0.118132i$
0.403	125	$0.363836 - 0.118129i$
0.301	167	$0.363836 - 0.118129i$
0.200	251	$0.363836 - 0.118129i$

These have converged to a slightly different answer, so the range of 25 in α was clearly not large enough for 6 s.f. accuracy. To see whether the grid is big enough yet, we increase the range of the grid again, to 100.

Grid-spacing $\Delta\alpha$	Number of grid points n	Largest mathematical eigenvalue
1.000	101	$0.363590 - 0.117720i$
0.500	201	$0.363836 - 0.118132i$
0.400	251	$0.363836 - 0.118127i$
0.299	335	$0.363836 - 0.118133i$
0.200	501	$0.363836 - 0.118129i$

These answers have not changed from the previous set to within 6 s.f., so we conclude that a range in α of 50 is sufficient.

In general, the spacing of the gridpoints is much more important than the size of the grid, provided that this covers a sufficient range in α. We want to choose the minimum n possible, since large n greatly increases the running-time of the code. Typically, we take $n = 251$, $\Delta\alpha = 0.2$ as suitable values.

Effect of range in l

We now consider the range in l.

Range in l	Largest mathematical eigenvalue
-10 to 20	$0.363850 - 0.118173i$
-20 to 20	$0.363829 - 0.118147i$
-20 to 30	$0.363836 - 0.118129i$
-30 to 30	$0.363835 - 0.118129i$
-30 to 40	$0.363837 - 0.118126i$
-40 to 40	$0.363837 - 0.118126i$
-40 to 50	$0.363837 - 0.118126i$

We have found that summing over l from -20 to 30 is generally sufficient for 6 s.f. accuracy. It is worth pointing out that including a larger range of radial harmonics does not necessarily lead to higher accuracy. Harmonics with large l are in general very small, but need a large number of points in the orbit integration to be accurately evaluated. If $a_{\rm acc}$ and $b_{\rm acc}$ are set too low for the given range in l, an unfortunate effect may occur in which the Fourier coefficients for extreme l are vastly overestimated, and distort the answer.

Effect of number of Gauss-Laguerre abscissae

The effect of $n_{\rm GL}$, the number of Gauss-Laguerre abscissae, on the transfer function was considered in section 5.3. Here we look at the effect of $n_{\rm GL}$ on the mathematical eigenvalue.

$n_{\rm GL}$	Largest mathematical eigenvalue
3	$0.363822 - 0.118148i$
4	$0.363840 - 0.118126i$
5	$0.363838 - 0.118128i$
6	$0.363836 - 0.118129i$
7	$0.363837 - 0.118128i$
8	$0.363837 - 0.118129i$
9	$0.363837 - 0.118128i$
10	$0.363837 - 0.118125i$

The eigenvalue is not very sensitive to the number of Gauss-Laguerre abscissae; even $n_{\rm GL} = 3$ suffices for 4 s.f. accuracy. We usually use $n_{\rm GL} = 9$ to be sure of high accuracy.

Effect of $a_{\rm acc}$ and $b_{\rm acc}$

$a_{\rm acc}$ and $b_{\rm acc}$ control the number of steps n_ψ in the Runge-Kutta integration used in finding Q_{lm}: we set $n_\psi = a_{\rm acc} \exp(b_{\rm acc}\tilde{U})$. The effect of these factors on Q_{lm} was considered in 5.2; now we investigate their effect on the eigenvalue itself.

a_{acc}	b_{acc}	Largest mathematical eigenvalue
5	0	$0.363836 - 0.118129i$
10	0	$0.363836 - 0.118129i$
50	0	$0.363836 - 0.118129i$
5	1.5	$0.363836 - 0.118129i$
10	1.5	$0.363836 - 0.118129i$
50	1.5	$0.363836 - 0.118129i$
5	3.0	$0.363836 - 0.118129i$
10	3.0	$0.363836 - 0.118129i$
50	3.0	$0.363836 - 0.118129i$

The eigenvalue is not at all sensitive to n_ψ. In fact changes only appear in the 8th decimal place. Changes in b_{acc}, predictably, have a greater effect than changes in a_{acc}.

Effect of factor f_σ

Most of the other accuracy parameters have a clearly-defined effect on the accuracy. For example, larger n and smaller $\Delta\alpha$ increase the accuracy. It is not immediately obvious, however, what value of f_σ is best, nor do we expect the answer to converge as we change f_σ. In order to understand the effect of f_σ, the code was first run with very high accuracy: $n = 501$, $\Delta\alpha = 0.1$, $l_{\text{min}} = 30$, $l_{\text{max}} = 50$, $n_{\text{GL}} = 9$, $a_{\text{acc}} = 50$, $b_{\text{acc}} = 3.0$. For $f_\sigma = 0.8$, the largest mathematical eigenvalue is $0.3638369 - 0.1181266i$. For $f_\sigma = 1.0$, it is $0.3638369 - 0.1181267i$. We conclude that the mathematical eigenvalue is $0.363837 - 0.118127i$ to 6 s.f.

Running the code with less accuracy, we look for the value of f_σ which gives an answer closest to the true answer. Using our canonical set of accuracy parameters, $n = 125$, $\Delta\alpha = 0.2$, $l_{\text{min}} = 20$, $l_{\text{max}} = 30$, $n_{\text{GL}} = 6$, $a_{\text{acc}} = 10$, $b_{\text{acc}} = 1.5$, we find:

f_σ	Largest mathematical eigenvalue
1.0	$0.363827 - 0.118129i$
0.9	$0.363838 - 0.118128i$
0.8	$0.363836 - 0.118129i$
0.7	$0.363838 - 0.118127i$
0.6	$0.363674 - 0.118126i$

Clearly there is little to choose between $f_\sigma = 0.7$, 0.8 and 0.9. We use $f_\sigma = 0.8$.

We usually use relatively fast, low-accuracy parameters for the first few calculations, to locate the (1,0) eigenvalue roughly. Typical values might be $n = 125$, $\Delta\alpha = 0.4$, $l_{\text{min}} = 10$, $l_{\text{max}} = 20$, $n_{\text{GL}} = 6$, $f_\sigma = 0.8$, $a_{\text{acc}} = 10$, $b_{\text{acc}} = 1.5$. We then use slow, high-accuracy values when homing in on the exact position of the unit eigenvalue. Typical values might be $n = 251$, $\Delta\alpha = 0.2$, $l_{\text{min}} = 30$, $l_{\text{max}} = 40$, $n_{\text{GL}} = 9$, $f_\sigma = 0.8$, $a_{\text{acc}} = 20$, $b_{\text{acc}} = 2.5$.

5.7 The mode-finding algorithm

Having established that we can find the mathematical eigenvalue to good accuracy, we now consider how to locate the unit eigenvalues which correspond to self-consistent modes. The mathematical eigenvalue depends on several quantities. First, there are the parameters describing the disk to which the perturbation is applied. The disk parameter β, the cut-out indices N and M, and the core angular momentum $\tilde{L}_{\rm c}$ together determine the active surface density, while the anisotropy parameter γ or radial velocity dispersion $\tilde{\sigma}_u$ describes the temperature of the disk. The pattern applied to that disk is described by its rotational symmetry m, its growth rate s and its pattern speed $\Omega_{\rm p}$.

We shall investigate modes of a given azimuthal symmetry in a given disk, meaning that m, β, N, M and $\tilde{L}_{\rm c}$ are held fixed. This leaves us free to adjust γ, s and $\Omega_{\rm p}$. A disk at a given temperature, if it admits a mode at all, will do so only for a particular growth rate and pattern speed. We may therefore hold γ fixed, and adjust s and $\Omega_{\rm p}$ until a mode is found. This is in fact a search in only one dimension, since the mathematical eigenvalue λ depends on growth rate and pattern speed only through the complex frequency $\omega = m\Omega_{\rm p} + is$. We use the Newton-Raphson method in one dimension to find such modes.

When the disk is sufficiently hot, it is completely stable. As the disk is cooled, instabilities set in through the marginal modes, for which the growth rate s is zero. We are therefore often interested in the marginally stable modes. To find these, we set s to some vanishingly small value, and perform a two-dimensional search in $\Omega_{\rm p}$ and γ. The diagram in fig. 5.9 shows curves of constant $\Omega_{\rm p}$ and constant γ, at a fixed value of s.

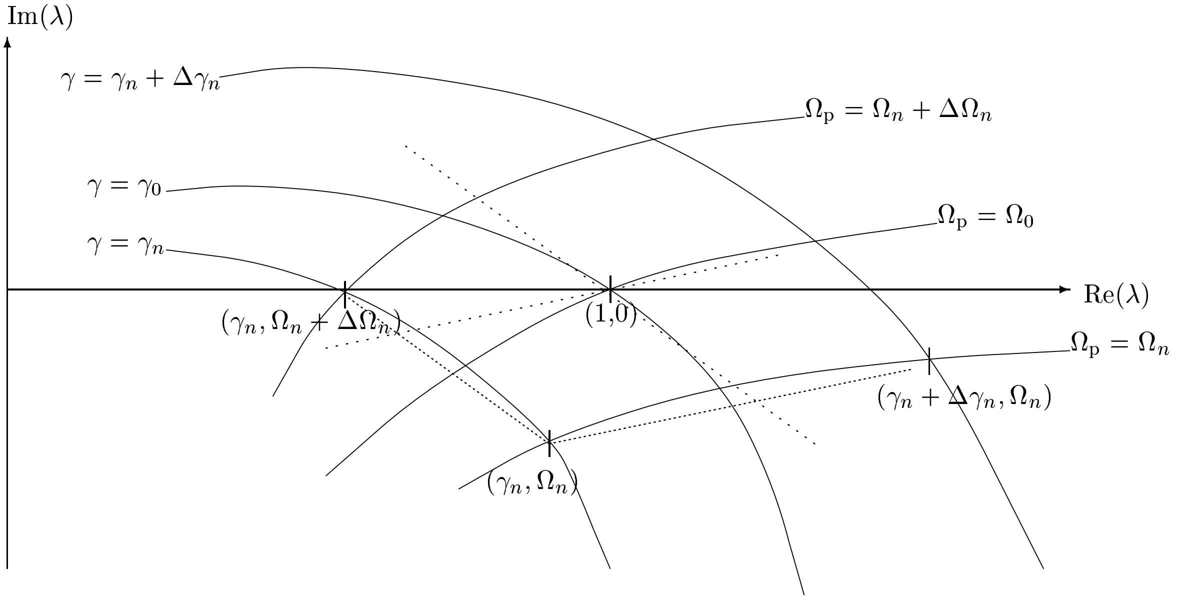


Figure 5.9: The mode-finding algorithm. Solid lines show curves of constant temperature and pattern speed in the plane of the mathematical eigenvalue.

We used the Newton-Raphson method in two dimensions. Suppose that the solution $\lambda = 1$ is at $\gamma = \gamma_0$ and $\Omega_{\rm p} = \Omega_0$. Suppose we have an approximation to the solution: (γ_n, Ω_n). We evaluate the eigenvalue for these values of γ and $\Omega_{\rm p}$, and obtain some value λ, which differs from unity. In order to probe how λ depends on γ and $\Omega_{\rm p}$, the eigenvalue is next evaluated at two further points: $(\gamma_n + \Delta\gamma_n, \Omega_n)$ and $(\gamma, \Omega_n + \Delta\Omega_n)$. To obtain the next approximation to the solution, $\gamma_{n+1}, \Omega_{n+1}$, we pretend that λ depends linearly on γ and $\Omega_{\rm p}$. The dotted lines in fig. 5.9 indicate the corresponding approximations to the curves of constant γ and Ω_p. Under this approximation,

$$\begin{aligned}\lambda(\gamma_{n+1},\Omega_{n+1}) = \lambda(\gamma_n,\Omega_n) &- \frac{\gamma_{n+1}-\gamma_n}{\Delta\gamma_n}\left[\lambda(\gamma_n,\Omega_n)-\lambda(\gamma_n+\Delta\gamma_n,\Omega_n)\right] \\ &- \frac{\Omega_{n+1}-\Omega_n}{\Delta\Omega_n}\left[\lambda(\gamma_n,\Omega_n)-\lambda(\gamma_n,\Omega_n+\Delta\Omega_n)\right].\end{aligned} \tag{5.26}$$

Requiring $\lambda(\gamma_{n+1},\Omega_{n+1}) = 1$ in this equation, and comparing real and imaginary parts, we obtain expressions for γ_{n+1} and Ω_{n+1}:

$$\gamma_{n+1} = \gamma_n + \Delta\gamma_{n+1}, \qquad \Omega_{n+1} = \Omega_n + \Delta\Omega_{n+1}, \tag{5.27}$$

where

$$\begin{aligned}\Delta\gamma_{n+1} = -\frac{\Delta\gamma_n}{\Delta}\{&[1-\lambda_\mathrm{R}(\gamma_n,\Omega_n)]\left[\lambda_\mathrm{I}(\gamma_n,\Omega_n)-\lambda_\mathrm{I}(\gamma_n,\Omega_n+\Delta\Omega_n)\right] \\ &+\lambda_\mathrm{I}(\gamma_n,\Omega_n)\left[\lambda_\mathrm{R}(\gamma_n,\Omega_n)-\lambda_\mathrm{R}(\gamma_n,\Omega_n+\Delta\Omega_n)\right]\},\end{aligned} \tag{5.28}$$

and

$$\begin{aligned}\Delta\Omega_{n+1} = \frac{\Delta\Omega_n}{\Delta}\{&[1-\lambda_\mathrm{R}(\gamma_n,\Omega_n)]\left[\lambda_\mathrm{I}(\gamma_n,\Omega_n)-\lambda_\mathrm{I}(\gamma_n+\Delta\gamma_n,\Omega_n)\right] \\ &+\lambda_\mathrm{I}(\gamma_n,\Omega_n)\left[\lambda_\mathrm{R}(\gamma_n,\Omega_n)-\lambda_\mathrm{R}(\gamma_n+\Delta\gamma_n,\Omega_n)\right]\},\end{aligned} \tag{5.29}$$

where

$$\begin{aligned}\Delta = &\left[\lambda_\mathrm{R}(\gamma_n,\Omega_n)-\lambda_\mathrm{R}(\gamma_n+\Delta\gamma_n,\Omega_n)\right]\left[\lambda_\mathrm{I}(\gamma_n,\Omega_n)-\lambda_\mathrm{I}(\gamma_n,\Omega_n+\Delta\Omega_n)\right] \\ &- \left[\lambda_\mathrm{I}(\gamma_n,\Omega_n)-\lambda_\mathrm{I}(\gamma_n+\Delta\gamma_n,\Omega_n)\right]\left[\lambda_\mathrm{R}(\gamma_n,\Omega_n)-\lambda_\mathrm{R}(\gamma_n,\Omega_n+\Delta\Omega_n)\right]\end{aligned} \tag{5.30}$$

and λ_R, λ_I denote the real and imaginary parts respectively of the eigenvalue. This process is repeated until the root has been found to the desired accuracy.

We used a different method to check that the modes obtained in this way were indeed self-consistent solutions of the integral equation (3.45). We calculated the transfer function $\mathcal{S}_m(\alpha,\alpha')$ over a grid of α and α', using the growth rate and pattern speed at which a mode was known to exist. This enabled us to calculate the response $A_\mathrm{res}(\alpha)$ to an imposed density pattern $A_\mathrm{imp}(\alpha')$. We iterated this procedure several times, so that

$$A_{n+1}(\alpha) = \int_{-\infty}^{+\infty} d\alpha' A_n(\alpha')\mathcal{S}_m(\alpha,\alpha'). \tag{5.31}$$

Romberg integration (Press *et al.* 1989) was used to perform the integration, without employing any smoothing procedures or Lagrange interpolation. If the initially-imposed density transform, $A_0(\alpha')$, was a mode obtained using EISPACK, then the same density pattern was recovered at each successive iteration. Furthermore, if initially a single leading log-spiral component was imposed (so that $A_0(\alpha') = \delta(\alpha_0-\alpha')$), then the pattern evolved until after about 100 iterations it had stabilised at the eigenfunction corresponding to the mode. This process therefore confirms our methods for locating modes. It relies on the same routines to calculate the transfer function, but acts as an additional check on the techniques used to build the matrix, find its eigenvalues and eigenvectors, and locate a unit eigenvalue.

5.8 Comparison with numerical simulation studies

Thanks to the help of David Earn, we were able to compare our analysis with the results from a numerical simulation. Earn studied a general power-law disk as described in his thesis (Earn 1993)[1]. His method involves expanding the potential in a set of basis functions, much as we have done using logarithmic spirals. Preliminary results were given in his thesis, but Earn later concluded that insufficient basis functions had been used. Earn has kindly provided us with revised values for three modes in a doubly cut-out disk with a rising rotation curve.

Earn (1993) studied a disk with $\beta = -0.191225$, $N = 4$, $M = 6$, $\gamma = 2.49$. The disk was given a quiet start as described by Sellwood (1983). The simulations were fitted with normal modes using the method of Sellwood and Athanassoula (1986). For the case $\tilde{R}_{\rm c} = 10$, Earn found two modes: a dominant mode with $\tilde{s} = 0.15$ and a secondary mode with $\tilde{s} = 0.09$. To find the dominant modes, we varied the growth rate s and pattern speed $\Omega_{\rm p}$ until the largest mathematical eigenvalue was unity. To find the secondary mode, we varied s and $\Omega_{\rm p}$ until the second mathematical eigenvalue was unity. The results are compared in the table below. Earn's results are quoted with their estimated error; below are the results of our analysis.

$\tilde{R}_{\rm c}$	Pattern speed $\tilde{\Omega}_{\rm p}$		Growth rate $\tilde{s}$	
10	0.39 ± 0.01	2.6%	0.15 ± 0.01	6.7%
(dominant)	0.396435	2.5%	0.127701	13%
10	0.28 ± 0.03	11%	0.09 ± 0.02	22%
(secondary)	0.324033	14%	-0.037818	140%
3	0.345 ± 0.001	0.29%	0.105 ± 0.005	4.8%
	0.359633	4.3%	0.095138	9.5%

[Earn's simulation: 25 $m = 2$ functions active, 1.2 million particles. Linear analysis: $n = 251$, $\Delta\alpha = 0.1$, $l_{\rm min} = 40$, $l_{\rm max} = 60$, $n_{\rm GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2.0$.]

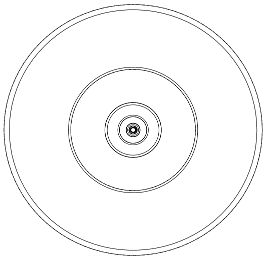
The upper percentage in each cell is the error quoted by Earn. The lower percentage is the discrepancy between our results and those of Earn. We see that in most cases the discrepancy is at least roughly consistent with the error quoted by Earn. The major exception is the secondary mode, where our analysis reports a negative growth rate (equivalently no mode). There are significant differences in the treatment adopted by Earn and the present analysis. For instance, Earn uses an inner and outer energy cut-off as well as the angular momentum cut-offs. Orbits that venture outside a certain radial range (in the present case, $0.1 < R/R_0 < 10$) are ignored. This has the effect of "nailing down" mass due to these orbits. In the circumstances, the agreement between our methods is encouraging.

---

[1] Earn's notation is different from ours. The quantity denoted β in Earn's thesis is $-\frac{1}{2}$ of our β. Earn's N_0, N_1 (eqs. (7.12) and (7.13) in his thesis) correspond to our N_β, M_β, defined in eq. (2.85). Earn uses the letter q for the anisotropy parameter (which we call γ), and uses γ for the growth rate (which we call s).

Chapter 6

Global Axisymmetric Modes



6.1 Introduction

In this chapter, we consider the simplest case of axisymmetric perturbations to axisymmetric disks. Toomre's (1964) famous study of the axisymmetric stability of disks using local theory is summarised. Then a complete global stability analysis is presented for both the self-consistent and the cut-out disks. The analysis is simplified in the axisymmetric case because there is then no dependence on pattern speed. This is intuitively obvious from the symmetry – nothing can be changed by rotating an axisymmetric pattern – and has been proved rigorously by Kalnajs (1971).

There are two kinds of axisymmetric instabilities. The first are axisymmetric Jeans modes, which cause the disk to break up into hoops of compressed and rarefied density. Cold disks are always violently unstable to such clumping. As the velocity dispersion is increased, there comes a point when pressure can stabilise the axisymmetric Jeans modes. This critical temperature is calculated via a global stability analysis. It is found that local theory always gives good – sometimes excellent – agreement with the more complicated global analysis. The second kind of axisymmetric stabilities are breathing modes, which cause dispersal or concentration of the disk. The gaseous power-law disks are known to be susceptible to such instabilities (Lemos *et al.* 1991; Syer & Tremaine 1996). Our global stability analysis shows them to be absent in the stellar dynamical power-law disks. Disks of stars are more stable than disks of gas to breathing modes.

6.2 Local stability

Toomre (1964) first discussed the stability of razor-thin disks to axisymmetric distortions. At low wavenumber (high wavelength), the centrifugal forces arising from the rotation of the disk are sufficient to stabilise the disk against the tendency to gravitational clumping. Conversely, for disturbances of short wavelength, the stars' radial velocity is enough to carry them out of the potential disturbance before it has managed to grow appreciably.

Toomre (1964) (see also Binney & Tremaine 1987, ch. 5, sec. 3) gives the following simple argument to explain the instability at low wavenumber. Consider a small circular patch, of area $(\Delta R)^2$, located on a thin stellar sheet rotating with angular velocity $\Omega_0(R)$. We shall assume for the moment that the disk is cold, so that all the stars are on circular orbits. Now suppose that the sides of the patch are reduced by a factor (1-ϵ). The mass $M = \Sigma(R)(\Delta R)^2$ contained within the patch is unchanged.

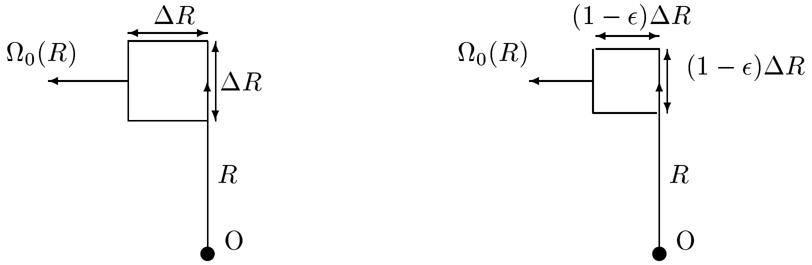


Figure 6.1: A patch of matter in a disk. On the left the patch is shown in its original state; on the right, after being compressed so its sides are reduced by a factor $(1-\epsilon)$.

The force on a unit mass at the edge of the patch pulling it towards the centre of the patch was originally $GM/(\Delta R)^2$. After the compression, this force has increased by an amount $\Delta F_{\rm g} \approx \epsilon GM/\Delta R^2 = \epsilon G\Sigma(R)$. However, this tendency towards clumping is resisted by centrifugal forces due to the rotation of the disk. The angular momentum per unit mass of a star on the edge of the original patch, about the centre of the patch, is of order $L \approx \Omega_0(R)(\Delta R)^2$. The centrifugal force per unit mass is of order $\Omega_0^2\Delta R \approx L^2/(\Delta R)^3$. We expect that L will be conserved, so when the patch is compressed, the centrifugal force increases by an amount $\Delta F_{\rm c} \approx \epsilon L^2/(\Delta R)^3 = \epsilon\Omega_0^2\Delta R$. If $\Delta F_{\rm c}$ is less than $\Delta F_{\rm g}$, the patch continues to shrink; the initial perturbation is amplified, and the disk is unstable. If $\Delta F_{\rm c}$ exceeds $\Delta F_{\rm g}$, the centrifugal force is sufficient to overcome the clumping. The compressed patch expands back again, and the disk is stable. The condition for stability is therefore $\Delta F_{\rm c} \gtrsim \Delta F_{\rm g}$, or

$$\Delta R \gtrsim \frac{G\Sigma}{\Omega_0^2}. \tag{6.1}$$

Thus the wavelength of the perturbation must exceed a critical value $\lambda_{\rm crit}$ for the disk to be stable. Toomre (1964), with a more rigorous treatment, obtains for $\lambda_{\rm crit}$ and the corresponding wavenumber [1] $\alpha_{\rm crit}$:

$$\lambda_{\rm crit} = \frac{4\pi^2 G\Sigma}{\kappa_0^2}, \qquad \alpha_{\rm crit} = \frac{\kappa_0^2 R}{2\pi G\Sigma}, \tag{6.2}$$

where κ_0 is the epicyclic frequency, i.e. the frequency of radial excursions about the circular orbit. Substituting for the surface density of the self-consistent power-law disks (2.1), we find that the disk is stable if

$$\alpha \leq \alpha_{\rm crit} = (2-\beta)\frac{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}. \tag{6.3}$$

---

[1] Here α is our dimensionless logarithmic wavenumber, which is $2\pi R/\lambda$ (see eq. (3.24)). It is not the same as Toomre's (1964) α, which is $2\pi/\lambda$, and thus has dimensions of reciprocal length.

The self-consistent power-law disks are totally scale-free, so the critical wavenumber is independent of radius. The dependence of $\alpha_{\rm crit}$ on the disk parameter β is shown in fig. 6.2. Note that the vertical axis is logarithmic.

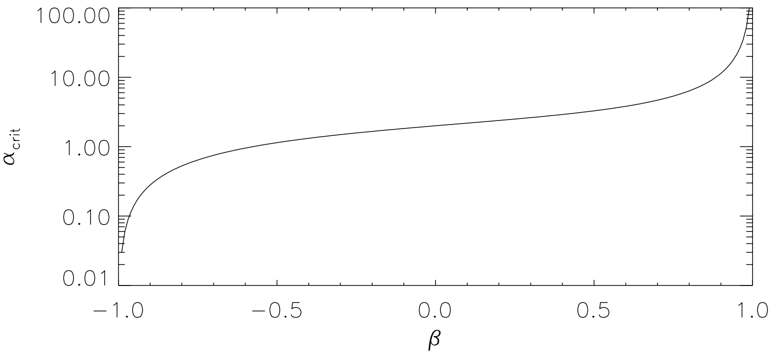


Figure 6.2: The dependence of the critical wavenumber $\alpha_{\rm crit}$ on β

The wavenumber rises (wavelength falls) as β increases. Physically, we can see why this occurs by considering eq. (6.1) and the arguments used to derive it. Since there is no dependence on radius, it is simplest to consider a patch situated at the reference radius R_0. The circular velocity here is v_β, which in our units is independent of β (section 2.2), so all disks have the same circular velocity at this point. In disks with rising rotation curves, relatively less of the density is concentrated inside $R = R_0$ (fig. 2.1), so the surface density needs to be greater at R_0 in order to achieve the same circular velocity. Thus the density of the patch Σ_0 is greater for negative β (as is readily verified from eq. (2.4)). We saw that the force tending to compress the patch, $\Delta F_{\rm g} \propto G\Sigma_0$, is opposed by a restoring centrifugal force due to the angular momentum of stars about the centre of the patch. This restoring force depends on the circular frequency at the patch: $\Omega_0 = v_\beta / R_0$. Thus $\Delta F_{\rm c} \propto v_\beta^2 \Delta R / R_0^2$. The gravitational force is greater for disks with rising rotation curves. To restore the balance, we must increase the size of the patch ΔR. Thus longer wavelengths are necessary for stabilisation by centrifugal forces when the rotation curve is rising.

The discussion so far has been for a cold disk. Simple physical arguments have indicated that the disk is stable to disturbances of sufficiently long wavelength, but unstable to perturbations of wavelength less than a critical value. This instability at short wavelengths can be overcome by heating the disk. If the disk is warm, then its stars have random motions, described by $\tilde{\sigma}_u$, which tend to carry them outside the region in which the perturbation is growing. This limits their participation in the co-operative effects which cause instability. The disk is stable if the random velocities are large enough to carry stars out of the compressed patch within the characteristic growth time τ of the perturbation: $\sigma_u \gtrsim \Delta R/\tau$. To estimate τ, we write $\Delta R \propto \exp(-t/\tau)$. Then $d^2\Delta R/dt^2 = -\Delta R/\tau^2$. Equating this to the force $GM/(\Delta R)^2$ acting on a unit mass at the edge of the patch, we obtain $\tau \approx (\Delta R/G\Sigma)^{1/2}$. The condition for stability is thus (Binney & Tremaine 1987, ch. 5)

$$\Delta R \lesssim \frac{\sigma_u^2}{G\Sigma}. \tag{6.4}$$

This places a lower limit on the wavelength of an unstable perturbation. If this coincides with (or exceeds) the upper limit provided by (6.1), no unstable disturbances will be possible. We thus arrive at an order of magnitude expression for Toomre's stability criterion: the minimum velocity dispersion necessary for stability is approximately

$$\sigma_{u,\mathrm{min}} \approx \frac{G\Sigma}{\Omega_0}. \tag{6.5}$$

Toomre (1964) carried out an exact calculation for a thin, warm, rotating disk composed of a continuum of stars with an assumed Maxwellian distribution of velocities. His analysis is valid if the velocity dispersion is small compared to the circular speed and the disturbances are of short wavelength; that is, $\sigma_u << v_{\mathrm{circ}}$ and $\alpha >> 1$. He found that marginally stable perturbations were described by the following implicit equation for the dimensional wavenumber $k = 2\pi/\lambda$ (also derived by Kalnajs in 1963)

$$\frac{|k|\sigma_u^2}{2\pi G\Sigma} + e^{-k^2\sigma_u^2/\kappa_0^2} I_0\left(\frac{k^2\sigma_u^2}{\kappa_0^2}\right) = 1. \tag{6.6}$$

I_0 is a modified Bessel function, $I_\nu(z) = e^{-i\pi\nu/2} J_\nu(ze^{i\pi/2})$ (see Binney & Tremaine 1987, app. 1.C.) Eq. (6.6) can be solved to give the marginal stability curve for a particular surface density.

To investigate the solutions to this equation, we define a response function

$$\mathcal{R}(\alpha, \tilde{\sigma}_u) = \left[|\alpha|\tilde{\sigma}_u^2 \frac{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]} + \exp\left(-(\alpha\tilde{\sigma}_u)^{\frac{2}{2-\beta}}\right) I_0\left((\alpha\tilde{\sigma}_u)^{\frac{2}{2-\beta}}\right)\right]^{-1}. \tag{6.7}$$

This is the reciprocal of the left-hand side of (6.6), expressed in terms of α and $\tilde{\sigma}_u$. Solutions to (6.6) thus correspond to $\mathcal{R}(\alpha, \tilde{\sigma}_u) = 1$. We plot $\mathcal{R}(\alpha, \tilde{\sigma}_u)$ against wavenumber α for several values of the velocity dispersion $\tilde{\sigma}_u$ in fig. 6.3. For all $\tilde{\sigma}_u$, there is a solution at $\alpha = 0$, as is apparent from eq. (6.7). These are the *breathing modes* (Lemos *et al.* 1991), for which the density perturbation has no radial nodes. Let us remark that local theory may not be a good guide for such long-wavelength disturbances. A full global analysis is really needed to ascertain the existence of the breathing modes. For sufficiently high values of $\tilde{\sigma}_u$, there are no solutions other than the breathing mode. The disk is stable to the Jeans instability. As $\tilde{\sigma}_u$ is reduced, we encounter a value $\tilde{\sigma}_{u,\mathrm{min}}$ for which the disk is just stable: the curve $\mathcal{R}(\alpha, \tilde{\sigma}_u)$ touches the line $\mathcal{R} = 1$ only once, at the most unstable wavenumber α_{u}. For $\tilde{\sigma}_u < \tilde{\sigma}_{u,\mathrm{min}}$, the curve $\mathcal{R}(\alpha, \tilde{\sigma}_u)$ crosses the line $\mathcal{R} = 1$ twice, so there are two solutions in addition to the breathing mode. This is illustrated in fig. 6.4, which focuses on the region of low wavenumber close to $\mathcal{R} = 1$. It shows the curve $\mathcal{R}(\alpha, \tilde{\sigma}_{u,\mathrm{min}})$ (dashed line) and marks the most unstable wavenumber α_{u}.

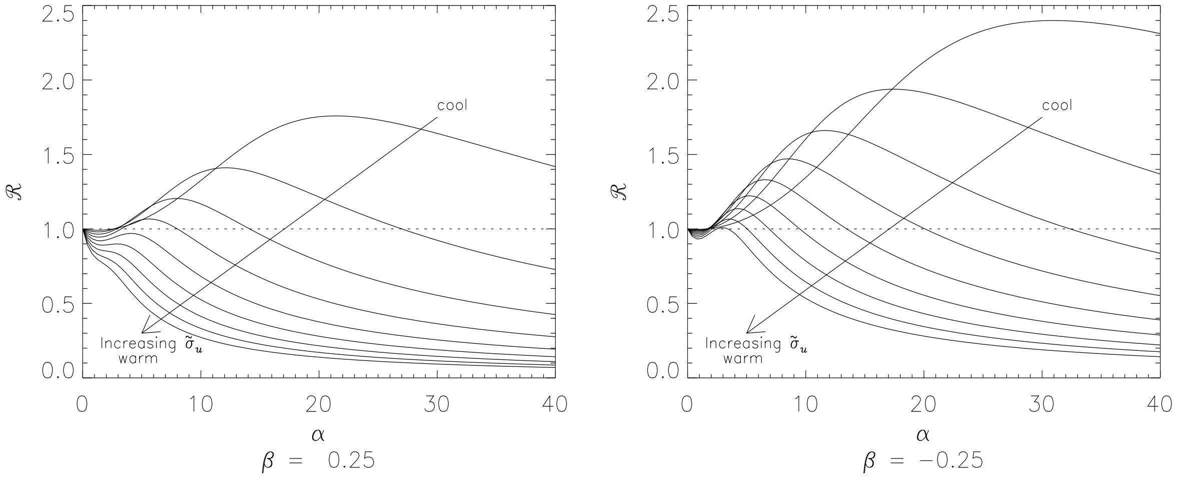


Figure 6.3: The function $\mathcal{R}(\alpha, \tilde{\sigma}_u)$ plotted against wavenumber α for 9 values of the velocity dispersion: $\tilde{\sigma}_u = 0.10$, 0.15, 0.20, 0.25 ... 0.5. The line $\mathcal{R} = 1$, corresponding to solutions of (6.6), is marked with a dotted line. The left-hand plot shows $\beta = 0.25$, for which $\tilde{\sigma}_{u,\mathrm{min}} = 0.283$ and $\alpha_{\mathrm{u}} = 4.55$. The right-hand plot shows $\beta = -0.25$, for which $\tilde{\sigma}_{u,\mathrm{min}} = 0.509$ and $\alpha_{\mathrm{u}} = 2.87$.

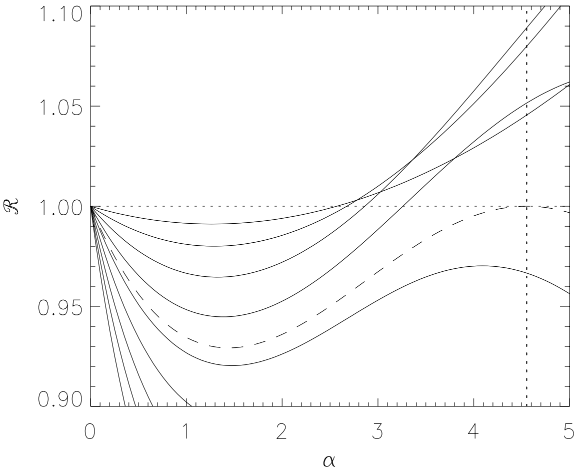


Figure 6.4: The function $\mathcal{R}(\alpha, \tilde{\sigma}_u)$ plotted against wavenumber α for $\beta = 0.25$ and 9 values of the velocity dispersion: $\tilde{\sigma}_u = 0.10$, 0.15, 0.20, 0.25 ... 0.5. The dashed curve is that with $\tilde{\sigma}_u = \tilde{\sigma}_{u,\mathrm{min}}$. The most unstable wavenumber α_{u} is marked with a dotted line. For $\beta = 0.25$, $\tilde{\sigma}_{u,\mathrm{min}} = 0.283$ and $\alpha_{\mathrm{u}} = 4.55$.

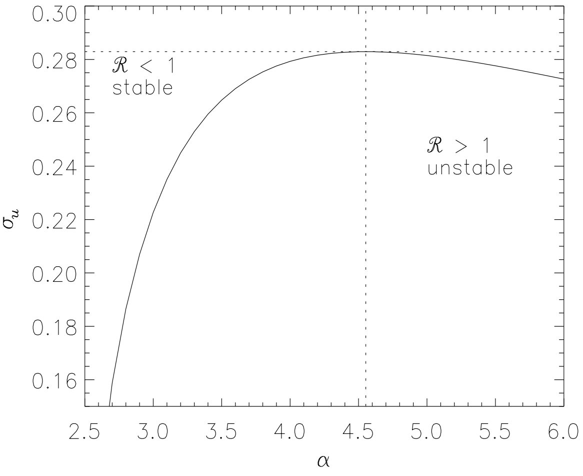


Figure 6.5: The local marginal stability curve $\mathcal{R} = 1$ for disks with $\beta = 0.25$. The minimum temperature needed for stability and the most unstable wavenumber are marked with dotted lines. For $\beta = 0.25$, $\tilde{\sigma}_{u,\mathrm{min}} = 0.283$ and $\alpha_{\mathrm{u}} = 4.55$.

The solutions of eq. (6.6) define the *marginal stability curve* for a particular surface density. Fig. 6.5 shows this for $\beta = 0.25$. As we saw in figs. 6.3 and 6.4, for sufficiently high values of the velocity dispersion $\tilde{\sigma}_u$, there are no solutions (other than the breathing mode). As the disk is cooled, it becomes unstable at the minimum velocity dispersion $\tilde{\sigma}_{u,\mathrm{min}}$ and most unstable wavenumber α_{u}, marked by dotted lines on the above plots. For still lower values of $\tilde{\sigma}_u$, there are two solutions for each $\tilde{\sigma}_u$.

Fig. 6.3 indicates that disks with rising rotation curves (negative β) are less stable than those with falling rotation curves. The curves in fig. 6.3 are shifted upwards for $\beta = -0.25$ relative to $\beta = 0.25$. Thus a curve such as that with $\tilde{\sigma}_u = 0.3$, which for $\beta = 0.25$ never crosses the line $f = 1$, intersects $f = 1$ twice for $\beta = -0.25$. $\tilde{\sigma}_u = 0.3$ is enough to stabilise the disk with $\beta = 0.25$, but not that with $\beta = -0.25$. To see in more detail how $\tilde{\sigma}_{u,\min}$ and $\alpha_{\rm u}$ depend on β, we rewrite (6.6) in terms of

$$x = \frac{\alpha}{\alpha_{\rm crit}}, \qquad y = \frac{\alpha_{\rm crit}\sigma_u}{R\kappa_0} = \frac{\sigma_u \kappa_0}{2\pi G\Sigma}, \tag{6.8}$$

where $\alpha_{\rm crit}$ is defined by (6.2). Eq. (6.6) then becomes

$$|x|y^2 + e^{-x^2y^2} I_0(x^2y^2) = 1. \tag{6.9}$$

This defines a marginal stability curve in the x-y plane separating stable regions from unstable, analogous to that shown in fig. 6.5. The maximum value of the curve – call it the point $(x_{\max}, y_{\max})$ – defines both the velocity dispersion $\sigma_{u,\min}$ at which the disk is locally just stable and the most unstable wavenumber $\alpha_{\rm u}$. The power of Toomre's approach is that these can be calculated completely generally without making assumptions about the form of Σ and κ_0. Eq. (6.9) can be solved numerically to give $x_{\max} = 1.8217367$ and $2\pi y_{\max} = 3.3582830$. So for the self-consistent power-law disks, local theory predicts that the wavenumber of the first mode to appear as the disk is cooled is

$$\alpha_{\rm u} = 1.8217367\,(2-\beta)\frac{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}. \tag{6.10}$$

Local theory predicts stability when

$$\sigma > \tilde{\sigma}_{u,\min} = \frac{3.3582830}{2\pi\sqrt{2-\beta}}\frac{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}. \tag{6.11}$$

The following graph shows the dependence of the dimensionless velocity dispersion $\tilde{\sigma}_{u,\min}$ on β. Note that the vertical axis is logarithmic.

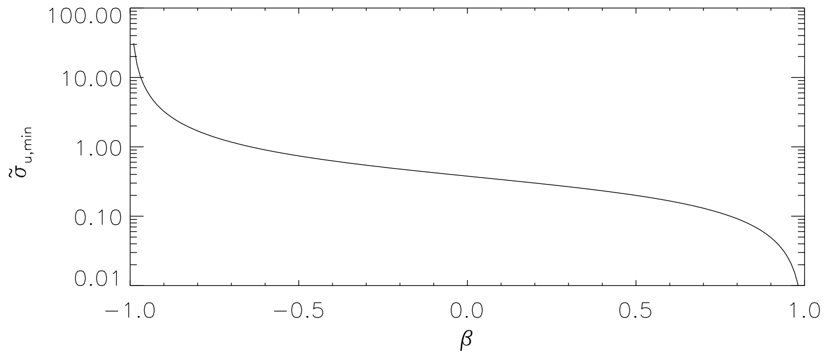


Figure 6.6: The dependence of the minimum dimensionless velocity dispersion $\tilde{\sigma}_{u,\min}$ on β

The minimum dimensionless velocity dispersion falls steeply with increasing β. Mathematically, this is due to the gamma functions in (6.11). We can see physically why it occurs by returning to the arguments used in deriving the order-of-magnitude expression (6.5). We saw in the derivation of (6.1) that a cold disk is always unstable to gravitational clumping on a sufficiently small scale, since the centrifugal force can stabilise perturbations only down to a finite wavelength λ_{crit}. A warm disk can become stable to these low-wavelength perturbations, if its stars have enough random velocity σ_u to carry them out of the perturbation before it has grown significantly. The larger the perturbation, the further the star has to travel in order to be clear; on the other hand, larger perturbations also grow more slowly. The growth time of the perturbation increases with wavelength only as $\sqrt{\lambda}$, so as the wavelength of the perturbation increases, stars need more random velocity in order to escape: $\sigma_u \approx \sqrt{\lambda}$. As in the discussion following fig. 6.2, we consider a patch situated at the reference radius R_0, and define the surface density here to be unity. We saw then that for disks with rising rotation curves (negative β), the circular frequency Ω_0 is smaller and the centrifugal force weaker, and thus the patch remains unstable up to longer wavelengths than for disks with falling rotation curves. As we have just seen, the stars need to move faster in order to stabilise longer wavelengths. Thus the random motion needed to ensure stability at all wavelengths is greater for rising than for falling rotation curves. In fact, this tendency is exacerbated by the dimensionless variables we use. The dimensionless velocity dispersion $\tilde{\sigma}_u$ is obtained by comparing the dimensional velocity dispersion σ_u to the reference velocity v_β (2.70). For negative β, $\sigma_{u,\text{min}}$ is larger and v_β is smaller. This explains the steep fall in $\tilde{\sigma}_{u,\text{min}}$ as β increases.

Here we have used the dimensionless velocity dispersion $\tilde{\sigma}_{u,\text{min}}$ to compare the stability of different disks. It is not immediately clear that this is the best parameter to use for the comparison, since the conversion of $\tilde{\sigma}_u$ into a physical velocity dispersion depends on β. The magnitude of the physical velocity dispersion at a particular point in various disks depends on the units used. If we use our usual units, in which $v_\beta = R_0 = 1$, then the dependence of the minimum dimensional velocity dispersion $\sigma_{u,\text{min}}$ on β is given by

$$\sigma_{u,\text{min}} = \frac{3.3582830 v_\beta}{2\pi\sqrt{2-\beta}} \frac{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]} \left(\frac{R_0}{R}\right)^{\beta/2}. \tag{6.12}$$

In units in which $\Sigma_0 = R_0 = 1$, the dependence on β is

$$\sigma_{u,\text{min}} = \frac{3.3582830}{\sqrt{2-\beta}} \sqrt{\frac{G\Sigma_0 R_0}{2\pi}} \sqrt{\frac{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}} \left(\frac{R_0}{R}\right)^{\beta/2}. \tag{6.13}$$

Figs. 6.7 and 6.8 show how the dimensional velocity dispersion $\sigma_{u,\text{min}}$ varies with radius, in disks with different values of β. In fig. 6.7, units where $v_\beta = 1$ are employed; in fig. 6.8, $G\Sigma_0 = 1$ (c.f. figs. 2.3 and 2.4). Except at the very centre of the disk, where the velocity dispersion becomes infinite, $\sigma_{u,\text{min}}$ is always larger for disks with rising rotation curve. Thus there is no unit-dependent ambiguity: we can definitely state that disks with rising rotation curves are less stable. In the remainder of this thesis, we shall always use units in which $v_\beta = 1$.

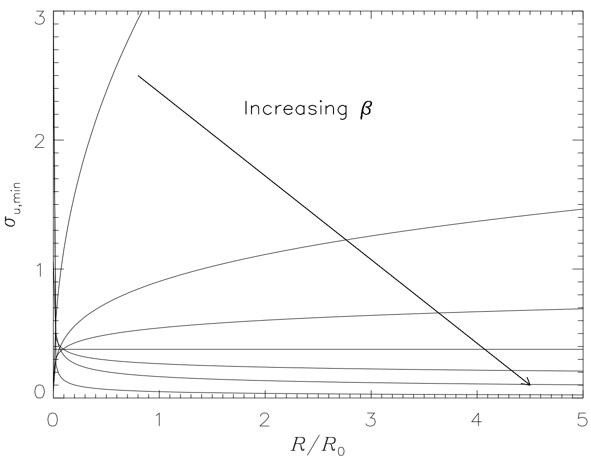


Figure 6.7: The minimum dimensional velocity dispersion $\sigma_{u,\min}$, in units where $v_\beta = R_0 = 1$. $\sigma_{u,\min}$, given in eq. (6.12), is plotted against radius for 7 different values of β, from $\beta = -0.9$ to $\beta = +0.9$.

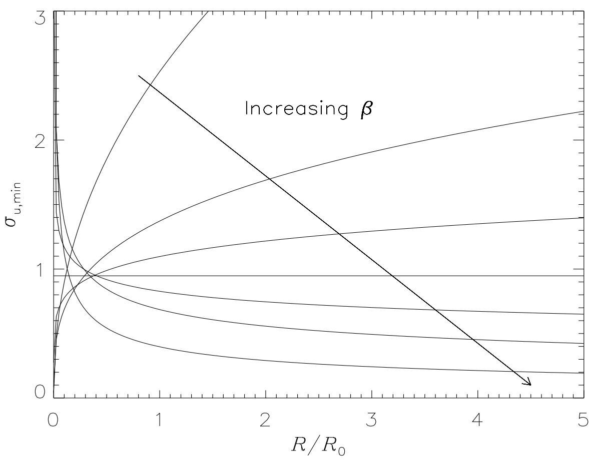


Figure 6.8: The minimum dimensional velocity dispersion $\sigma_{u,\min}$, in units where $G\Sigma_0 = R_0 = 1$. $\sigma_{u,\min}$, given in eq. (6.13), is plotted against radius for 7 different values of β, from $\beta = -0.9$ to $\beta = +0.9$.

Toomre (1964) introduced the quantity Q to describe the stability. Q, the stability parameter, is the ratio of the velocity dispersion at a point to that necessary for local axisymmetric stability. Thus in general

$$Q \equiv \frac{\sigma_u}{\sigma_{u,\min}} = \frac{1}{3.3582830} \frac{\sigma_u \kappa_0}{G\Sigma}. \tag{6.14}$$

Subsequently, when discussing cut-out disks, we shall often find it useful to relate the velocity dispersion $\tilde{\sigma}_u$ in a particular disk to the $\tilde{\sigma}_{u,\min}$ (6.11) necessary to ensure local axisymmetric stability in the self-consistent disk. We shall refer to this ratio as Q_{s}

$$Q_{\mathrm{s}} \equiv \tilde{\sigma}_u \frac{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]} \frac{2\pi\sqrt{2-\beta}}{3.3582830}. \tag{6.15}$$

This is not quite Toomre's Q, in that we are comparing the actual velocity dispersion to that needed for stability in a *different* disk. Q_{s} is related to the usual Q by

$$Q = Q_{\mathrm{s}} \frac{\Sigma_{\mathrm{eq}}}{\Sigma_{\mathrm{active}}} \tag{6.16}$$

6.3 Global stability to an axisymmetric perturbation

We now begin our global stability analysis. In the following sections, we investigate the global stability of the self-consistent and cut-out disks, using the integral equation machinery developed in the preceding chapter. In the case of an axisymmetric perturbation, the integral equation admits certain symmetries. The first simplification is that the frequency is purely imaginary: $\omega = is$. There is no dependence on the pattern speed, since nothing is changed by rotating an axisymmetric pattern. Secondly, the integral equation is equivalent to one with a Hermitian kernel, and hence must have purely real eigenvalues.

As a preliminary to proving this, we first note the symmetries of the angular momentum function and the Fourier coefficient in the axisymmetric case. From the definition of F_{lm} (3.77), we have

$$F_{l0} = \frac{1}{2\pi} \left| \frac{2 + 2\gamma - \beta\gamma}{\beta\tilde{U}^2 + \beta - 2} \right| \int_0^\infty \frac{l\tilde{\kappa} e^{-i\eta \frac{2}{2-\beta} \ln \tilde{L}_z}}{l\tilde{\kappa} - i\tilde{s}\tilde{L}_z^{\frac{2+\beta}{2-\beta}}} \tilde{H}(\tilde{L}_z) \frac{d\tilde{L}_z}{\tilde{L}_z}. \tag{6.17}$$

It is straightforward to see that, whatever the form of the cut-out function,

$$F_{l0}(\eta) = F^*_{-l0}(-\eta). \tag{6.18}$$

Similarly, from the definition of Q_{lm} given in eq. (5.1), we have

$$Q_{l0}(\alpha) = \frac{1}{\pi} \int_0^\pi \exp\left\{ \left(i\alpha - \tfrac{1}{2}\right)\tilde{X} \right\} \cos l\chi \, d\chi. \tag{6.19}$$

and hence the symmetries

$$Q^*_{l0}(\alpha) = Q_{l0}(-\alpha) \qquad\qquad Q_{-l0}(\alpha) = Q_{l0}(\alpha) \tag{6.20}$$

(The first of these in fact holds for general m.) We now recast the integral equation (5.6) into a form with a Hermitian kernel. Following Zang (1976), we define a modified transfer function and density transform

$$T_0(\alpha, \alpha') = \sqrt{\frac{K(\alpha, 0)}{K(\alpha', 0)}} \mathcal{S}_0(\alpha, \alpha') \qquad\qquad B(\alpha) = \sqrt{K(\alpha, 0)} A(\alpha) \tag{6.21}$$

The integral equation (5.6) now becomes

$$\lambda B(\alpha) = \int_{-\infty}^{+\infty} d\alpha' B(\alpha') T_0(\alpha, \alpha'), \tag{6.22}$$

The value of this transformation is that $T_0(\alpha, \alpha')$ is Hermitian. We have, from eqs. (6.21) and (3.79)

$$\begin{aligned} T_0(\alpha, \alpha') = {} & 2\pi G\tilde{C}R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} \sqrt{K(\alpha, 0) K(\alpha', 0)} \int d\tilde{U} \, \mathcal{J}_0(\tilde{U})\tilde{U} \\ & \times \left| \frac{\beta\tilde{U}^2 + \beta - 2}{2\beta} \right|^{\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}} \sum_{l=-\infty}^{+\infty} Q_{l0}(\alpha') Q^*_{l0}(\alpha) F_{l0}(\alpha - \alpha'). \end{aligned}$$

We make the transformations $\alpha \to \alpha'$, $\alpha' \to \alpha$, $l \to -l$, and take the complex conjugate, to obtain

$$\begin{aligned} T^*_0(\alpha', \alpha) = {} & 2\pi G\tilde{C}R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} \sqrt{K(\alpha', 0) K(\alpha, 0)} \int d\tilde{U} \, \mathcal{J}_0(\tilde{U})\tilde{U} \\ & \times \left| \frac{\beta\tilde{U}^2 + \beta - 2}{2\beta} \right|^{\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}} \sum_{l=+\infty}^{-\infty} Q^*_{-l0}(\alpha) Q_{-l0}(\alpha') F^*_{-l0}(\alpha' - \alpha). \end{aligned}$$

We now use the symmetries $Q_{-l0}(\alpha) = Q_{l0}(\alpha)$ and $F^*_{-l0}(\alpha' - \alpha) = F_{l0}(\alpha - \alpha')$ to derive the Hermitian nature of T_0:

$$T^*_0(\alpha', \alpha) = T_0(\alpha, \alpha') \tag{6.23}$$

The modified integral equation (6.22) is thus a homogeneous, linear Fredholm integral equation with a Hermitian kernel. It must therefore have real eigenvalues and orthogonal eigenfunctions (Tricomi 1985). Each eigenvector B_n must satisfy the modified integral equation (6.22) with its corresponding eigenvalue λ_n.

$$\lambda_n B_n(\alpha) = \int_{-\infty}^{+\infty} d\alpha' B_n(\alpha') T_0(\alpha, \alpha'). \tag{6.24}$$

We then multiply this equation by the complex conjugate of another eigenvector B_m, and integrate over all α to arrive at

$$\lambda_n \int_{-\infty}^{+\infty} d\alpha B_m^*(\alpha) B_n(\alpha) = \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} d\alpha' B_m^*(\alpha) B_n(\alpha') T_0(\alpha, \alpha'). \tag{6.25}$$

If we interchange n and m and take the complex conjugate, we obtain

$$\lambda_m^* \int_{-\infty}^{+\infty} d\alpha B_n(\alpha) B_m^*(\alpha) = \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} d\alpha' B_n(\alpha) B_m^*(\alpha') T_0^*(\alpha, \alpha'). \tag{6.26}$$

We now interchange the dummy variables α and α' and swap the order of integration, and then use the fact that the kernel is Hermitian and hence $T_0^*(\alpha', \alpha) = T_0(\alpha, \alpha')$. This yields

$$\lambda_m^* \int_{-\infty}^{+\infty} d\alpha B_n(\alpha) B_m^*(\alpha) = \int_{-\infty}^{+\infty} d\alpha \int_{-\infty}^{+\infty} d\alpha' B_n(\alpha') B_m^*(\alpha) T_0(\alpha, \alpha'). \tag{6.27}$$

We now subtract (6.27) from (6.25) to obtain

$$(\lambda_n - \lambda_m^*) \int_{-\infty}^{+\infty} d\alpha B_m^*(\alpha) B_n(\alpha) = 0. \tag{6.28}$$

Thus if $n \neq m$, the eigenvectors are orthogonal, while if $n = m$, the integrand is positive definite and hence the eigenvalue is purely real (except for the trivial case $B_n(\alpha) = 0$). This occurs because axisymmetric eigenvalues have no dependence on pattern speed. In general, the mathematical eigenvalue λ is an analytic function of the complex frequency ω. For $m = 0$, the frequency is purely imaginary: from covering the complex plane, it collapses onto the imaginary axis. As a result, the eigenvalue collapses onto the real axis. Our problem thus becomes one-dimensional. Mathematically, the proof that the eigenvalues are real depends on the symmetry $F_{l0}(\eta) = F_{-l0}^*(-\eta)$ (6.18), which in turn depends on the fact that ω is imaginary.

There is a further symmetry. If we make the transformations $\alpha \to -\alpha$, $\alpha' \to -\alpha'$, $l \to -l$ in (6.3), and take the complex conjugate, we obtain

$$\begin{aligned} T_0^*(-\alpha, -\alpha') = {} & 2\pi G \tilde{C} R_0^{\gamma+1} v_\beta^{\frac{2}{\beta}(1+\gamma)} \sqrt{K(-\alpha, 0) K(-\alpha', 0)} \int d\tilde{U}\, \mathcal{J}_0(\tilde{U}) \tilde{U} \\ & \times \left| \frac{1}{2} \left(\tilde{U}^2 + 1 - \frac{2}{\beta} \right) \right|^{\frac{1}{\beta} + \frac{\gamma}{\beta} - \frac{\gamma}{2}} \sum_{l=-\infty}^{+\infty} Q_{-l0}^*(-\alpha') Q_{-l0}(-\alpha) F_{-l0}^*(-\alpha + \alpha'). \end{aligned}$$

Using the symmetry properties of the Fourier coefficients (6.20) and the angular momentum function (6.18), along with that of the Kalnajs gravity factor $K(\alpha, m) = K(-\alpha, m)$ (3.34), we obtain the result

$$T_0^*(-\alpha, -\alpha') = T_0(\alpha, \alpha'). \tag{6.29}$$

This has the consequence that the eigenvalues are degenerate: if $B(\alpha)$ is an eigenvector with eigenvalue λ, then so is $B^*(-\alpha)$. To see this, we make the transformations $\alpha \to -\alpha$, $\alpha' \to -\alpha'$ in the modified integral equation (6.22), and take the complex conjugate. We obtain

$$\lambda B^*(-\alpha) = -\int_{\infty}^{-\infty} d\alpha' B^*(-\alpha') T_0^*(-\alpha, -\alpha') = \int_{-\infty}^{+\infty} d\alpha' B^*(-\alpha') T_0(\alpha, \alpha'), \tag{6.30}$$

i.e. $B^*(-\alpha)$ also satisfies the integral equation (6.22). In terms of our original density transforms, $A(\alpha)$ and $A^*(-\alpha)$ are a degenerate pair with the same eigenvalue λ. This is expected for axisymmetric perturbations. When $m = 0$, the logarithmic “spirals” are just rings, and there is no difference between log-spirals with wavenumber $\pm\alpha$. Thus the integral equation cannot favour one sign of the wavenumber over another.

6.4 Neutral modes in the self-consistent disk

The self-consistent disk has no intrinsic length- or time-scales. A neutral mode contributes no length- or time-scales of its own, meaning that the problem is entirely self-similar, so the response to a single log-spiral is itself a log-spiral. In Chapter 4, we derived an expression for the neutral response function $\mathcal{R}$ (4.40). The form of this function for $m = 0$ is shown in fig. 6.9. This figure is analogous to fig. 6.3, which plotted the response function for the local disk. We see that the structure of the curves is very similar. Once again, increased velocity dispersion always tends to stabilise the disk.

The major difference between fig. 6.9 and fig. 6.3 is that the local response curves have a solution at $\alpha = 0$, whereas the global curves do not. The global analysis does not admit breathing modes. Breathing modes are analogous to the instability of stars with $\gamma < 4/3$, where γ is the ratio of the heat capacity at constant pressure to the heat capacity at constant volume (see Binney & Tremaine 1987, ch. 5, p. 302). These are stars dominated by radiation pressure. In the stellar dynamics of spherical systems, $\gamma = 5/3$, as for a monatomic gas. Thus breathing modes do not exist in spherical stellar systems. It appears that a similar result holds for our disks. The indication by local theory that breathing modes exist (section 6.2) was misleading. This is not surprising, since the theory had been taken so far beyond the small-wavelength regime in which it is trustworthy.

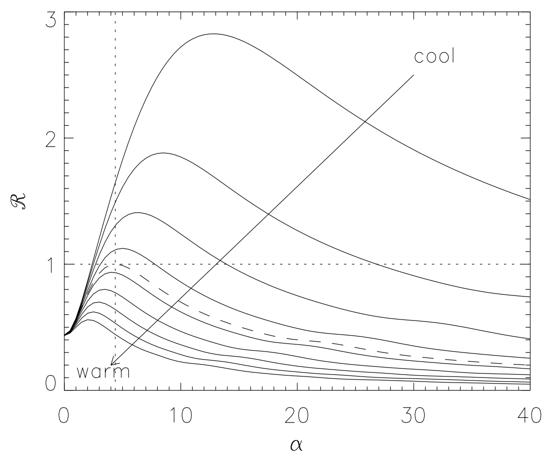


Figure 6.9: The response function $\mathcal{R}(\alpha, \tilde{\sigma}_u)$ for the $\beta = 0.25$ disk. $\mathcal{R}$ is plotted as a function of wavenumber α for 9 values of the velocity dispersion: $\tilde{\sigma}_u$ =0.1, 0.15, 0.2, 0.25 .. 0.5. Also shown, with a dashed line, is the curve with $\tilde{\sigma}_u = \tilde{\sigma}_{u,\mathrm{min}}$. The most unstable wavenumber α_u is marked with a dotted line. For $\beta = 0.25$, $\tilde{\sigma}_{u,\mathrm{min}} = 0.281$ and $\alpha_\mathrm{u} = 4.38$.

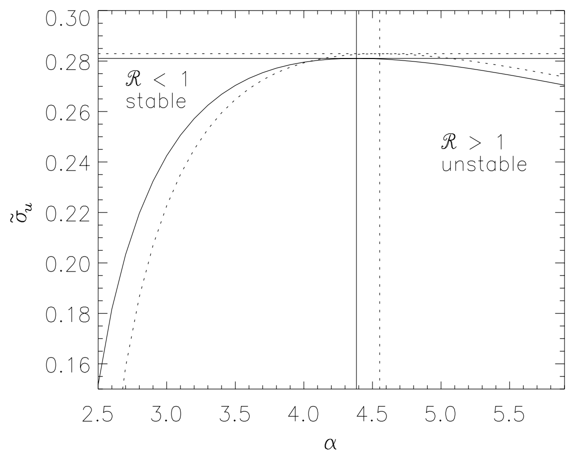


Figure 6.10: The marginal stability curve $\mathcal{R} = 1$ for the $\beta = 0.25$ disk. The solid lines show the result of the global calculation; the dotted lines show results using local theory. The temperature and wavenumber at which instability sets in are marked. These are $\tilde{\sigma}_{u,\mathrm{min}} = 0.281$ for global stability, $\tilde{\sigma}_{u,\mathrm{min}} = 0.283$ for local stability; $\alpha_\mathrm{u} = 4.38$ for global stability, $\alpha_\mathrm{u} = 4.55$ for local stability.

As Binney & Tremaine (1987) explain, breathing modes can occur in gas disks because the gas molecules have internal degrees of freedom, which absorb energy released in changes such as contraction. This energy is then not available to the translational degrees of freedom which contribute to the pressure. There is thus insufficient pressure to resist gravitational collapse. Conversely, when a stellar disk contracts, all the energy released contributes to increased random velocity, which tends to counteract the effect of the increased gravity. Thus stellar disks are more stable than gaseous disks to breathing modes.

We plot the marginal stability curve $\mathcal{R} = 1$ in the plane of dimensionless wavenumber α and velocity dispersion $\tilde{\sigma}_u$, just as performed in section 6.2 for local theory. Fig. 6.10 compares the marginal stability curves obtained using local theory (dotted lines) and global theory (solid lines). The agreement is good, especially close to the critical temperature.

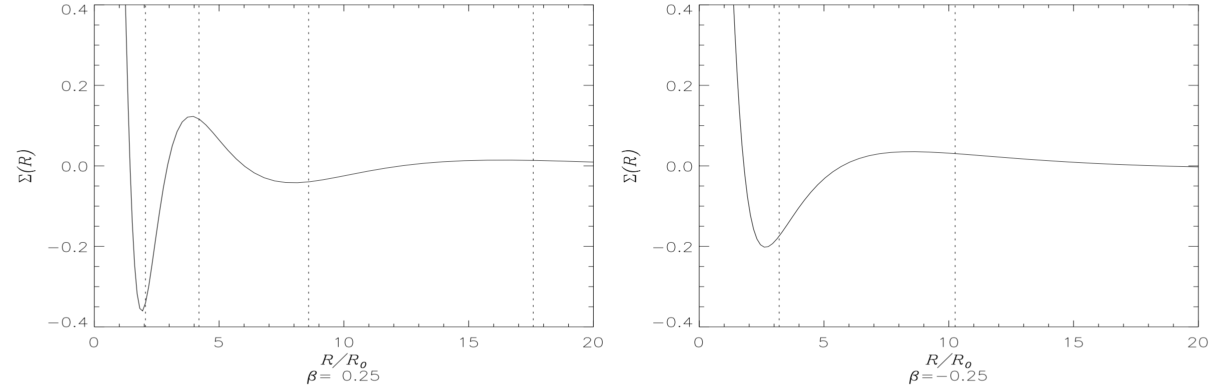


Figure 6.11: The most unstable $m = 0$ modes in self-consistent disks with $\beta = 0.25$ and $\beta = -0.25$. The graphs show the real part of the density: $\Sigma(R) = (R_0/R)^{3/2} \cos(\alpha_\mathrm{u} \ln(R/R_0))$. For $\beta = 0.25$, $\alpha_\mathrm{u} = 4.38$; for $\beta = -0.25$, $\alpha_\mathrm{u} = 2.70$. The vertical dotted lines mark the maxima and minima of the cos term, which occur at $R = \exp(n\pi/\alpha)$.

Fig. 6.11 shows the first neutral mode to set in, for disks with $\beta = \pm 0.25$. For the falling rotation curve ($\beta = 0.25$), the most unstable wavenumber is larger than for the rising rotation curve ($\beta = -0.25$). Thus the maxima of the oscillatory component in $\Sigma(R)$ are more closely spaced for the falling rotation curve.

Figs. 6.12 and 6.13 compare the results obtained with a global calculation (solid lines) to those from local theory (dotted lines). The results used in these graphs are tabulated in Appendix E. Fig. 6.12 shows the minimum temperature needed for axisymmetric stability. In the left-hand plot, this is shown as the minimum velocity dispersion $\tilde{\sigma}_{u,\mathrm{min}}$. As Toomre and Zang found for $\beta = 0$ (Zang 1976), the agreement is excellent. The right-hand plot presents the same data in terms of Toomre's stability parameter Q_{s}. This is the ratio of the global $\tilde{\sigma}_{u,\mathrm{min}}$ to that calculated by local theory. It is clearer in this plot that the local and global results coincide most closely at $\beta = 0$, where the rotation curve is flat. This is probably because Toomre's local analysis assumes that the stellar disk has a Maxwellian velocity distribution. This is roughly true for the Toomre-Zang disks (see eq. (2.60)), but not for the other power-law disks.

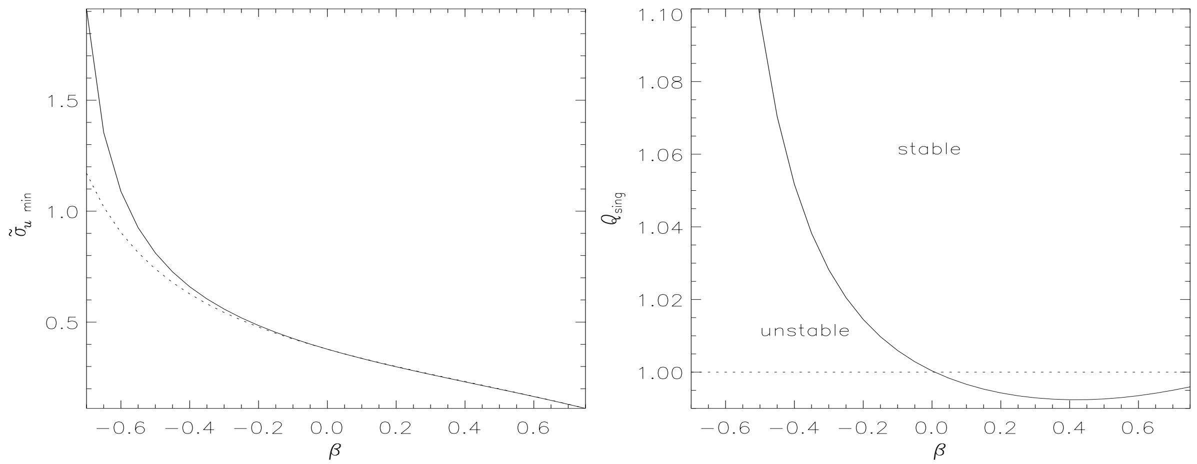


Figure 6.12: Minimum temperature for global axisymmetric stability plotted against β. In the left-hand plot, this is presented as the minimum velocity dispersion $\tilde{\sigma}_u$, and in the right-hand plot as the stability parameter Q_{s}, necessary to stabilise the disk to axisymmetric perturbations. The dotted line shows the minimum temperature needed for local stability.

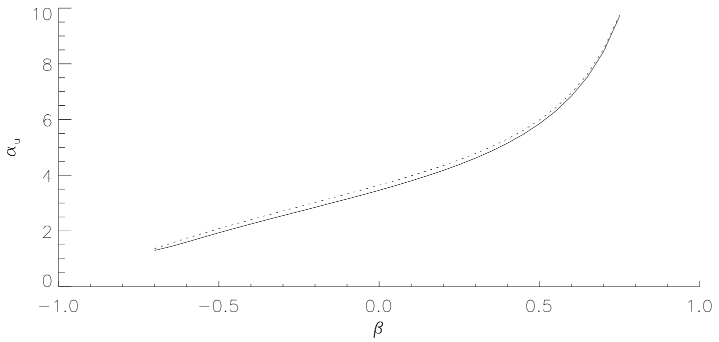


Figure 6.13: Most unstable wavenumber α_{u} plotted against β, for an axisymmetric perturbation in a self-consistent disk. The solid line shows the results of the global calculation, the dotted line those for local theory. The right-hand plot shows the same data expressed in terms of the most unstable wavelength $\lambda_{\mathrm{u}} \equiv 2\pi R/\alpha_{\mathrm{u}}$.

Local theory is seriously discrepant only when $\beta \lesssim -0.25$. Toomre's local analysis assumes that the velocity dispersion is small compared to the circular velocity, and that the wavelength of any disturbance is small compared to the radius R at the point under consideration. In our notation, these two conditions become $\tilde{\sigma}_u << 1$ and $\alpha >> 1$. We see from figs. 6.12 and 6.13 that these conditions cease to be valid when $\beta \approx -0.25$. As β becomes more negative, the wavelength of the unstable perturbation becomes larger, undermining local theory. The local calculation then underestimates the velocity dispersion needed for axisymmetric stability. Disks with rising rotation curves remain globally unstable even when they are hot enough to be locally stable.

6.5 Global stability of the cut-out disks

How does this marginal stability analysis for the self-consistent disk compare with results from the cut-out disks? The most obvious difference is that the self-consistent disk is self-similar, so that instabilities set in at all radii at once. When we introduce a cut-out, we break this self-similarity. In the cut-out regions of the disk, the active surface density is now less than the equilibrium density. Using (2.87) as an approximation for the density of a doubly cut-out disk, the stability criterion $\sigma_{\mathrm{u,min}} = 3.36\, G\Sigma/\kappa_0$ becomes

$$\begin{aligned}\tilde{\sigma}_{u,\mathrm{min}} = \; & \frac{3.3582830}{2\pi\sqrt{2-\beta}} \frac{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]} \\ & \times \frac{R^{N(2+\beta)/2}(R_0\tilde{R}_{\mathrm{c}})^{M(2+\beta)/2}}{\left[R^{N(2+\beta)/2} + R_0^{N(2+\beta)/2}\right]\left[R^{M(2+\beta)/2} + (R_0\tilde{R}_{\mathrm{c}})^{M(2+\beta)/2}\right]}. \end{aligned} \tag{6.31}$$

This tends to the self-consistent disk result if $R_0 << R << R_0\tilde{R}_{\mathrm{c}}$. The cut-out function lowers the density Σ in certain regions of the disk, and thus tends to reduce the minimum velocity dispersion needed for local stability in these regions. The minimum velocity dispersion needed to stabilise the disk *everywhere* to axisymmetric perturbations is set principally by the regions of the disk which have been least affected by the cut-out. Thus we expect to find that the presence of the cut-out does not greatly affect $\tilde{\sigma}_{u,\mathrm{min}}$. This conclusion holds because axisymmetric log-spiral waves do not propagate through the disk. We can see this mathematically by considering the real part of each log-spiral component (3.25):

$$\mathrm{Re}(\Sigma^{\alpha m}_{\mathrm{imp}}) = \Sigma_{\mathrm{p}} e^{st} \left(\frac{R_0}{R}\right)^{3/2} \cos\left[m(\theta - \Omega_{\mathrm{p}} t) + \alpha \ln\frac{R}{R_0}\right]. \tag{6.32}$$

Crests of the density wave occur when the argument of the cosine is equal to $2n\pi$. Along a line of constant θ, the radial position of the nth crest varies as

$$R_{\mathrm{crest}} = R_0 \exp\frac{1}{\alpha}[2n\pi - m(\theta - \Omega_{\mathrm{p}} t)]. \tag{6.33}$$

For $m = 0$, the crests of the density wave remain in the same position. The wave grows or decays, but does not propagate through the disk. Thus the axisymmetric perturbation is relatively unaffected by the presence of the cut-out. For non-zero values of m, the waves do propagate. They are greatly affected by the presence of the inner density cut-off, which presents a boundary where reflection and refraction can occur (see section 7.2). In subsequent chapters, we shall see that the self-consistent and cut-out disks have very different stability to non-axisymmetric modes.

The limit of vanishing growth rate

In the axisymmetric case, the limit of vanishing growth rate is rather delicate, and worth some discussion. As for the scale-free disk, the results of setting s equal to zero are not the same as the limiting value of the results as $s \to 0$. The general expression for the angular momentum function for the inner cut-out disk is derived in Appendix C. This expression has no limiting value as $s \to 0$ for $m = 0$. From (C.36), we obtain

$$\lim_{\tilde{s}\to 0^+} F_{l0}(\eta \neq 0) = -\frac{2-\beta}{2+\beta}\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| \frac{i}{1-e^{2\pi\hat{\eta}}}\left\{\lim_{\tilde{s}\to 0^+} e^{-i\hat{\eta}\ln\frac{l\tilde{\kappa}}{i\tilde{s}}} + \frac{(1-e^{\pi\hat{\eta}})}{N(1-e^{-\pi\hat{\eta}/N})}\right\}. \tag{6.34}$$

Taking the principal value of the logarithm, we find that

$$\ln\frac{l\tilde{\kappa}}{i\tilde{s}} = \ln\left|\frac{l\tilde{\kappa}}{\tilde{s}}\right| + i\pi\left(1+\mathrm{sign}(l)\times\frac{1}{2}\right), \tag{6.35}$$

and hence the limit in (6.34) is

$$\lim_{\tilde{s}\to 0^+} e^{-i\hat{\eta}\ln\frac{l\tilde{\kappa}}{i\tilde{s}}} = e^{\pi\hat{\eta}(1+\mathrm{sign}(l)\times\frac{1}{2})}\lim_{\tilde{s}\to 0^+} e^{-i\hat{\eta}\ln||l|\tilde{\kappa}/\tilde{s}|} \tag{6.36}$$

For very negative values of η, the $e^{\pi\eta}$ term in (6.36) ensures that the limit is approximately zero. But for positive η, the expression for F_{lm} fails to converge to any definite limit. This might be expected to present problems for numerical implementation. Fortunately, it turns out that the mathematical eigenvalue converges even though F_{lm} does not.

The following table shows the results for a particular F_{lm} as $s \to 0^+$. The results for each of the two terms inside the braces in (6.34) are shown separately. We consider an $m = 0$ perturbation in an inner cut-out disk with $\beta = 0.25$, $N = 2$ and $\gamma = 10.99$ (corresponding to $Q_\mathrm{s} = 1$). We study the angular momentum function at $\eta = 10$, $\tilde{U} = 0.5$, $l = 2$.

Growth rate s	Contribution from pole at $\ln\frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}$ $/10^{-6}$	Contribution from N poles at $(2j-1)\frac{2\pi}{N}$ $/10^{-6}$	$F_{lm}/10^{-6}$
5×10^{-1}	$+6.81052+5.69650i$	$-5.41809i$	$+6.81052+0.278414i$
10^{-1}	$+4.26127-7.31278i$	$-4.41846i$	$+4.26127-11.7312i$
10^{-2}	$-7.49074-3.90493i$	$-4.24236i$	$-7.49074-8.14729i$
10^{-3}	$-3.54807+7.66604i$	$-4.22551i$	$-3.54807+3.44052i$
10^{-4}	$+7.82449+3.18343i$	$-4.22384i$	$+7.82449-1.04041i$
10^{-5}	$+2.81173-7.96562i$	$-4.22367i$	$+2.81173-12.1893i$
10^{-6}	$-8.08910-2.43381i$	$-4.22365i$	$-8.08910-6.65746i$

The contribution from the first pole converges to a constant magnitude, but its phase fails to converge. This means that the answer for F_{lm} fails to converge. However, the largest mathematical eigenvalue for this perturbation does converge as $s \to 0$, as demonstrated below. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\min} = 30$, $l_{\max} = 40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 2.5$.]

Growth rate s	Largest mathematical eigenvalue λ
5×10^{-1}	0.469668
10^{-1}	0.776788
10^{-2}	0.938473
10^{-3}	0.974632
10^{-4}	0.984755
10^{-5}	0.988532
10^{-6}	0.990427
10^{-7}	0.991034
10^{-8}	0.991142

The mathematical eigenvalue is clearly converging to a value close to unity. This is presumably because the eigenvalue is dominated by the matrix elements along the diagonal $\eta = 0$, where the angular momentum function diverges (eq. (6.34)). Here the troublesome term (6.36) is simply unity. We now look at this convergence in more detail. We work out the eigenvalue using four different sets of numerical accuracy parameters:

Accuracy set	n	$\Delta\alpha$	$l_{\min}$	$l_{\max}$	n_{GL}	f_σ	a_{acc}	b_{acc}
Accuracy set 1	65	0.3	10	20	6	0.8	10	1.5
Accuracy set 2	125	0.2	20	30	9	0.8	10	1.5
Accuracy set 3	251	0.1	30	40	9	0.8	20	2.0
Accuracy set 4	301	0.1	40	50	9	0.8	20	2.5

The mathematical eigenvalues obtained with the different accuracy sets were:

s	Accuracy set 1	Accuracy set 2	Accuracy set 3	Accuracy set 4
10^{-1}	$.776249 + .000019i$	$.776782 + .000000i$	$.776788 + .000000i$	$.776789 + .000000i$
10^{-2}	$.938149 + .000048i$	$.938461 + .000000i$	$.938473 + .000000i$	$.938473 + .000000i$
10^{-3}	$.972224 - .000409i$	$.974504 - .000026i$	$.974632 + .000000i$	$.974632 - .000000i$
10^{-4}	$.976550 - .000768i$	$.983786 - .000137i$	$.984755 - .000000i$	$.984755 - .000001i$
10^{-5}	$.976652 - .000787i$	$.985489 - .000255i$	$.988532 - .000000i$	$.988532 - .000004i$
10^{-6}	$.976653 + .000787i$	$.985562 - .000268i$	$.990427 + .000714i$	$.990427 + .000714i$
10^{-7}	$.976653 + .000787i$	$.985561 - .000268i$	$.991059 + .000217i$	$.991059 + .000217i$

Both real and imaginary parts of the mathematical eigenvalue λ are recorded. We know that the eigenvalue should be real, so the size of $|\mathrm{Im}(\lambda)|$ is a useful clue as to the accuracy. As we expect, $|\mathrm{Im}(\lambda)|$ decreases as the accuracy parameters are made more strict. However, $|\mathrm{Im}(\lambda)|$ increases as the growth rate tends to zero, and this cannot be overcome by tightening the accuracy parameters, since there is already no difference between accuracy sets 3 and 4. We can be confident that the limiting value of the mathematical eigenvalue as $s \to 0$ is 0.991 to 3 d.p.; the finite size of $\mathrm{Im}(\lambda)$ indicates that we cannot claim higher accuracy. The eigenvalues shown in the table were calculated using EISPACK (section 5.5). They were also checked with the power method; there was no difference to within the size of $|\mathrm{Im}(\lambda)|$, except that the sign of the $\mathrm{Im}(\lambda)$ was sometimes different from that reported by EISPACK.

We also find, as reported by Toomre and Zang (1976), that the smaller eigenvalues become larger fractions of the dominant eigenvalue as s is decreased. They suggest that this effect is linked to the existence of a continuum of modes in the self-consistent disk. For example, for accuracy set 4, the first six eigenvalues reported by EISPACK for successively smaller growth rates are

$s = 0.2$	$s = 0.1$:	$s = 0.01$	$s = 0.001$	$s = 10^{-7}$
0.666429	0.776789	0.938473	0.974632	0.991059
0.665676	0.776777	0.938468	0.974632	0.991034
0.552118	0.656545	0.861037	0.934636	0.984787
0.549628	0.655977	0.861011	0.934633	0.984719
0.473906	0.568637	0.785410	0.886627	0.975259
0.469213	0.566558	0.785301	0.886614	0.975163

We also notice that the successive pairs of eigenvalues are close together, in accordance with the theoretical expectation that the eigenvectors occur in degenerate pairs.

For the self-consistent disk exposed to an axisymmetric perturbation with vanishing growth rate, the transfer function is proportional to a Dirac delta-function. Once the self-similarity has been broken by applying a cut-out, a delta-function no longer occurs in the analysis. However, the transfer function $\mathcal{S}_m(\alpha, \alpha')$ still becomes more peaked about the diagonal $\alpha = \alpha'$ as the growth rate $s \to 0$. The following contour plots show the real and imaginary parts of the $m = 0$ transfer function for an inner cut-out disk. Positive contours are drawn with solid lines, and negative with dotted. These plots also demonstrate the Hermitian nature of the transfer function. (Strictly speaking, it is the modified transfer function $T_0(\alpha, \alpha')$ (6.21) which is Hermitian, but close to the diagonal $T_0(\alpha, \alpha') \approx \mathcal{S}_0(\alpha, \alpha')$, and thus $\mathcal{S}_0\,(\alpha, \alpha') \approx \mathcal{S}_0^*\,(\alpha', \alpha)$).

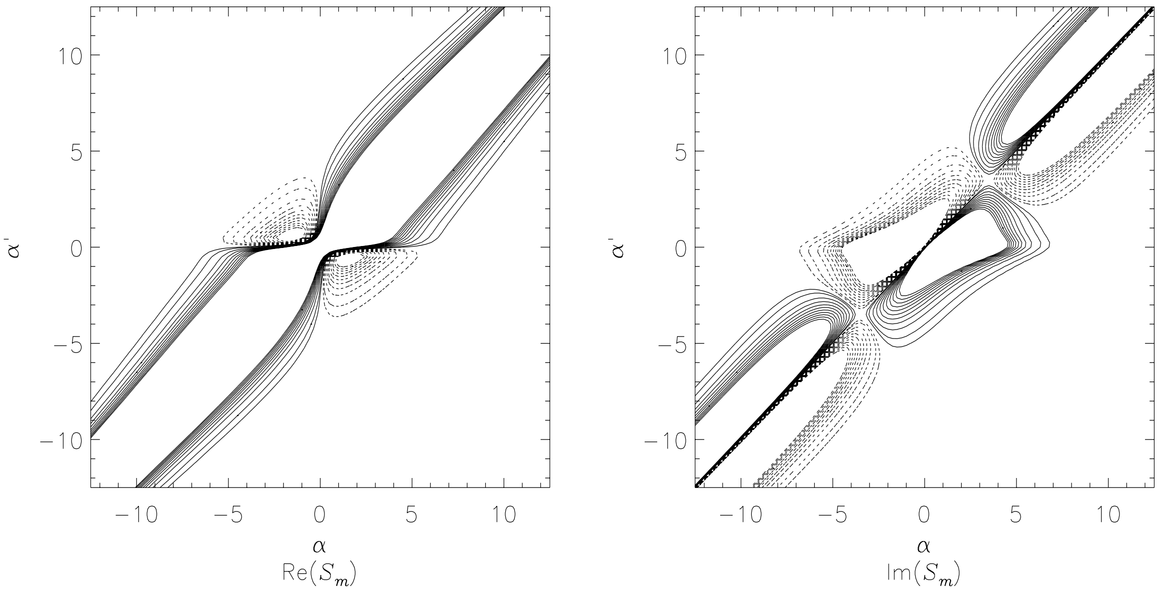


Figure 6.14: The transfer function $\mathcal{S}_0(\alpha, \alpha')$ for $s = 1.5$. Contours are drawn at $\pm 5\%, 10\%, 20\%, 30\%...100\%$ of the maximum value of $|\mathrm{Re}(\mathcal{S}_0)|$ in the left-hand plot, and of $|\mathrm{Im}(\mathcal{S}_0)|$ in the right-hand plot. Positive values are drawn with solid contours, and negative with dotted. [$\beta = 0.25$, $N = 2$, $\tilde{\sigma}_u = 0.283$, $\gamma = 11.0$, $m = 0$, $\Omega_{\mathrm{p}} = 0.5$, $s = 1.5$]

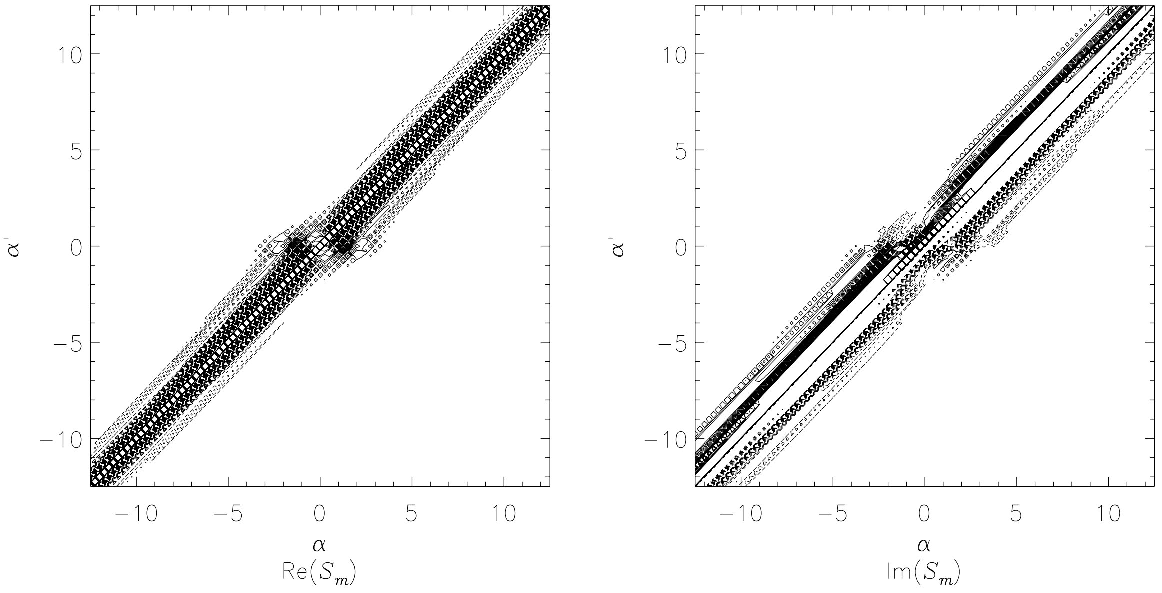


Figure 6.15: The transfer function $\mathcal{S}_0(\alpha,\alpha')$ for $s = 10^{-7}$. Contours are drawn at $\pm 5\%, 10\%, 20\%, 30\%...100\%$ of the maximum value of $|\mathrm{Re}(\mathcal{S}_0)|$ in the left-hand plot, and of $|\mathrm{Im}(\mathcal{S}_0)|$ in the right-hand plot. Positive values are drawn with solid contours, and negative with dotted. [$\beta = 0.25$, $N = 2$, $\tilde{\sigma}_u = 0.283$, $\gamma = 11.0$, $m = 0$, $\Omega_\mathrm{p} = 0.5$, $s = 10^{-7}$]

We see that for vanishing growth rate, most contributions come from small values of η, where the imposed and response wavenumbers are similar. In the expression for the angular momentum function for the inner cut-out disk (C.36), we can then write

$$N(1 - e^{-\pi\hat{\eta}/N}) \approx \pi\hat{\eta}. \tag{6.37}$$

The dependence on the inner cut-out function N is cancelled out in the matrix elements close to the diagonal, where the transfer function is largest. We thus expect the stability of the inner cut-out disk to be virtually independent of N.

For the doubly cut-out disk, the angular momentum function is given by (C.41). It is straightforward to take the limit $m = 0$, $s \to 0$ and obtain:

$$\begin{aligned} F_{lm}(\eta \neq 0) = \frac{2-\beta}{2+\beta}\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| \frac{i}{(1-e^{2\pi\hat{\eta}})} \times \Bigg\{ \frac{\tilde{L}_\mathrm{c}^{M_\beta}}{N} \sum_{j=1}^{N} \frac{\exp\left\{\frac{2j-1}{N}\pi\hat{\eta}\right\}}{\left[\exp\left\{(2j-1)\frac{M}{N}i\pi\right\} + \tilde{L}_\mathrm{c}^{M_\beta}\right]} \\ - \frac{\tilde{L}_\mathrm{c}^{N_\beta}}{M} \exp\left\{-i\hat{\eta}\frac{2+\beta}{2-\beta}\ln\tilde{L}_\mathrm{c}\right\} \sum_{k=1}^{M} \frac{\exp\left\{\frac{2k-1}{M}\pi\hat{\eta}\right\}}{\left[e^{-(2k-1)\frac{N}{M}i\pi} + \tilde{L}_\mathrm{c}^{N_\beta}\right]} \Bigg\}. \end{aligned} \tag{6.38}$$

Unless $\tilde{L}_\mathrm{c}$ is very large, no such cancellation of N and M occurs. We therefore expect the results for the doubly cut-out disk to depend on the cut-out indices.

6.6 Marginal modes

Having convinced ourselves that we can find the mathematical eigenvalue to good accuracy, even in the rather strange limit $s \to 0$, we proceed to consider the marginal stability of the cut-out disks. First we look in general terms at how the mathematical eigenvalue depends on growth rate and temperature.

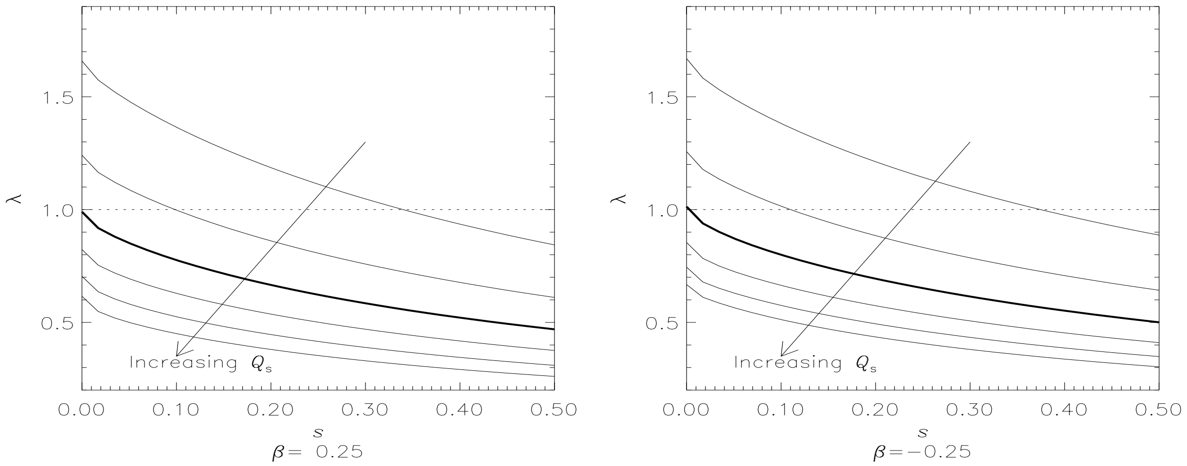


Figure 6.16: Dependence of the largest mathematical eigenvalue on growth rate s and temperature $Q_{\rm s}$. Curves are drawn for six different values of $Q_{\rm s}$: $Q_{\rm s} = 0.6$, 0.8, 1.0, 1.2, 1.4, 1.6. The $Q_{\rm s} = 1$ curve is drawn with a bolder line. For the $\beta = 0.25$ disk, $Q_{\rm s} = 1$ corresponds to a velocity dispersion of $\tilde{\sigma}_u = 0.283$, and for $\beta = -0.25$, to $\tilde{\sigma}_u = 0.509$. The curves are for an inner cut-out disk with $N = 2$.

For sufficiently small $Q_{\rm s}$, we see that a unit eigenvalue can always be found for sufficiently large s: when the disk is cool, it is susceptible to fast-growing modes. As the disk is heated, the growth rate of the unstable mode decreases. When the disk is sufficiently hot, the disk is stable and no modes exist. The temperature at which the disk is marginally stable is given by the $Q_{\rm s}$ for which the $s = 0^+$ eigenvalue is unity. We see that this value of $Q_{\rm s}$ is close to unity, in accordance with our argument that the cut-out disks should have stability properties close to those of their self-consistent counterparts. We investigated the dependence of the marginal stability upon β by setting $s = 10^{-7}$ and using the Newton-Raphson method to find the velocity dispersion $\tilde{\sigma}_{u,\rm min}$ necessary for a self-consistent mode.

The inner cut-out disks

Tables of the marginal modes are given in Appendix E. The results are shown in fig. 6.17. The left-hand plot shows the velocity dispersion $\tilde{\sigma}_{u,\rm min}$ as a function of β. The results from the present global calculation for an inner cut-out disk are shown with a solid line. Four different values of N were used, but they made so little difference to the results that the curves are scarcely distinguishable. For comparison, the results for the self-consistent disk are also shown (with dotted and dashed lines for local and global results respectively), reproducing fig. 6.12.

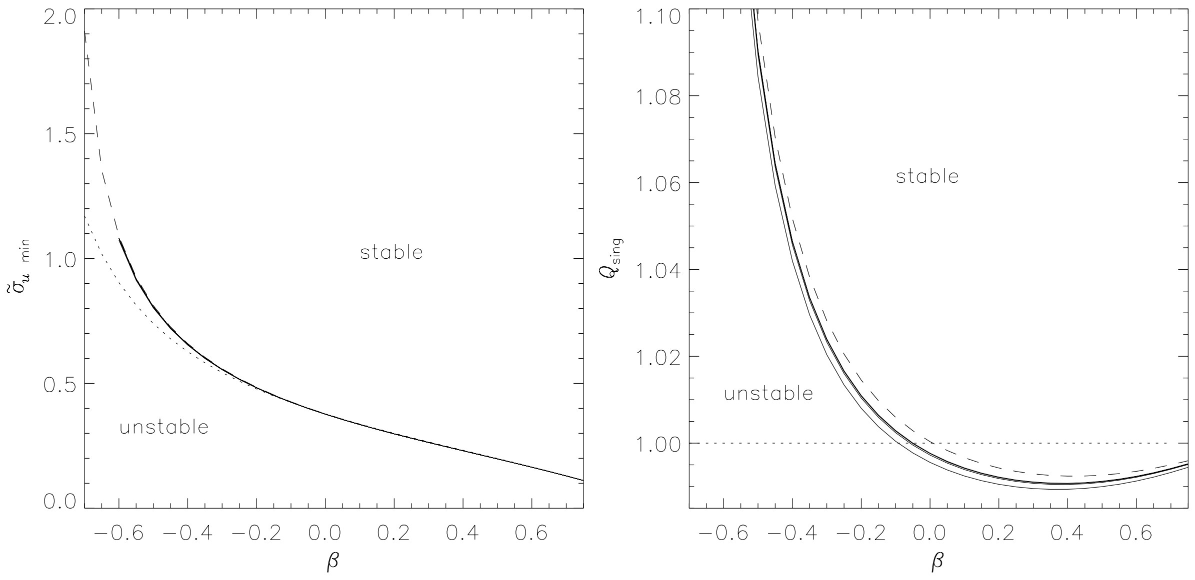


Figure 6.17: Minimum temperature for axisymmetric stability plotted against β, for disks with various cut-out functions. The left-hand plot shows the minimum velocity dispersion $\tilde{\sigma}_u$; the right-hand one shows the same data in terms of the stability parameter $Q_{\rm s}$. The solid lines indicate results for an inner cut-out disk. Results for $N = 1$, 2, 3, 4 are plotted in both graphs, but coincide so closely as to be virtually indistinguishable in the left-hand graph. The dotted and dashed lines show the results for the self-consistent disk derived from local and global theory respectively. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\rm min} = -40$, $l_{\rm max} = +50$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$, $\epsilon_\sigma = 10^{-7}$.]

As expected, the inner cut-out disks have almost exactly the same stability as the self-consistent disk. This is in accordance with Toomre and Zang's finding that the stability criteria for the self-consistent and $N = 2$ cut-out disks agreed "to within a few tenths of one percent" (Zang 1976). The inner cut-out does have a slight effect on the global stability, which is more easily seen when the results are expressed in terms of $Q_{\rm s}$. This is the ratio of the velocity needed for global axisymmetric stability in the present cut-out disk to that needed for local axisymmetric stability in the corresponding self-consistent disk (see eq. (6.15)). This is shown in the right-hand plot of fig. 6.17. The dotted line shows $Q_{\rm s} = 1$, which is by definition the result for local theory in the self-consistent disk. The dashed line shows the result of a global calculation in the self-consistent disk, and the solid lines those of a global calculation in the cut-out disks. For both rising and falling rotation curves, the inner cut-out function tends to make the disk slightly more stable, in that the temperature can be lowered slightly further before axisymmetric instabilities set in. As expected, the value of the cut-out index has little effect. The lowest curve in the right-hand plot of fig. 6.17 is for $N = 1$; $N = 2$, 3, 4 are practically indistinguishable.

For the self-consistent disk, the neutral modes are single log-spirals. In accordance with this, the density transforms for the cut-out disks are strongly peaked. Fig. 6.18 shows density transforms of marginal modes in the $N = 2$ disks with $\beta = \pm 0.25$. Also shown are the most unstable wavenumbers for the corresponding self-consistent disks. We define the dominant wavenumber to be that at which the amplitude of the density transform $A(\alpha)$ is maximum. We then define the most unstable wavenumber $\alpha_{\rm u}$ for the cut-out disks to be the dominant wavenumber of the marginal mode. Fig. 6.18 suggests that this wavenumber is typically close to the most unstable wavenumber of the corresponding self-consistent disk.

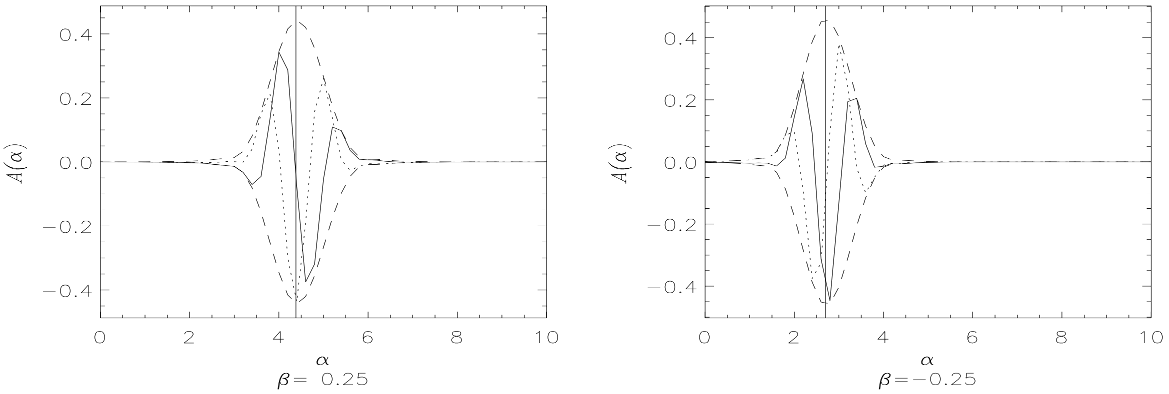


Figure 6.18: Density transforms $A(\alpha)$ for marginal modes in the inner cut-out disk with $N = 2$. The left-hand plot is for $\beta = 0.25$ and the right-hand one for $\beta = -0.25$. In each case, the solid line is $\mathrm{Re}(A(\alpha))$, the dotted line is $\mathrm{Im}(A(\alpha))$ and the dashed line is the envelope $\pm|A(\alpha)|$. The vertical line marks the α_u for the corresponding self-consistent disk. For $\beta = 0.25$, $\alpha_\mathrm{u} = 4.38$, $\tilde{\sigma}_{u,\mathrm{min}} = 0.280$, $\gamma_\mathrm{max} = 11.2$; for $\beta = -0.25$, $\alpha_\mathrm{u} = 2.70$, $\tilde{\sigma}_{u,\mathrm{min}} = 0.517$, $\gamma_\mathrm{max} = 3.24$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -30$, $l_\mathrm{max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 1.5$, $\epsilon_\sigma = 10^{-7}$.]

Fig. 6.19 confirms this. The solid lines shows the dominant wavenumber plotted against β for various inner cut-out disks. The dotted line shows the most unstable wavenumber for the self-consistent disk, reproducing fig. 6.13. The results for the self-consistent and cut-out disks are very similar across the full range of β. Note that our method of solving the integral equation involves discretising in wavenumber. The α are thus available to the nearest $\Delta\alpha$, which accounts for the slightly wobbly appearance of the curves for the cut-out disks.

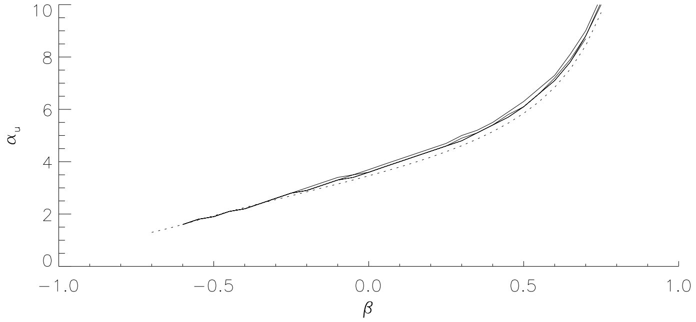


Figure 6.19: Most unstable wavenumber α_u plotted against β, for disks with various inner cut-out functions. The solid lines show the most unstable wavenumber for inner cut-out disks with $N = 1$, 2, 3 and 4. The dotted line shows the most unstable wavenumber for the self-consistent disk. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.1$, $l_\mathrm{min} = -30$, $l_\mathrm{max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 1.5$, $\epsilon_\sigma = 10^{-7}$.]

The doubly cut-out disks

Tables of the marginal modes for doubly cut-out disks are given in Appendix E. Twelve disks were studied: those with inner cut-out index $N = 1, 2, 3, 4$, in each case with outer cut-out index $M = 2, 4, 6$, and $\tilde{R}_\mathrm{c} = 100$. The results are summarised in fig. 6.20. The presence of the outer cut-out does not change the basic shape of the curve. As expected from the form of the angular momentum function (6.38), there is more dependence on the value of the cut-out indices N and M than there was for the inner cut-out disk in fig. 6.17. Again, it is helpful to consider this data in terms of Toomre's stability parameter Q_s. We compare the velocity dispersion needed for global stability in the various cut-out disks with that needed for local stability in the self-consistent disk. This ratio Q_s is shown in the right-hand plot of fig. 6.20. The results obtained previously for the self-consistent disk are also shown.

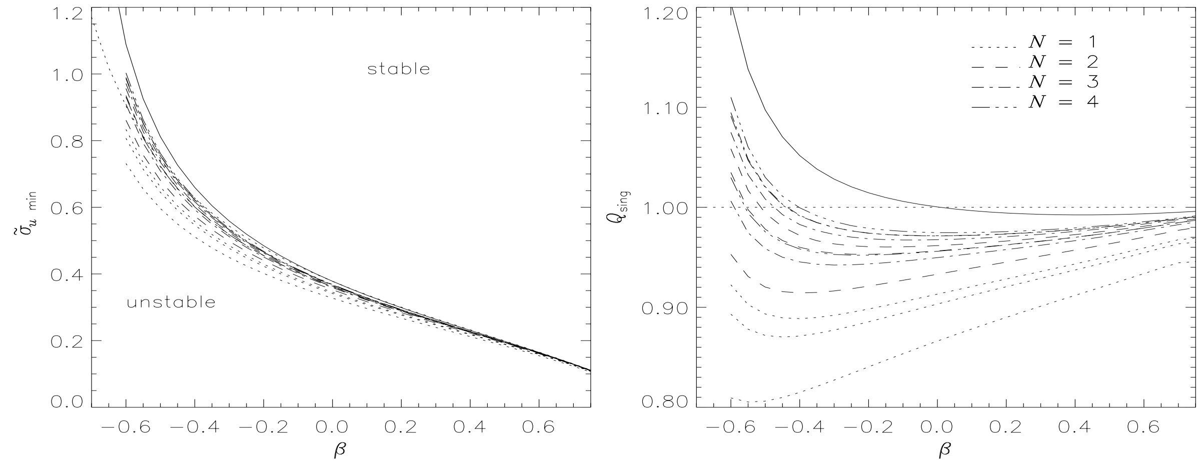


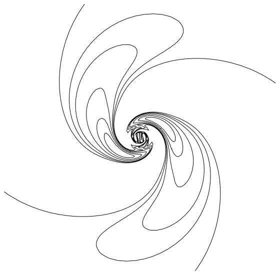
Figure 6.20: Minimum temperature for axisymmetric stability plotted against β, for disks with various cut-out functions. The left-hand plot shows the minimum velocity dispersion $\tilde{\sigma}_u$; the right-hand one shows the same data in terms of the stability parameter Q_s. The solid line shows the global, and the dotted line the local, stability of the self-consistent disk. The broken lines are for various doubly cut-out disks. The inner cut-out indices are $N = 1$, 2, 3, 4, with in each case $M = 2$, 4, 6, $\tilde{R}_\mathrm{c} = 100$. In the $\tilde{\sigma}_{u,\mathrm{min}}$ plot, they coincide too closely to be easily distinguished. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.1$, $l_\mathrm{min} = -40$, $l_\mathrm{max} = +50$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 1.5$, $\epsilon_\sigma = 10^{-7}$.]

The additional outer cut-out function tends to make the disk more stable, in that less heat is needed to stabilise the doubly cut-out disk. The outer cut-out has most effect for rising rotation curves. It is interesting that the presence of the outer cut-out makes the inner cut-out more significant. Whereas for the inner cut-out disk, the value of N had little effect, here N has a noticeable effect on the stability. All the cut-out disks are more stable than the self-consistent disk, but those where the cut-out is gentle (i.e. low N) are more stable than those with a sharper cut-out (high N). A similar effect was seen with the outer cut-out. For each N, the curve with $M = 6$ lies above the curve with $M = 4$ which lies above the curve with $M = 2$, so the more gently tapered disks are more stable.

6.7 Summary

In this Chapter, we have studied the stability of the power-law disks to axisymmetric modes. We first considered the local axisymmetric stability, and found that disks with rising rotation curves are inherently less stable than those with rising rotation curves. We then studied the global stability of both the self-consistent and the cut-out power-law disks. We found that the presence of a cut-out made very little difference to the stability of the disk. Local theory is a remarkably accurate predictor of global axisymmetric stability.

Chapter 7
Global Bisymmetric Modes



7.1 Introduction

In this chapter, we investigate the stability of the power-law disks to bisymmetric perturbations. The situation is more complicated than for the axisymmetric perturbations we examined in Chapter 6. The pattern can now rotate, introducing an extra degree of freedom into the problem. This is reflected in the fact that the mathematical eigenvalue is now (in general) complex, rather than purely real as for axisymmetric perturbations.

7.2 Wave mechanics

Disks sustaining self-consistent modes are like a resounding gong or a clanging cymbal. When a gong is struck, a standing wave pattern is excited in it. This pattern can be regarded as being made up of many travelling waves, which reflect off the boundaries of the gong. The travelling waves superpose in such a way as to create a relatively persistent standing wave pattern. A similar effect occurs in galactic disks. We can regard our global mode as being made up of the combined effects of many waves propagating through the disk. These waves may be leading or trailing, travelling inwards or outwards. They reflect off the inner cut-out and the Lindblad resonances in much the same way as waves in the gong reflect off its edges.

The behaviour of these waves may be investigated analytically if we make the assumption that the waves are tightly-wound, i.e. $\alpha >> 1$. This method is also known as the WKB approximation, because of its similarities to the Wentzel-Kramers-Brillouin approximation in quantum mechanics. In the case of tight winding, successive spiral arms are close together, and long-range gravitational effects are negligible in comparison. Thus the WKB approach is a local theory, analogous to that discussed in section 6.2 for axisymmetric perturbations. However, we shall see that the local approach is very much less successful in predicting the global response of the disk to non-axisymmetric than to $m = 0$ perturbations.

Lin & Shu (1966) and Kalnajs (1965) independently derived a dispersion relation relating the frequency of tightly-wound waves to their wavenumber α.

$$\frac{m^2(\Omega_0 - \Omega_{\rm p})^2}{\kappa_0^2} = 1 - \frac{2\pi G\Sigma}{\kappa_0^2}\frac{|\alpha|}{R}\mathcal{F} \tag{7.1}$$

$\mathcal{F}$ is the *reduction factor* describing how the effects of the spiral perturbation are lessened for stars with random motion in the radial direction (see Appendix D). Eq. (7.1) contains the quantity $\varpi = m(\Omega_{\rm p} - \Omega_0)/\kappa_0$. This has a simple physical interpretation. $m(\Omega_{\rm p} - \Omega_0)$ is the frequency at which a star on a circular orbit encounters successive maxima of the perturbing potential. ϖ is the ratio of this forcing frequency to the natural radial frequency of the star, κ_0. For the power-law disks, this ratio is

$$\varpi = \frac{m(\Omega_{\rm p} - \Omega_0)}{\kappa_0} = \frac{m}{\sqrt{2-\beta}}\left\{\tilde{\Omega}_{\rm p}\left(\frac{R}{R_0}\right)^{\frac{2+\beta}{2}} - 1\right\} \tag{7.2}$$

The case $\varpi = 0$ corresponds to co-rotation, when the pattern speed of the waves is matched to the circular velocity of stars at this radius. The radii at which $|\varpi| = 1$ are the *Lindblad resonances.* These represent the positions at which the forcing frequency is equal to the natural frequency. We naturally expect the response of the disk to be greatest at these radii. As we shall see, in a very hot stellar disk stationary disturbances can exist only at the Lindblad resonances. However, as the disk is cooled, gravitational interactions between the stars become more important, and reduce the natural radial frequency below κ_0. This means that for a sufficiently cold disk, waves can exist when $|\varpi| < 1$. In a stellar disk, no mechanisms exist to increase the natural radial frequency above κ_0, and so no modes can exist where $|\varpi| > 1$. Eq. (7.3) gives the positions of the three resonances for the power-law disks.

$$\frac{R_{\rm ILR}}{R_0} = \left(\frac{m - \sqrt{2-\beta}}{m\tilde{\Omega}_{\rm p}}\right)^{\frac{2}{2+\beta}}, \qquad \frac{R_{\rm OLR}}{R_0} = \left(\frac{m + \sqrt{2-\beta}}{m\tilde{\Omega}_{\rm p}}\right)^{\frac{2}{2+\beta}}, \qquad \frac{R_{\rm CR}}{R_0} = \tilde{\Omega}_{\rm p}^{\frac{-2}{2+\beta}}. \tag{7.3}$$

Lin & Shu (1966) calculated the reduction factor using the Schwarzschild distribution function (D.27). In Appendix D, the reduction factor is calculated for our power-law distribution function (given by eqs. (2.54) and (2.81)). Fig. 7.1 compares the reduction factors obtained using the Schwarzschild distribution function (solid line), the self-consistent power-law distribution function (2.54) (dotted line) and a cut-out distribution function (2.81) (dashed line). The different distribution functions cause only very slight differences in the reduction factor. The Schwarzschild reduction factor is numerically much faster to evaluate, since two out of the three necessary integrations can be carried out analytically (see Appendix D). In the following discussion, we therefore use the Schwarzschild reduction factor, confident that this will not qualitatively affect our results.

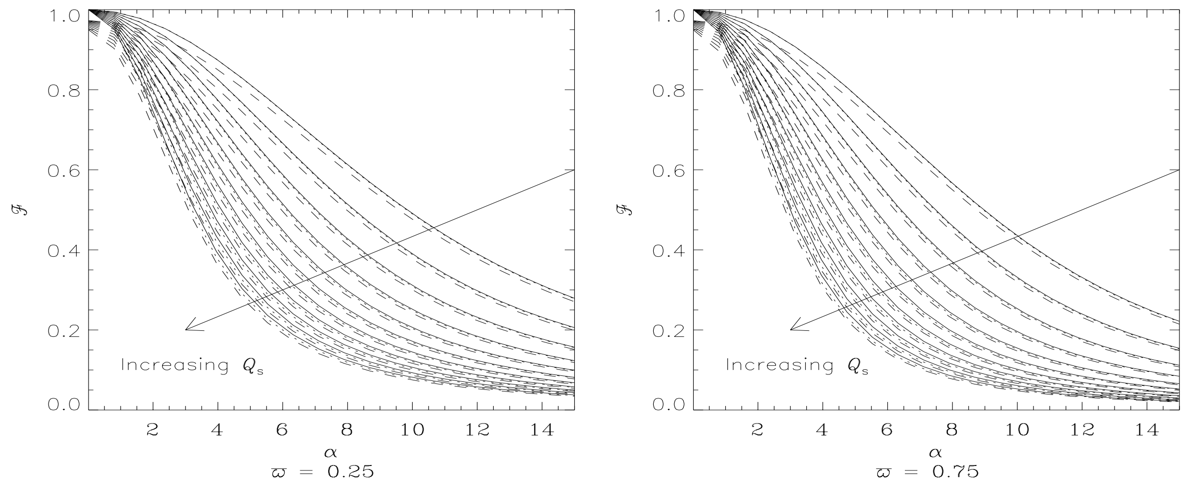


Figure 7.1: The reduction factor obtained with three different distribution functions. The reduction factor $\mathcal{F}$ is plotted against wavenumber α for several different temperatures Q_{s}. In the left-hand plot, $\varpi = 0.25$; in the right-hand plot $\varpi = 0.75$. In each case, $\mathcal{F}$ is drawn for three different distribution functions. The solid line shows the reduction factor obtained with the Schwarzschild distribution function, the dotted line that obtained with the self-consistent power-law distribution function, and the dashed line that obtained with an inner cut-out distribution function with $N = 2$. 11 values of Q_{s} are shown, from $Q_{\mathrm{s}} = 0.5$ to $Q_{\mathrm{s}} = 1.5$ in steps of 0.1. $\beta = 0.25$.

When the Schwarzschild distribution function is used, it is convenient to write the dispersion relation (7.1) in terms of the ratio ϖ, and the general variables x, y (6.8) and α_{crit} (6.2), which we introduced in section 6.2 when discussing the local stability to axisymmetric perturbations. x expresses the wavenumber of the density wave as a multiple of α_{crit}, the wavenumber below which a cold disk is stable to axisymmetric perturbations. y contains the dependence on temperature.

The Schwarzschild reduction factor is given by (Lin & Shu 1966; Binney & Tremaine 1987, app. 6.A; eq. (D.28))

$$\mathcal{F}_{\blacktriangledown} = \frac{1-\varpi^2}{\sin\pi\varpi}\int_0^{\pi} d\chi\, e^{-\zeta(1+\cos\chi)}\sin\varpi\chi\sin\chi, \tag{7.4}$$

where $\zeta = x^2y^2$. The dispersion relation (7.1) can then be written simply as

$$\varpi^2 = 1 - |x|\mathcal{F}_{\blacktriangledown}(\varpi, x^2y^2). \tag{7.5}$$

This may be solved to yield curves of ϖ against x for different values of y. Fig. 7.2 shows such curves. The different values of y are expressed in terms of Toomre's stability parameter Q_{s}. As we saw in section 6.2, the criterion for local axisymmetric stability, $Q_{\mathrm{s}} = 1$, corresponds to $y = 3.3582830/2\pi$. For all temperatures, the solutions fall into two classes: the long-wave solutions at small α, where ϖ falls as wavenumber increases, and the short-wave solutions at large α, where ϖ increases with wavenumber. The long-wave solutions are not trustworthy, since they violate the tight-winding approximation on which this treatment depends.

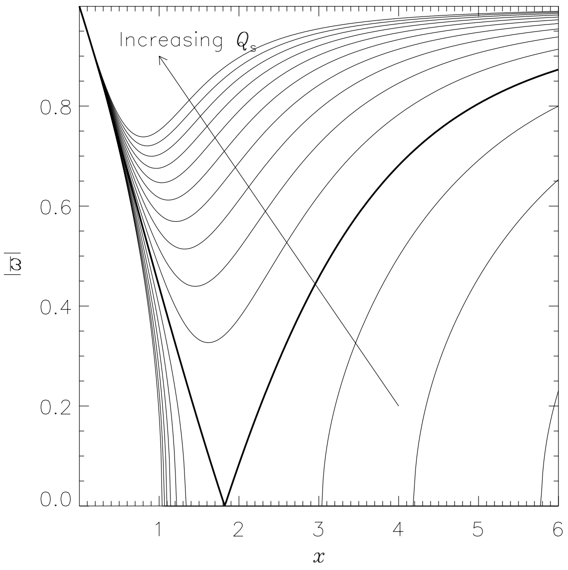


Figure 7.2: The dispersion relation for tightly-wound waves in general disks. The ratio ϖ of the forcing frequency to the natural radial frequency of the star is plotted against x, the ratio of the wavenumber to the critical wavenumber. Curves are drawn for 17 values of Toomre's stability parameter Q_s, from $Q_\mathrm{s} = 0.4$ to $Q_\mathrm{s} = 2.0$ in steps of 0.1. The curve with $Q_\mathrm{s} = 1$ is drawn with a bolder line.

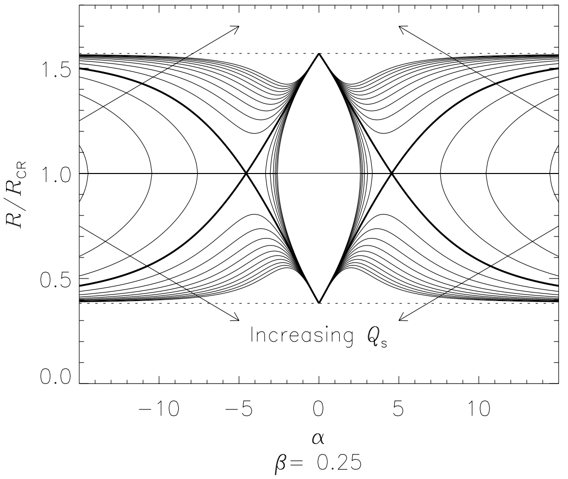


Figure 7.3: The dispersion relation for tightly-wound waves in a power-law disk. The position of the wave in the disk is plotted against wavenumber for $m = 2$ waves in the $\beta = 0.25$ disk. Both leading and trailing waves, inside and outside co-rotation radius, are shown. Curves are drawn for 17 values of Toomre's stability parameter Q_s, from $Q_\mathrm{s} = 0.4$ to $Q_\mathrm{s} = 2.0$ in steps of 0.1. The arrows indicate the direction of increasing Q_s on each branch. The curves with $Q_\mathrm{s} = 1$ are drawn with a bolder line. The solid horizontal line marks co-rotation, and the dotted lines the Lindblad resonances.

The case $Q_\mathrm{s} = 1$ – drawn with a bolder line in fig. 7.2 – demarcates two different behaviours. For Q_s greater than 1, there are solutions at all wavenumbers, but none close to co-rotation. So for hot disks, waves of all wavenumbers are possible, but there is a "forbidden region" around the co-rotation radius. For Q_s less than 1, there is no forbidden region: waves can exist anywhere between the Lindblad resonances. However, there are wavenumbers at which no solutions exist. The short- and long-wave solutions are not continuous.

Using the expressions for the circular and epicyclic frequency (2.41), we can derive the pattern speed and wavenumber of waves in the power-law disks. The value of the pattern speed depends on where the wave is in the disk: we have (7.2)

$$\tilde{\Omega}_\mathrm{p} = \left(\frac{R_0}{R}\right)^{\frac{2+\beta}{2}} \left(\frac{m + \varpi\sqrt{2-\beta}}{m}\right). \tag{7.6}$$

Alternatively, we can express the position of the wave in the disk in terms of pattern speed. The position of the wave is given in terms of the co-rotation radius (7.3) by

$$\frac{R}{R_\mathrm{CR}} = \frac{R}{R_0}\tilde{\Omega}_\mathrm{p}^{2/(2+\beta)} = \left(\frac{m + \varpi\sqrt{2-\beta}}{m}\right)^{\frac{2}{2+\beta}}. \tag{7.7}$$

Similarly, the wavenumber is given by

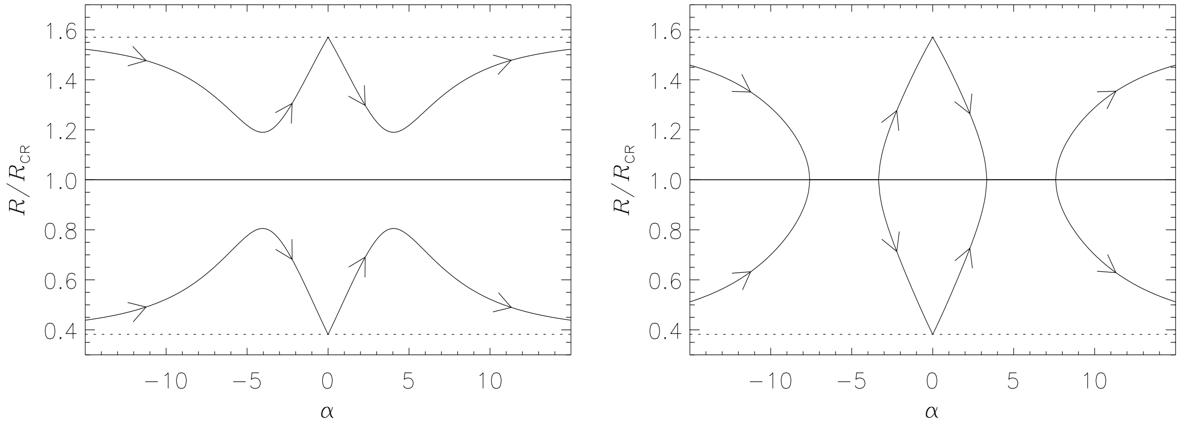
$$\alpha = (2-\beta)x\frac{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}. \tag{7.8}$$

Eq. (7.5) is independent of the sign of ϖ and x (since the Schwarzschild reduction factor has the symmetry $\mathcal{F}_{\blacktriangledown}(\varpi, \zeta) = \mathcal{F}_{\blacktriangledown}(-\varpi, \zeta)$). Solutions therefore come in quartets with $\pm\varpi, \pm x$. The sign of the wavenumber distinguishes leading ($\alpha < 0$) from trailing waves ($\alpha > 0$). From eq. (7.7), we see that the sign of ϖ distinguishes waves on either side of co-rotation. Fig. 7.3 shows the position of the wave in the disk, relative to co-rotation, plotted against the wavenumber α. All four solutions (leading/trailing, inside/outside co-rotation) are shown. As in fig. 7.2, curves are drawn for disks of several different temperatures.

This figure illustrates that the waves are confined between the Lindblad resonances. In disks with $Q_{\rm s} > 1$, the forbidden region confines the waves either inside or outside co-rotation. The hotter the disk, the more narrowly the wave is confined. In very hot disks, waves can exist only at the Lindblad resonances. In cooler disks, the waves can exist anywhere in the region between the Lindblad resonances. To study the propagation of the waves, we need to know their direction of travel. The waves travel through the disk with a group velocity (Toomre 1969; Binney & Tremaine 1987)

$$v_{\rm g} = \frac{\partial \omega(k, R)}{\partial k} = m v_\beta \frac{R}{R_0} \frac{\partial \tilde{\Omega}_{\rm p}(\alpha, R)}{\partial \alpha} = m v_\beta \left(1 + \frac{\beta}{2}\right) \left(\frac{R_0}{R_{\rm CR}}\right)^{\beta/2} \frac{\partial}{\partial \alpha} \left(\frac{R}{R_{\rm CR}}\right). \qquad (7.9)$$

We can therefore deduce the sign of the group velocity from fig. 7.3. The group velocity is directed outwards when the curves slope upwards, and inwards when they slope downwards. Fig. 7.4 shows waves for $Q_{\rm s} > 1$, $Q_{\rm s} < 1$, with the direction of travel indicated with arrows.



$Q_{\rm s} = 1.1$ $Q_{\rm s} = 0.9$

Figure 7.4: The dispersion relation for tightly-wound waves in a power-law disk with $\beta = 0.25$. The position of the wave in the disk is plotted against wavenumber for $m = 2$ waves in disks with $Q_{\rm s} = 1.1$ and 0.9. Both leading and trailing waves, inside and outside co-rotation radius, are shown. The arrows indicate the direction of the group velocity on each branch. The solid horizontal line marks co-rotation, and the dotted lines the Lindblad resonances.

We first consider disks with $Q_{\rm s} > 1$, as in the left-hand plot of fig. 7.4. Here, a packet of leading waves initially close to the inner Lindblad resonance travels outwards towards co-rotation, becoming less tightly-wound as they do so. The packet is "reflected" from the forbidden region around co-rotation, and propagates back towards the centre again. The WKB theory indicates that the packet is reflected by the inner Lindblad resonance, having changed the sense of winding to become trailing.[1] Once again, the waves reflect off the forbidden zone. They travel back towards the inner Lindblad resonance, becoming ever more tightly wound. Mark (1974) showed that they are eventually absorbed at the inner Lindblad resonance.

In disks with $Q_{\rm s} < 1$, the waves fall into two distinct classes: the long-wave ($|\alpha| \lesssim 1$) and short-wave ($|\alpha| >> 1$) solutions. Leading short-wave packets move towards co-rotation, whether they are initially inside or outside the co-rotation radius. They unwind as they travel, but never change from leading to trailing. The amount of unwinding depends on the temperature: at co-rotation the leading waves are looser in warmer disks. Trailing short-wave packets move away from co-rotation, becoming more tightly-wound as they do so. They end up at the inner or outer Lindblad resonances, depending on which side of co-rotation they were initially. This gives us another insight into why cool disks are less stable. In hot disks, waves inevitably end up as tightly-wound trailing waves, which are absorbed at the Lindblad resonances. In cold disks, the lacunae in allowed wavenumber separate the different solutions. This prevents the differential rotation from twisting all waves into tightly-wound trailing waves.

Fig. 7.3 suggests that leading long-wave packets move away from the co-rotation radius towards one of the Lindblad resonances. There they are reflected back towards co-rotation as loosely-wound trailing waves. However, the long-wave solutions do not satisfy the WKB approximation, and so we cannot rely on fig. 7.3 as an accurate guide to their behaviour. The inability of the WKB approach to explore the behaviour of the long-wave branch is a serious deficiency. The long-wave branch turns out to be of crucial importance in understanding the global behaviour of the disks. As the long-wave leading waves change over into trailing waves, their amplitude is greatly enhanced, an effect known as *swing amplification*. This mechanism was noticed independently by Goldreich & Lynden-Bell (1965a) and Julian & Toomre (1966), and reviewed and revived by Toomre (1981). It depends on a brief match in velocities between a spiral arm as it unwinds from leading to trailing and the epicyclic motion of individual stars. This can greatly enhance the perturbing effect of the density wave on the star's orbit, and lead to rapid growth of the amplitude of the wave as it unwinds. For swing amplification to be effective, the disk must be reasonably cool. High values of $Q_{\rm s}$ make the disk insensitive to the gravitational interactions between stars on which swing amplification depends. Further, swing amplification only functions when the arclength between adjacent spiral arms, $2\pi R/m$, is sufficiently small relative to the critical wavelength $\lambda_{\rm crit}$ at which Jeans instabilities first set in. The ratio of these two quantities is the parameter X (Toomre 1981; Binney & Tremaine 1987). For the power-law disks, this parameter is

$$X = \frac{\alpha_{\rm crit}}{m} = \frac{2-\beta}{m} \frac{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}. \tag{7.10}$$

The critical wavenumber $\alpha_{\rm crit}$ was plotted in fig. 6.2 on logarithmic axes for $|\beta| < 1$. Fig. 7.5 shows the

---

[1] Of course, the tight-winding approximation on which this treatment is based breaks down as the waves unwind from leading to trailing. It therefore cannot be relied upon to trace the evolution of the waves at this point. However, a more rigorous treatment (Goldreich & Tremaine 1978) indicates that the behaviour implied by the tight-winding approximation does occur provided that the parameter X, defined in eq. (7.10), is sufficiently large.

parameter X in the range of interest, $|\beta| < 0.5$. A rule of thumb is that swing amplification is powerful when $X \lesssim 3$ (Toomre 1981). This condition is satisfied throughout the range of β.

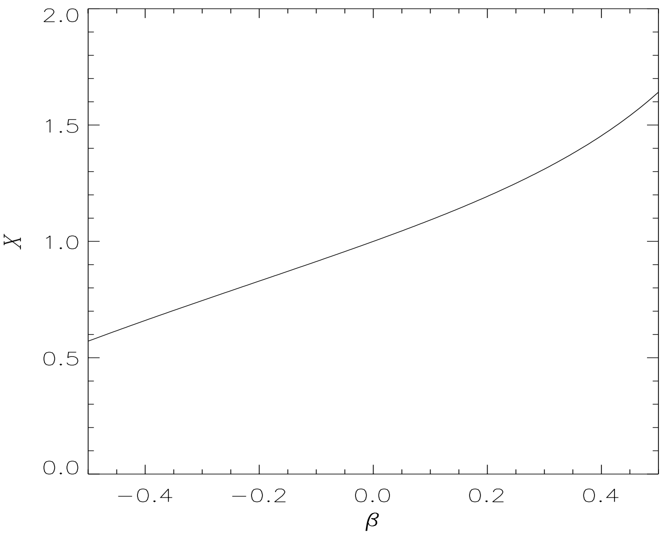


Figure 7.5: The dependence of the parameter $X = \alpha_{\rm crit}/m$ on β, for $m = 2$.

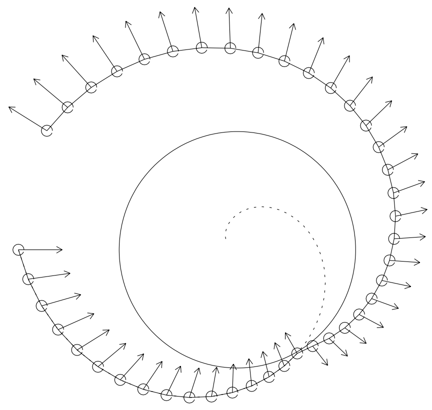


Figure 7.6: A model incoming trailing wave being reflected from the inner cut-out. Particles initially travelling inwards with velocity v are reflected from the inner cut-out. The dotted line shows where the particles would have been in the absence of a cut-out. In this sketch, the effect of differential rotation is not included.

Swing amplification on its own is not sufficient to make the disk unstable. Leading waves are greatly amplified as they are unwound into trailing waves, but the trailing waves then travel in towards the inner Lindblad resonance, becoming ever more tightly-wound, and are absorbed. We need a method of stoking the furnace, a feedback loop returning trailing waves as leading waves which can be swing-amplified. In our disks, this is provided by the inner cut-out. Without a detailed analysis, we can see that incoming trailing waves are reflected from the inner cut-out as outgoing leading waves. Fig. 7.6 sketches how this happens. It shows model "density waves", consisting of a set of particles travelling radially inwards with velocity v. The relative positions of the particles are arranged so that they lie on a log-spiral wavefront, with radius $R \propto \exp[(2\pi - m\theta)/\alpha]$. When the particles reach the inner cut-out, the direction of their velocity is reversed. Fig. 7.7 traces a set of incoming trailing wavefronts as they are reflected into outgoing leading waves.

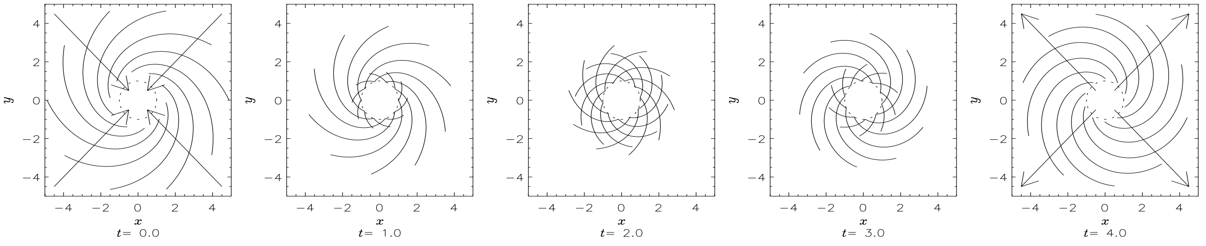


Figure 7.7: Incoming trailing waves being reflected from the inner cut-out. Several wavefronts are shown at successive time-steps. Incoming trailing waves are reflected from the inner cut-out (dotted circle) as outgoing leading waves. The arrows indicate the direction of propagation of the waves. The effect of differential rotation is included. The radial speed with which the waves propagate is chosen to be equal to the circular velocity at the cut-out. The unit of time is chosen to be the time taken to travel from the cut-out to the origin.

In terms of fig. 7.4, the feed-back loop provided by the inner cut-out turns the flat plots into tori, so that as waves leave the diagram on the right, they re-enter on the left. Of course, the inner cut-out is not a perfect mirror, but rather a beam-splitter. The reflectivity depends on the inner cut-out index. When N is small, the density departs only gradually from the self-similar case (fig. 2.13), and most of the incoming wave propagates straight through the cut-out radius $R = R_0$. When N is larger ($\gtrsim 2$), the abrupt discontinuity in the density presents a much more effective barrier to the waves' propagation. Almost all incoming waves are then reflected out, forming an effective feed-back loop. In doubly cut-out disks, the outer cut-out acts as a second reflector. Sketches similar to those in fig. 7.7 indicate that outgoing trailing waves return from an outer reflective boundary as incoming leading waves. However, the outer cut-out involves considerably gentler changes to the discontinuity than the inner cut-out (cf. figs. 2.13 and 2.16), and is thus an inefficient reflector. Accordingly, we shall see that the presence of an outer cut-out has a minimal effect on the stability of the disk.

The stability of the disk depends critically on the relative positions of the inner Lindblad resonance and the inner cut-out. If the inner Lindblad resonance lies outside the inner cut-out, it stabilises the disk, both directly by absorbing incoming trailing waves, and indirectly by preventing feedback to the swing amplifier. The position of the inner cut-out controls whether the waves are reflected as leading waves which can be swing-amplified, or absorbed as trailing waves by the inner Lindblad resonance. The situation is summarised in fig. 7.8.

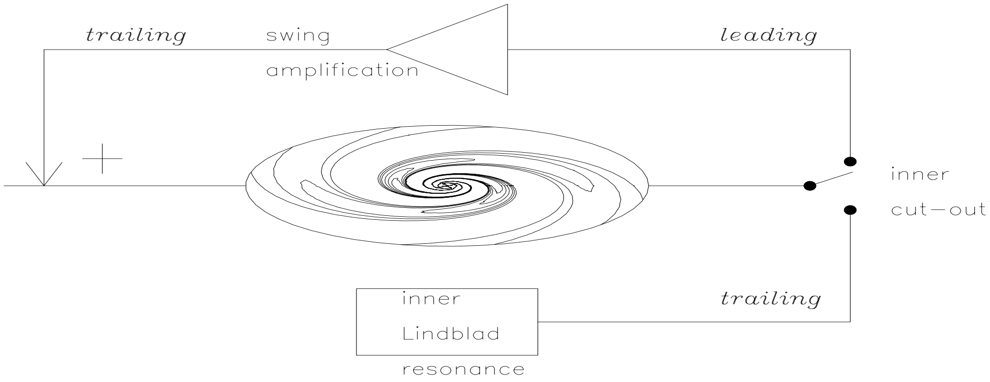


Figure 7.8: Sketch of the feedback loop and amplification. The position of the inner cut-out controls whether the waves are reflected as leading waves which can be swing-amplified, or absorbed as trailing waves by the inner Lindblad resonance.

We have thus assembled the components causing instabilities in our disks. A sharp inner cut-out protects incoming trailing waves from the damping influence of the inner Lindblad resonance, and reflects them out as leading waves. Differential rotation unwinds the leading waves into trailing waves. Swing amplification enhances the spiral arms as they pass from leading to trailing. These instabilities can be removed by either disabling the amplifier, or breaking the feedback circuit. Heating the disk is an example of the former: it weakens the self-gravity and reduces the swing amplification. Smoothing the inner cut-out, thus reducing the reflection and giving the waves access to the damping Lindblad resonance, is an example of the latter.

7.3 Global stability of the cut-out disks

We recall that the mathematical eigenvalue is defined by $A_{\mathrm{res}}(\alpha) = \lambda A_{\mathrm{imp}}(\alpha)$ (section 5.5). If the modulus of the complex mathematical eigenvalue $|\lambda|$ is greater than unity, the response is more vigorous than the imposed perturbation; if $|\lambda| < 1$, the response is less than the original disturbance. When the mathematical eigenvalue is unity, the response is equal in magnitude and phase to the imposed disturbance, and we have found a self-consistent mode. When investigating stability, therefore, we are principally concerned to understand the position of the eigenvalue in the complex plane, and to learn what choices of growth rate, pattern speed and temperature can bring it to the point $(1, 0)$.

In general, cool disks are less stable than warm ones. We thus expect the modulus of the mathematical eigenvalue to be greater for disks with low velocity dispersions. In Chapter 6, we saw that a disk could be made stable to axisymmetric disturbances by increasing the temperature sufficiently. We anticipate that a similar effect obtains for bisymmetric stability: by increasing the temperature sufficiently, we expect to be able to shrink the modulus of the mathematical eigenvalue so that no choice of growth rate and pattern speed will enable it to attain the value $(1, 0)$.

The mathematical eigenvalue depends on growth rate and pattern speed only through the single complex frequency $\omega = m\Omega_{\mathrm{p}} + is$. However, the growth rate s and pattern speed Ω_{p} have different effects on the disk's response. We expect to be able to excite slowly-growing disturbances more easily than faster ones. Thus we expect that the modulus of the mathematical eigenvalue will be greater for smaller growth rates. The pattern speed has a more complicated effect. First, it seems intuitively plausible that the pattern speed should affect the phase of the response. If the imposed pattern is rotating too fast for the disk, we expect the response to lag behind. This is expressed in the phase of the mathematical eigenvalue. Second, the pattern speed affects the position of the co-rotation radius R_{CR} and the inner and outer Lindblad resonances, R_{ILR} and R_{OLR}. As we have seen, any modes are expected to be confined to the region between the two Lindblad resonances. For $\tilde{\Omega}_{\mathrm{p}} \gtrsim 1 + \frac{1}{2}\sqrt{2-\beta}$, the outer Lindblad resonance lies inside the inner cut-out at $R = R_0$. The disturbance is thus confined to a part of the disk where there is not much matter free to respond. We expect the response, and hence the mathematical eigenvalue, to be small for large pattern speeds. For $\tilde{\Omega}_{\mathrm{p}} \lesssim 1 - \frac{1}{2}\sqrt{2-\beta}$, the inner Lindblad resonance lies beyond the inner cut-out. In this case we expect that the presence of the inner cut-out will have a minimal effect on the response, since the density experienced is similar to that in the self-consistent disk. As we saw in Chapter 4, each imposed log-spiral then excites only a pure log-spiral in response, and the eigenvalue is expected to be real. In a doubly cut-out disk, with sufficiently low pattern speed the inner Lindblad resonance would fall beyond the outer cut-off at $R = R_{\mathrm{c}}$. This would require $\tilde{\Omega}_{\mathrm{p}} \lesssim \left(1 - \frac{1}{2}\sqrt{2-\beta}\right) \tilde{R}_{\mathrm{c}}^{-\frac{2+\beta}{2}}$. In this circumstance we again expect the mathematical eigenvalue to be very small. In practice, we usually choose $\tilde{R}_{\mathrm{c}} = 100$, so this would require vanishingly small values of Ω_{p}. It turns out that the stability of the disk is little affected by the presence of an outer cut-out, so in what follows we concentrate on inner cut-out disks. The anticipated effect of pattern speed on the stability of an inner cut-out disk is summarised in the following table:

Pattern speed:	$\tilde{\Omega}_{\rm p} \lesssim 1 - \frac{1}{2}\sqrt{2-\beta}$	$1 - \frac{1}{2}\sqrt{2-\beta} \lesssim \tilde{\Omega}_{\rm p} \lesssim 1 + \frac{1}{2}\sqrt{2-\beta}$	$\tilde{\Omega}_{\rm p} \gtrsim 1 + \frac{1}{2}\sqrt{2-\beta}$
Position of $R_{\rm ILR}$ Position of $R_{\rm OLR}$	Outside inner cut-out Outside inner cut-out	Inside inner cut-out Outside inner cut-out	Inside inner cut-out Inside inner cut-out
Eigenvalue λ	As for self-consistent disk; roughly real	Affected by cut-out	Tends to zero

The effect of growth rate and pattern speed

Having discussed the behaviour of the mathematical eigenvalue from a theoretical point of view, we now examine the actual results obtained. In the following plots, the temperature of the disk has been chosen to be that at which the self-consistent disk would be locally stable to axisymmetric perturbations, i.e. $Q_{\rm s} = 1$ (see eq. (6.15)). As we saw in Chapter 6, this is close to the temperature required for the cut-out disks to be globally stable to axisymmetric perturbations. We take as an example the inner cut-out disk with $N = 2$. The following plots show how the largest eigenvalue depends on the growth rate and pattern speed, for different values of β. In these and subsequent plots, the diamonds indicate the data points actually calculated. The smooth curves joining the data points were obtained using a spline fitting procedure.

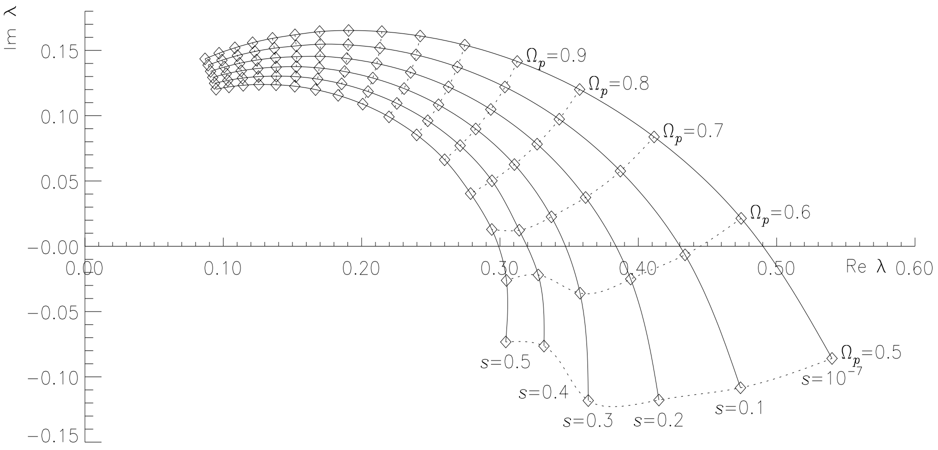


Figure 7.9: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = 0.25$. This is for $Q_{\rm s} = 1$, i.e. $\tilde{\sigma}_u = 0.283$ and $\gamma = 11.0$. The solid lines show curves of constant s at intervals of 0.1 from $s = 0^+$ to $s = 0.5$. The dotted lines show curves of constant $\Omega_{\rm p}$ at intervals of 0.1 from $\Omega_{\rm p} = 0.5$ to $\Omega_{\rm p} = 2.0$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$.]

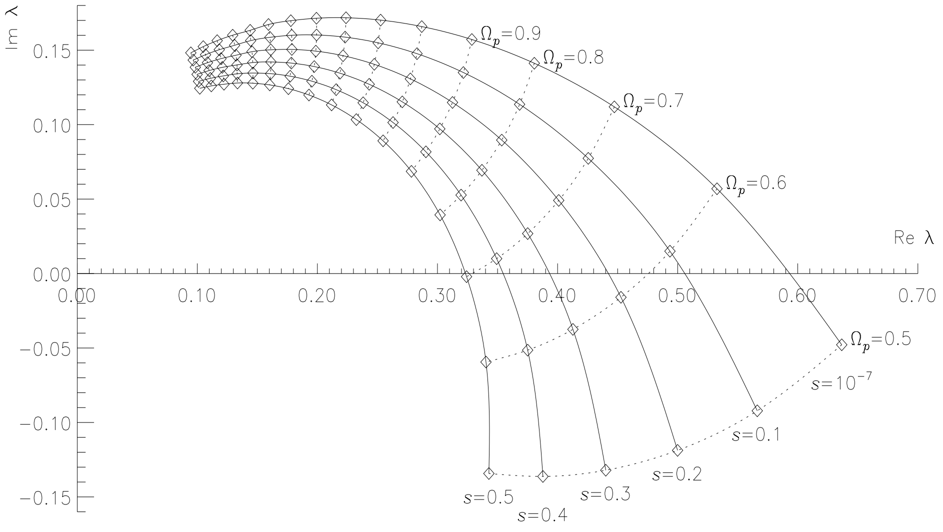


Figure 7.10: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = 0.0$. This is for $Q_{\rm s} = 1$, i.e. $\tilde{\sigma}_u = 0.378$ and $\gamma = 6.00$. These results were obtained using code written specially for the Toomre-Zang disk. The solid lines show curves of constant s at intervals of 0.1 from $s = 0^+$ to $s = 0.5$. The dotted lines show curves of constant $\Omega_{\rm p}$ at intervals of 0.1 from $\Omega_{\rm p} = 0.5$ to $\Omega_{\rm p} = 2.0$. Note that this figure corresponds to the figure on p. 113 of Zang's (1976) thesis, but there are visible discrepancies between the two. This is because Zang's figure was only intended as a guide to the shape of the curves, and was obtained using, as Zang says, "coarse results in which only a few radial harmonics are kept." [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = -30$, $l_{\max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$.]

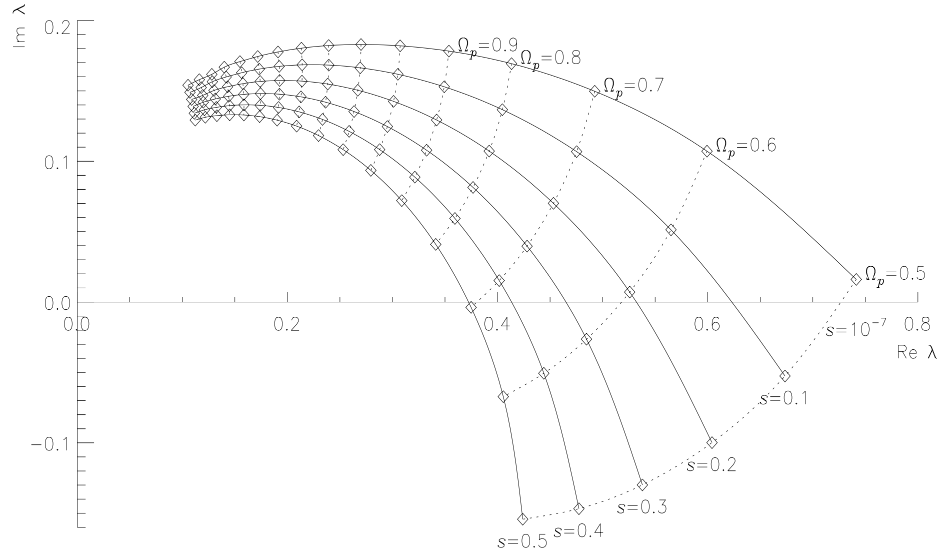


Figure 7.11: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = -0.25$. This is for $Q_{\rm s} = 1$, i.e. $\tilde{\sigma}_u = 0.509$ and $\gamma = 3.36$. The solid lines show curves of constant s at intervals of 0.1 from $s = 0^+$ to $s = 0.5$. The dotted lines show curves of constant $\Omega_{\rm p}$ at intervals of 0.1 from $\Omega_{\rm p} = 0.5$ to $\Omega_{\rm p} = 2.0$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = -30$, $l_{\max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$.]

These curves have the general behaviour expected from our theoretical discussion. Very roughly, the growth rate controls the modulus, and the pattern speed the phase, of the mathematical eigenvalue. The modulus decreases as the growth rate increases, and the phase increases as the pattern speed increases. At any particular pattern speed, the *marginal eigenvalue curve*, for which the growth rate is 0^+, has the greatest modulus. As the pattern speed is increased beyond ~ 1, the modulus of the mathematical eigenvalue declines. For large $\Omega_{\rm p}$, it looks as if the eigenvalue is turning towards to the origin, and the growth rate is becoming less significant. The disk responds to the perturbation only very weakly, since most of the mass has been removed from the region which would otherwise respond.

We note that the mapping of the s-Ω_p plane to the λ plane is conformal. As the kernel is analytic, the mathematical eigenvalue λ can be written as a power series in ω:

$$\begin{aligned}\lambda A(\alpha) &= \int_{-\infty}^{+\infty} d\alpha' A(\alpha') \mathcal{S}_m(\alpha, \alpha'; \omega) \\ &= \int_{-\infty}^{+\infty} d\alpha' A(\alpha') \left\{ \mathcal{S}_m(\alpha, \alpha'; \omega_0) + (\omega - \omega_0) \left.\frac{\partial \mathcal{S}_m}{\partial \omega}\right|_{\omega_0} + \frac{(\omega - \omega_0)^2}{2} \left.\frac{\partial^2 \mathcal{S}_m}{\partial \omega^2}\right|_{\omega_0} \cdots \right\}\end{aligned} \tag{7.11}$$

Equivalently, λ is an analytic function of ω (see e.g. Rudin (1976), ch. 8). This means that $\lambda(\omega)$ gives a conformal mapping of the upper half of the complex ω-plane (as the growth rate is positive) to the complex λ-plane. Conformal mappings preserve the angles of infinitesimal polygons. This means – as is evident in the figures – that curves of constant s and constant Ω_p cross at right-angles.

The following two figures demonstrate the conformal map from the ω-plane to the λ-plane and vice versa. The first figure, like those above, demonstrates how curves of constant pattern speed Ω_p and growth rate s map to the eigenvalue plane.

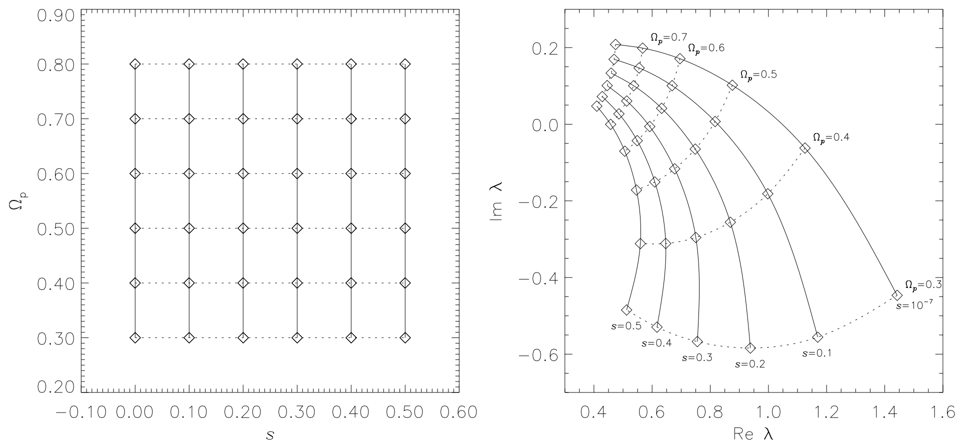


Figure 7.12: The mapping from frequency to eigenvalue. This is for a disk with $\beta = -0.50$, $N = 2$, $Q_\mathrm{s} = 1$. The solid lines show curves of constant s at intervals of 0.1 from $s = 0^+$ to $s = 0.5$. The dotted lines show curves of constant Ω_p at intervals of 0.1 from $\Omega_\mathrm{p} = 0.3$ to $\Omega_\mathrm{p} = 0.8$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -30$, $l_\mathrm{max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 1.5$.]

The second figure demonstrates how curves of constant λ map back onto the frequency plane.

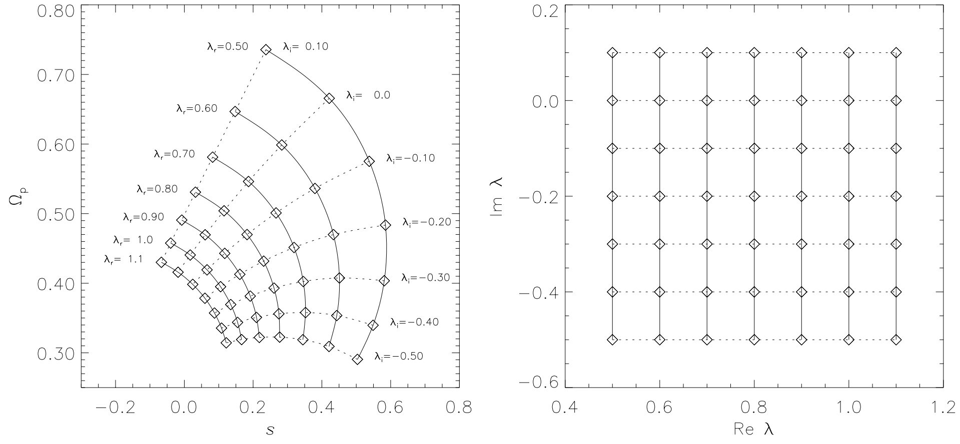


Figure 7.13: The mapping from eigenvalue to frequency. This is for a disk with $\beta = -0.50$, $N = 2$, $Q_\mathrm{s} = 1$. The solid lines show curves of constant $\mathrm{Re}(\lambda)$ at intervals of 0.1 from $\mathrm{Re}(\lambda) = 0.5$ to $\mathrm{Re}(\lambda) = 1.1$. The dotted lines show curves of constant $\mathrm{Im}(\lambda)$ at intervals of 0.1 from $\mathrm{Im}(\lambda) = -0.5$ to $\mathrm{Im}(\lambda) = 0.1$. [Numerical accuracy parameters: $n = 201$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -30$, $l_\mathrm{max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 15$, $b_\mathrm{acc} = 1.5$.]

Referring back to fig. 7.9, we see a problem which can occur. There is a discontinuity near $\Omega_\mathrm{p} = 0.5$, $s = 0.4$, although λ is expected to change smoothly. The discontinuity arises because here the second largest eigenvalue has overtaken the previously largest one, so that we are no longer following the same eigenvalue. The effect can be seen in fig. 7.14, which is a "close-up" on the region of the discontinuity.

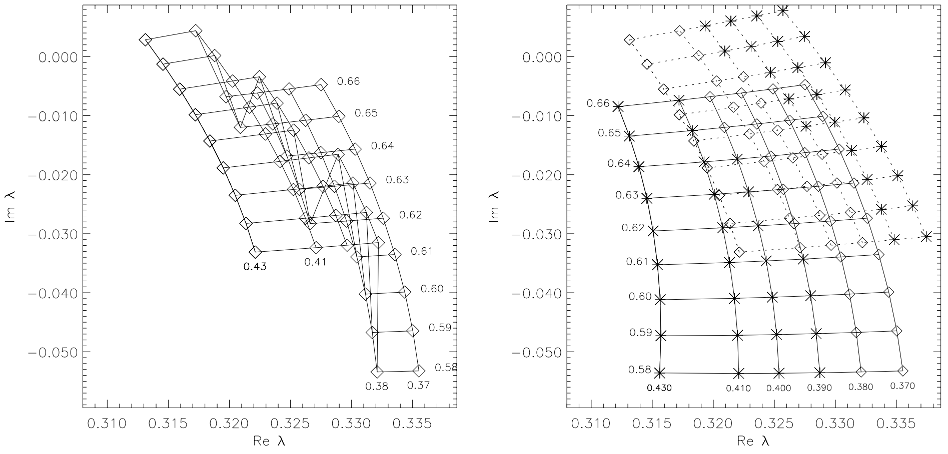


Figure 7.14: How eigenvalues can swap positions; see text for explanation. These results are for $m = 2$, $N = 2$ and $\beta = 0.25$. This is for $Q_\mathrm{s} = 1$, i.e. $\tilde{\sigma}_u = 0.283$ and $\gamma = 11.0$. [Numerical accuracy parameters: $n = 201$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -20$, $l_\mathrm{max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 1.5$.]

The left-hand plot shows the position of the largest eigenvalue, marked with a diamond, for six values of s from 0.37 to 0.43, and nine values of Ω_p from 0.58 to 0.66. Solid lines join together constant s and constant Ω_p. There is clearly a severe discontinuity, and the picture does not look like a conformal mapping from ω to λ. The right-hand plot shows how this comes about. Here, the position of the largest eigenvalue is marked with a diamond, and that of the second largest eigenvalue with an asterisk. The "nets" of solid and dotted lines follow the changes in a particular eigenvalue over the grid of Ω_p and s.

These nets give the conformal mapping from ω to λ; they are therefore smooth, and lines of constant s and constant $\Omega_{\rm p}$ intersect at right-angles. However, one sees that the two eigenvalues swap places over the region studied.

Now that we understand in general terms how the mathematical eigenvalue depends on growth rate, pattern speed and temperature, we wish to find sets of these variables for which the eigenvalue is unity. Figs. 7.9 - 7.11 strongly suggest that, for $Q_{\rm s} = 1$ disks with β in the range -0.25 to 0.25, no choice of growth rate and pattern speed results in an eigenvalue equal to unity. For each β, the curves of constant growth rate arc down and cross the real axis significantly to the left of unity. From theoretical considerations, we expect the mathematical eigenvalue to continue towards the origin as the pattern speed is increased beyond 2. With the constraints that, at any pattern speed, the marginal eigenvalue curve has the largest modulus, and that the curves of constant s cannot cross, it seems unlikely that any solutions can exist at pattern speeds lower than those shown in figs. 7.9 - 7.11. However, it is possible that, as $\Omega_{\rm p}$ is reduced further, the curves of constant growth rate bend back up again and touch the real axis again. Also, we have not yet examined the behaviour of other eigenvalues. Figs. 7.9 - 7.11 showed the behaviour of the dominant eigenvalue, i.e. that which has the largest modulus at a particular pattern speed and growth rate. For the range of pattern speeds investigated so far, the modulus of the dominant eigenvalue has not reached unity, and so we know that none of the lower eigenvalues can yield a mode. However, it is conceivable that, at some low value of the pattern speed, the dominant eigenvalue has a modulus in excess of unity and a phase presumably between $3\pi/2$ and 2π, while one of the lower eigenvalues has a modulus of unity and a phase of zero. It is therefore prudent to examine the full spectrum of eigenvalues.

Our finite matrix approximation to the kernel of the integral equation has n eigenvalues, where n is the number of points along each side of the grid in wavenumber space (see Chapter 5). To this extent, the eigenvalue spectrum is dependent on the grid size and spacing. However, once the number of grid points is large enough for accuracy, most of the eigenvalues accumulate at the origin. The positions of the relatively small proportion with larger modulus are independent of the precise choice of n. The set of 18 plots shown in fig. 7.15 demonstrates how the eigenvalue spectrum depends on growth rate and pattern speed. The three rows of fig. 7.15 show eigenvalue spectra for three different growth rates: $s = 10^{-7}$, 0.1, 0.2. The six columns show spectra for increasing pattern speeds: $\Omega_{\rm p} = 0.01$, 0.05, 0.1, 0.2, 0.5 and 1.0. In each plot, the position of the largest mathematical eigenvalue is marked with a line drawn to it from the origin.

At high pattern speed, the eigenvalues all have approximately the same phase. The phase of each eigenvalue is also roughly independent of growth rate for pattern speeds $\gtrsim 0.1$. As the pattern speed decreases, the phase of each eigenvalue decreases by an amount dependent on the magnitude of the eigenvalue. The largest eigenvalue moves clockwise more than the second largest, and so on. As the pattern speed decreases, the initially straight line formed by the eigenvalues shears out into a spiral. At smaller growth rates, the moduli of the eigenvalues are larger (as expected). At vanishing growth rate, the eigenvalues continue to be strung out in a spiral as the pattern speed is brought to zero. The tip of the spiral moves clockwise towards the real axis as $\Omega_{\rm p}$ is reduced. The behaviour of eigenvalues for higher

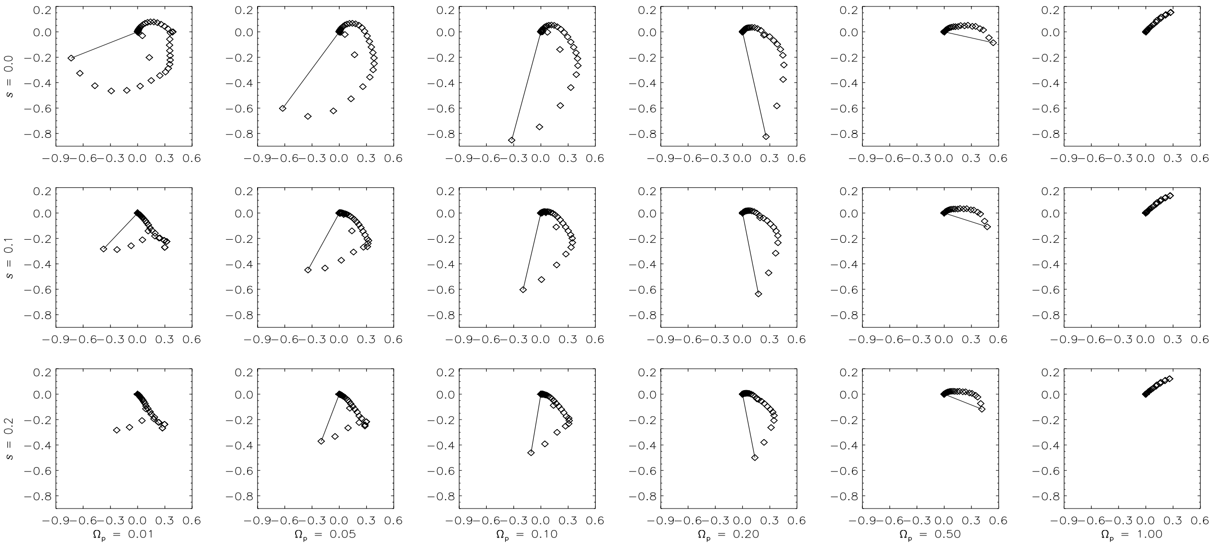


Figure 7.15: The eigenvalue spectra for perturbations with various growth rates and pattern speeds in an inner cut-out disk with $N = 2$, $\beta = 0.25$, $Q_{\mathrm{s}} = 1.0$. The three rows of plots show eigenvalue spectra for increasing growth rates: $s = 10^{-7}$, 0.1, 0.2. The six columns show spectra for increasing pattern speeds: $\Omega_{\mathrm{p}} = 0.01$, 0.05, 0.1, 0.2, 0.5 and 1.0. In each case, the position of the largest mathematical eigenvalue is marked with a line drawn to it from the origin. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\mathrm{min}} = -20$, $l_{\mathrm{max}} = +30$, $n_{GL} = 9$, $f_{\sigma} = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 2.0$.]

growth rates is different in this limit. The eigenvalues separate into two branches, forming a shape like a lambda. Most of the eigenvalues are clustered on a line of roughly constant phase. A few are strung out with more negative phases. They do not end up on the real axis as Ω_{p} is brought towards zero. Note that in the limit of vanishing pattern speed, large jumps can occur in the position of the largest mathematical eigenvalue, as it moves from one branch to another. This is one reason why it is best to study the full spectrum of eigenvalues in this limit.

We now trace the behaviour of the four largest marginal eigenvalues over a wide range of pattern speeds.

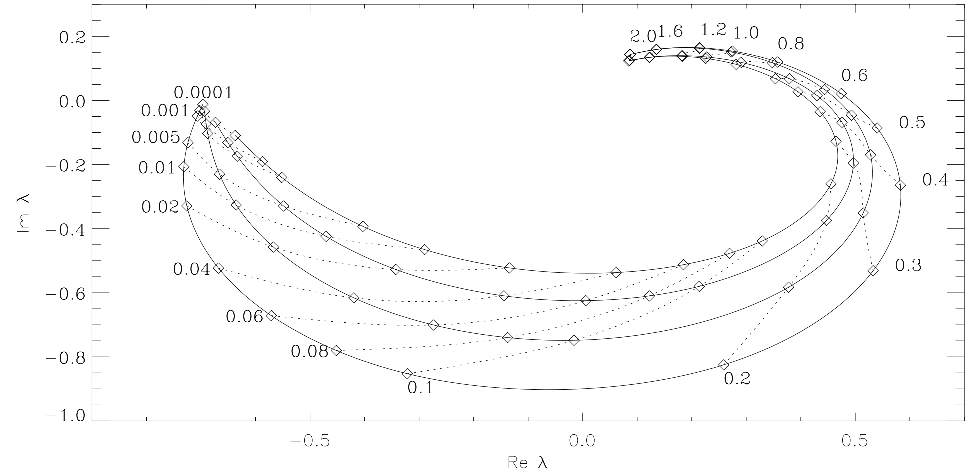


Figure 7.16: The dependence of the four largest eigenvalues on pattern speed, for a perturbation with vanishing growth rate in an inner cut-out disk with $N = 2$ and $\beta = 0.25$. This is for $Q_{\mathrm{s}} = 1$, i.e. $\tilde{\sigma}_u = 0.283$ and $\gamma = 11.0$. The solid lines are the first four marginal eigenvalue curves. The dotted lines show curves of constant Ω_{p}, as labelled. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\mathrm{min}} = -20$, $l_{\mathrm{max}} = +30$, $n_{GL} = 9$, $f_{\sigma} = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 1.5$.]

As the pattern speed is reduced, the phase continues to decrease monotonically, and the mathematical eigenvalues move on arcs around the origin, as shown in fig. 7.16. As the pattern speed tends to zero, they appear to end up on or close to the negative real axis. There is no question of their returning to cross the positive real axis for a second time. The behaviour of the four largest eigenvalues shown here is very similar. The phase of the larger eigenvalues consistently lags behind that of the smaller ones. (We saw in fig. 7.15 that smaller eigenvalues are successively less sensitive to pattern speed: their phase decreases more slowly as the pattern speed is reduced.) It appears, then, that we can rely on the largest eigenvalue to determine whether or not modes exist. We note also that for very large or small pattern speeds, the eigenvalues draw together. For large pattern speeds, they appear to come in degenerate pairs. For larger growth rates (not shown), the behaviour of the eigenvalues is qualitatively similar. Their tendency to draw together at low pattern speeds is less marked, and they do not approach the real axis as closely in this limit.

So it seems that, indeed, for disks in this range of β with $Q_{\rm s} = 1$, no choice of growth rate and pattern speed results in an eigenvalue equal to unity. An $N = 2$ disk which is hot enough to be stable to axisymmetric disturbances is also protected against bar-like modes. However, we note that as β is reduced in figs. 7.9 - 7.11, the set of curves tilts upwards, crossing the real axis at progressively higher values, and thus coming closer to possessing a unit eigenvalue. The disks with rising rotation curves (negative β) are less securely stable than their relatives with falling rotation curves (positive β). This accords with our finding in Chapter 6 that disks with rising rotation curve are less stable to axisymmetric perturbations. In fact, it turns out that disks with rising rotation curves are even less stable to bisymmetric perturbations than would be expected from the axisymmetric stability. Fig. 7.17 illustrates this point. It shows marginal eigenvalue curves for several different values of β in the range -0.5 to +0.35. The curves are drawn for $Q_{\rm s} = 1$, so all the disks shown are already stable to axisymmetric perturbations. Thus the disks with rising rotation curves have higher velocity dispersions than those with falling rotation curves, but even so they are more prone to bisymmetric disturbances. As β is decreased, the magnitude of the marginal eigenvalue at a given pattern speed increases. This means that eigenvalues in disks with rising rotation curves cross the real axis at higher values than for their cousins with falling rotation curves.

For values of $\beta < -0.470$, $Q_{\rm s} = 1$ is insufficient to stabilise the disk to $m = 2$ perturbations. Disks whose rotation curves rise faster than this are more susceptible to bar-like than to ring-like modes. As an example, fig. 7.18 shows the largest eigenvalue for $\beta = -0.50$. It is evident that the disk will admit an $m = 2$ mode with growth rate $s < 0.1$ and pattern speed $\Omega_{\rm p} \sim 0.45$. A Newton-Raphson algorithm indicates that this mode occurs at $\Omega_{\rm p} = 0.440732$, $s = 0.0175300$. The curves corresponding to this pattern speed and growth rate are shown in dashed lines on fig. 7.18. The density transform for the mode is shown in fig. 7.19. (Of course, the overall magnitude and phase of the function shown are irrelevant, since any complex multiple of an eigenvector is also an eigenvector.) We note that the density transform is peaked at $\alpha \approx 3$, indicating that the mode is close to a log-spiral with this wavenumber.

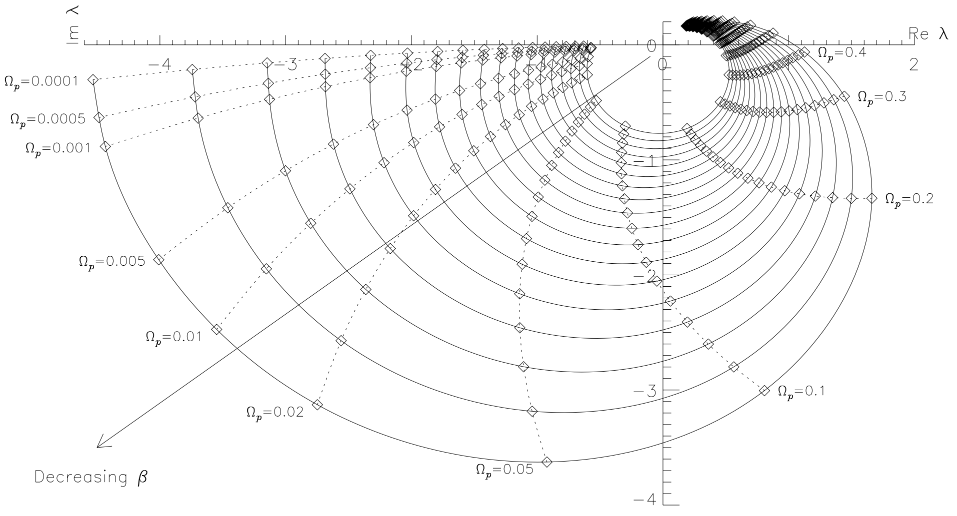


Figure 7.17: Marginal eigenvalue curves for inner cut-out disks with $N = 2$ and 18 values of β from $\beta = -0.5$ to $\beta = 0.35$ in steps of 0.05. The plot shows the largest mathematical eigenvalues for vanishing growth rate, $s = 10^{-7}$, and 20 values of the pattern speed: $\Omega_{\rm p} = 0.0001$, 0.0005, 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, ... 1.2, 1.3. The velocity dispersion for each β is that corresponding to $Q_{\rm s} = 1$. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

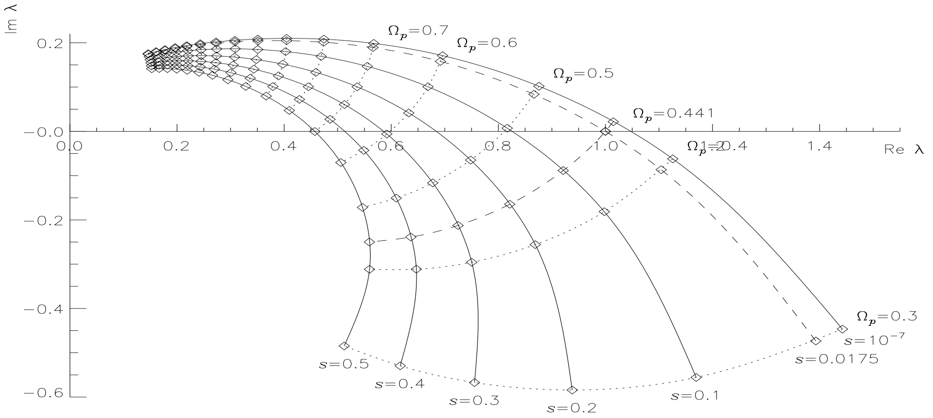


Figure 7.18: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = -0.50$. This is for $Q_{\rm s} = 1$, i.e. $\tilde{\sigma}_u = 0.740$ and $\gamma = 1.83$. The solid lines show curves of constant s at intervals of 0.1 from $s = 0^+$ to $s = 0.5$. The dotted lines show curves of constant $\Omega_{\rm p}$ at intervals of 0.1 from $\Omega_{\rm p} = 0.5$ to $\Omega_{\rm p} = 2.0$. The dashed lines mark the curves which pass through the unit eigenvalue. [Numerical accuracy parameters: $n = 201$, $\Delta\alpha = 0.2$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

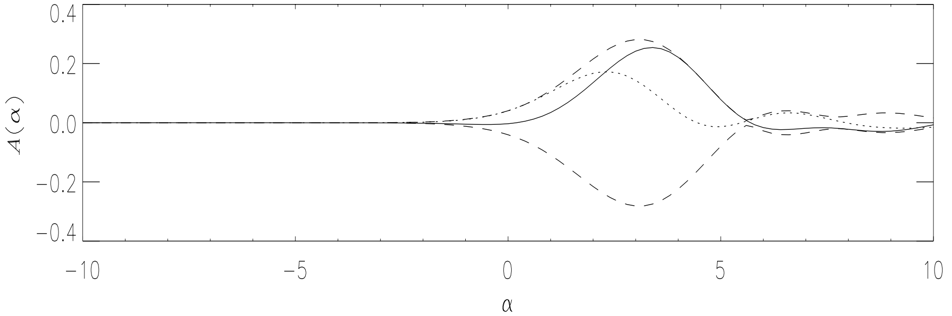


Figure 7.19: Density transform $A(\alpha)$ of a bisymmetric mode in an inner cut-out disk with $N = 2$, $\beta = -0.5$. The mode has pattern speed $\Omega_{\rm p} = 0.440732$ and growth rate $s = 0.0175300$. This is for $Q_{\rm s} = 1$, i.e. $\tilde{\sigma}_u = 0.740$ and $\gamma = 1.83$. The solid line is $\mathrm{Re}(A(\alpha))$, the dotted is $\mathrm{Im}(A(\alpha))$ and the dashed line shows the envelope $\pm|A(\alpha)|$. [Numerical accuracy parameters: $n = 101$, $\Delta\alpha = 0.2$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

The actual form of the density is shown in the left-hand contour plot in fig. 7.20. The right-hand plot shows the pure log-spiral with wavenumber $\alpha = 3$. The log-spiral describes the winding of the spiral arms well, but not the shape of the density within them. The mode has a maximum density perturbation just outside the inner Lindblad resonance, but extends out to the co-rotation radius. The log-spiral, in contrast, is concentrated at the centre of the disk.

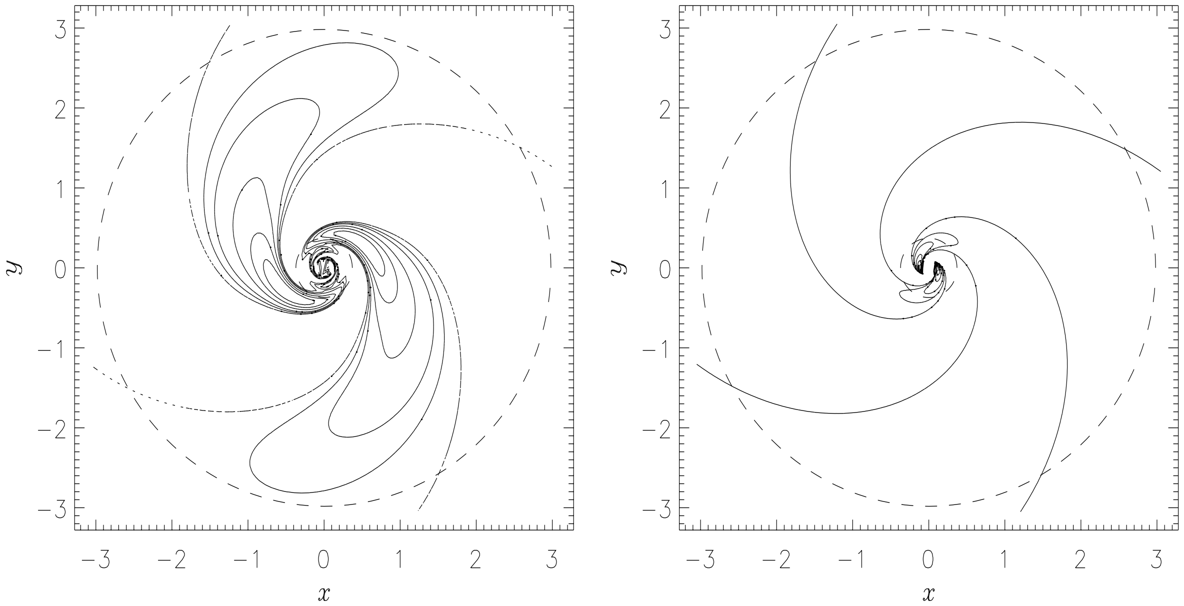


Figure 7.20: The left-hand plot is the density contour plot for a bisymmetric mode in an inner cut-out disk with $N = 2$, $\beta = -0.5$. This is for $Q_{\rm s} = 1$, i.e. $\tilde{\sigma}_u = 0.740$ and $\gamma = 1.83$. The density is plotted over the radial range from $0.3R_{\rm ILR}$ to $1.1R_{\rm CR}$. The solid lines plot 10%, 20%, 40%, 60% and 80% of the maximum density within this range; the dotted lines show the nodes, where the density perturbation is zero. The dashed circles indicate the inner Lindblad resonance, at 0.371, and the co-rotation radius, at 2.98. Units of $R_0 = 1$ are used. The right-hand plot shows the log-spiral with $m = 2$, $\alpha = 3$, over the same range in R. Again, contours are plotted at 10%, 20%, 40%, 60% and 80% of the maximum density within this range. [Numerical accuracy parameters: $n = 101$, $\Delta\alpha = 0.2$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

The effect of temperature

Raising the temperature of a stellar disk makes it more stable to any perturbations. Thus raising the velocity dispersion $\tilde{\sigma}_u$ moves the marginal stability curves closer to the origin, while lowering it moves them further out. At some value $\tilde{\sigma}_{u,\rm min}$, the marginal stability curve will pass through (1,0); this is the temperature at which the disk is just stable to bisymmetric perturbations. Figs. 7.21 - 7.23 show how the temperature affects the marginal stability curves for a disk with a falling, flat or rising rotation curve. The eigenvalue depends separately on the pattern speed and temperature, so the mapping from the $\Omega_{\rm p}$-$Q_{\rm s}$ plane to the λ plane is not conformal. This is evident from the figures, where lines of constant $\Omega_{\rm p}$ and constant $Q_{\rm s}$ clearly do not intersect at right-angles. Again, we see that disks with falling rotation curves are very stable to $m = 2$ perturbations. Even when the velocity dispersion is only 60% of that needed for axisymmetric stability, these disks remain stable to bisymmetric perturbations.

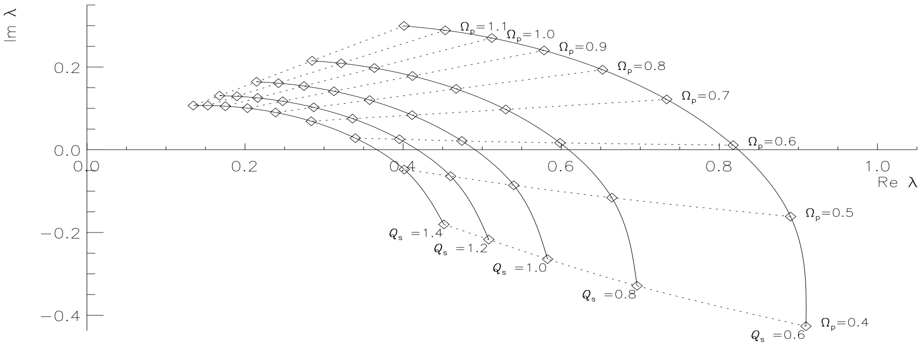


Figure 7.21: The dependence of the largest mathematical eigenvalue on temperature and pattern speed for $\beta = 0.25$. Curves are plotted for an $m = 2$, $s = 10^{-7}$ perturbation in an inner cut-out disk with $N = 2$. The temperature used for each curve is expressed in terms of Q_s: $\tilde{\sigma}_u = 0.283 Q_\mathrm{s}$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -30$, $l_\mathrm{max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 2$.]

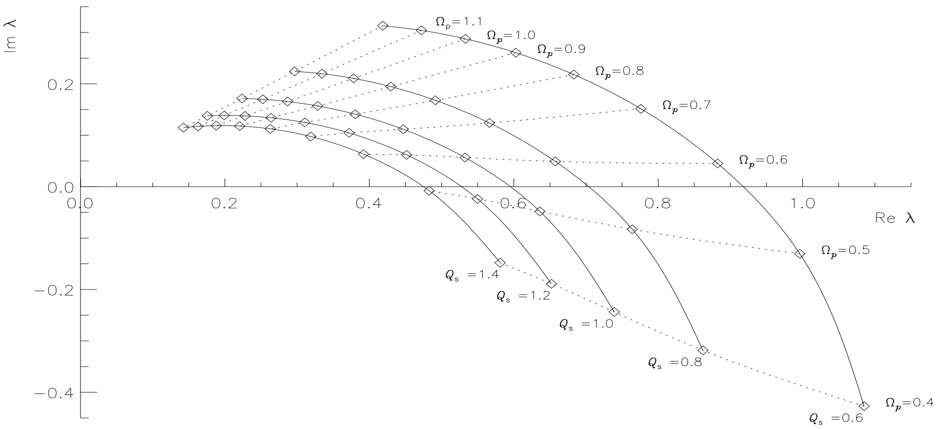


Figure 7.22: The dependence of the largest mathematical eigenvalue on temperature and pattern speed for $\beta = 0.00$. Curves are plotted for an $m = 2$, $s = 10^{-7}$ perturbation in an inner cut-out disk with $N = 2$. The temperature used for each curve is expressed in terms of Q_s: $\tilde{\sigma}_u = 0.378 Q_\mathrm{s}$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -30$, $l_\mathrm{max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 2$.]

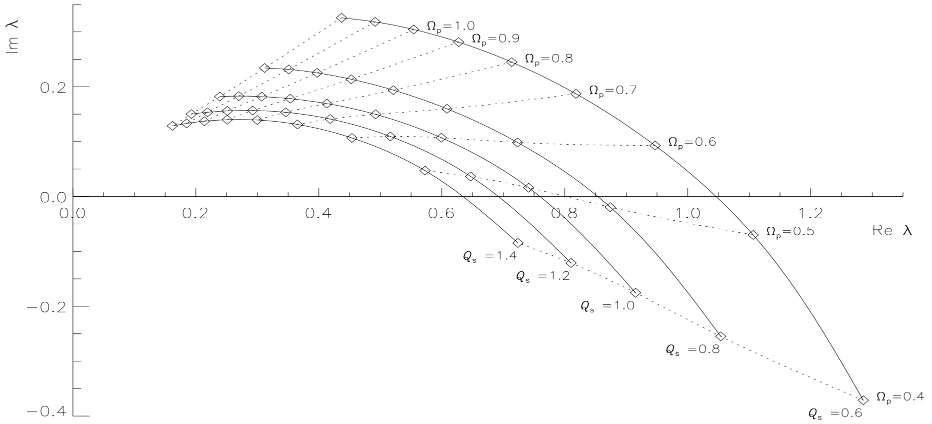


Figure 7.23: The dependence of the largest mathematical eigenvalue on temperature and pattern speed for $\beta = -0.25$. Curves are plotted for an $s = 10^{-7}$ perturbation in an inner cut-out disk with $N = 2$. The temperature used for each curve is expressed in terms of Q_s: $\tilde{\sigma}_u = 0.509 Q_\mathrm{s}$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -30$, $l_\mathrm{max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 2$.]

The effect of the inner cut-out

As expected from the discussion of tightly-wound waves in section 7.2, the inner cut-out has a profound influence on the stability of the disk. To illustrate this, fig. 7.24 shows the marginal stability curves in disks with different β and different values of the inner cut-out index N.

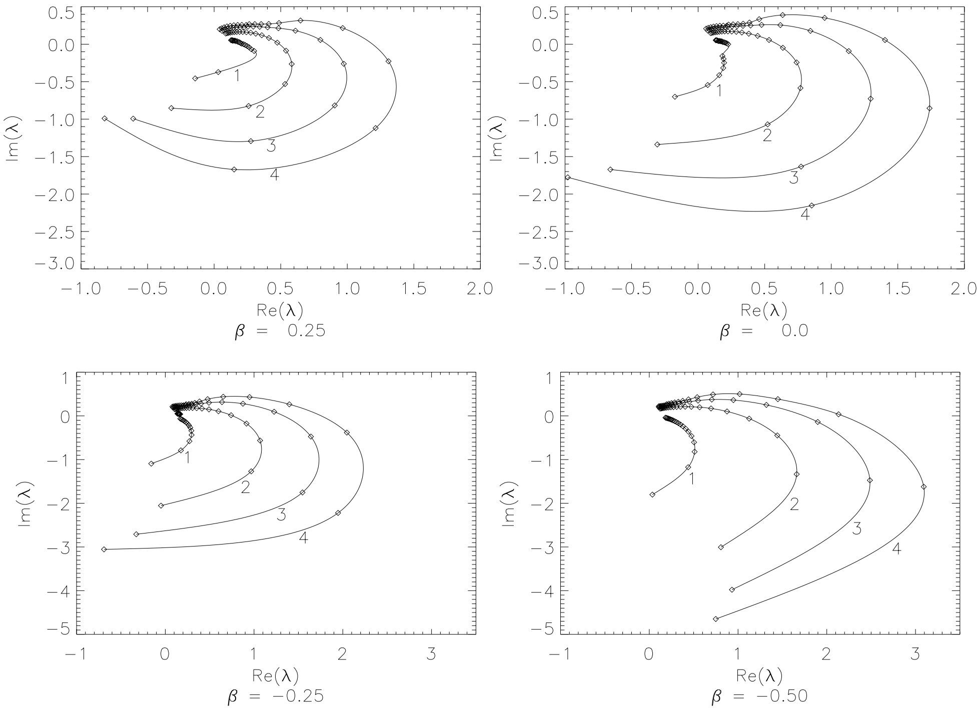
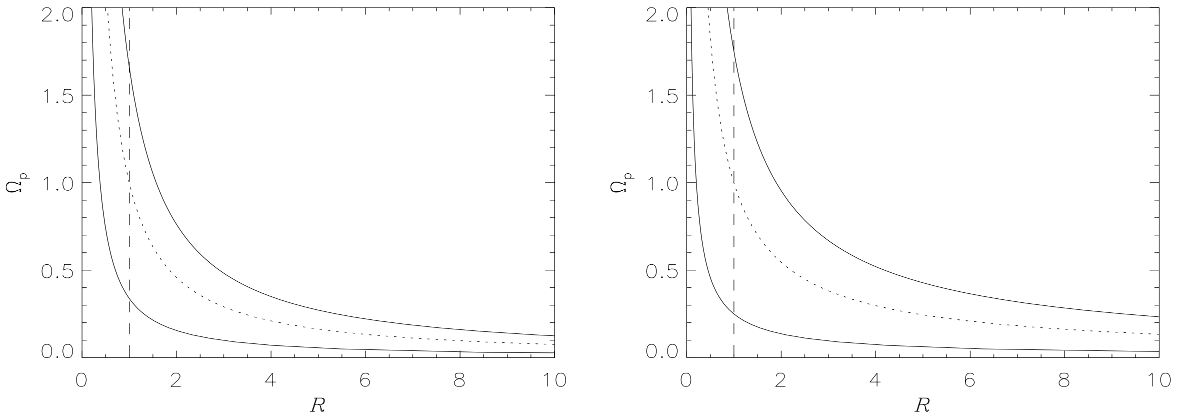


Figure 7.24: The dependence of the largest mathematical eigenvalue on pattern speed, for different inner cut-out indices N. Curves are plotted for a perturbation with vanishing growth rate in an inner cut-out disk. The first two plots are for $\beta = 0.25$ and $\beta = 0.00$; the second two are for $\beta = -0.25$ and $\beta = -0.50$. The curves are labelled with the value of the inner cut-out index N. In each case $Q_{\mathrm{s}} = 1$. The diamonds show the results for $\Omega_{\mathrm{p}} = 0.1,\ 0.2,\ 0.3\ \ldots\ 2.0$ [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = -20$, $l_{\max} = +30$, $n_{GL} = 9$, $f_{\sigma} = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 1.5$.]

Increasing the cut-out index N makes the disk much more susceptible to instabilities. Curves with higher N cross the real axis at successively higher values of λ. For these $Q_{\mathrm{s}} = 1$ results, the $N = 3$ and $N = 4$ disks are already unstable to $m = 2$. Disk with $N = 1$ are highly stable to $m = 2$ perturbations, and will not be discussed further in this chapter.

To understand why the inner cut-out has such a profound effect on the stability of the disks, we need to consider the position of the Lindblad resonances in relation to the inner cut-out. These are shown, with their dependence on pattern speed, in fig. 7.25. The solid lines show the positions of the Lindblad resonances, and the dotted line marks co-rotation. The dashed line marks $R = R_0$, the position of the inner cut-out.

For low values of the pattern speed, the inner Lindblad radius lies outside the inner cut-out radius R_0. Incoming waves are absorbed by the inner Lindblad resonance, which thus damps down disturbances in the disk, and reduces the possibility of a mode. However, once Ω_{p} exceeds ~ 0.25, the inner Lindblad resonance moves inside the inner cut-out radius. If the cut-out is gentle ($N = 1$), the incoming waves pass



β = 0.25

β = −0.25

Figure 7.25: The position of co-rotation and of the $m = 2$ Lindblad resonances for $\beta = \pm 0.25$. Units $v_\beta = R_0 = 1$ are used.

through it, reach the inner Lindblad resonance and are absorbed. If however the cut-out is sufficiently sharp (as for larger cut-out indices), it presents a barrier which reflects the incoming trailing waves. Instead of reaching the inner Lindblad resonance and being absorbed, they are reflected back as outgoing leading waves. The exceptional stability of $N = 1$ disks is mostly due to the fact that the surface density is disturbed much less abruptly than for higher cut-out indices, presenting less of a reflective barrier to incoming waves. An additional, related factor is the much greater surface density within the inner cut-out when $N = 1$. For $N > 1$, the surface density rapidly falls to zero within R_0. Waves are thus suppressed by simply having no medium through which to propagate! For $N = 1$, the active surface density diverges at the origin when the rotation curve is falling, and even when the rotation curve is rising the surface density remains large to much smaller radii (see fig. 2.12).

Marginal modes

We are frequently interested in marginal stability: the stability of the disk to disturbances with zero growth rate. To investigate this, we set $s = 0^+$, and keep only the temperature and pattern speed as variables. For sufficiently high temperature, the disk is stable; even the largest eigenvalue curve crosses the real axis at less than unity, and no self-consistent marginal modes are possible. How hot does the disk have to be for this stabilisation to occur? We can answer this by finding the temperature for which the marginal eigenvalue curve passes through $(1, 0)$. The numerical procedure for finding this unit eigenvalue is described in Chapter 5, section 5.7. The temperature γ and pattern speed Ω_p were adjusted until the difference between the largest mathematical eigenvalue and unity was less than some small quantity[2] ϵ_λ , which was usually chosen to be 10^{-6}. This means that the accuracy of γ_max and $\tilde{\sigma}_{u,\mathrm{min}}$ was not specified in advance. An estimate of this accuracy is provided by the size of the last step found necessary in the Newton-Raphson routine as the search homed in on the unit eigenvalue. This was typically less than 10^{-5} for $\tilde{\sigma}_u$ and Ω_p, and less than 10^{-2} for γ, indicating that γ_max and $\tilde{\sigma}_{u,\mathrm{min}}$ are accurate to at

---

[2] Specifically, the iteration was continued until $|\mathrm{Re}(\lambda) - 1|$ and $|\mathrm{Im}(\lambda)|$ were both less than ϵ_λ.

least 3 s.f. The numerical accuracy parameters used are given underneath the graphs below. These were chosen so as to ensure that the largest mathematical eigenvalue converged to at least 6 s.f. Tables of the data used in figures 7.26 and 7.27 are given in Appendix E. The $\beta = 0$ results agree with those given in Zang (1976).

Fig. 7.26 shows the minimum temperature needed for bisymmetric stability, as a function of β. The solid lines show the results for the inner cut-out disks with $N = 2, 3, 4$. The broken lines show results for the corresponding doubly cut-out disks with $M = 2, 4, 6$, $\tilde{R}_c = 10$. For comparison, the temperature necessary for global axisymmetric stability of the self-consistent disk is also shown.

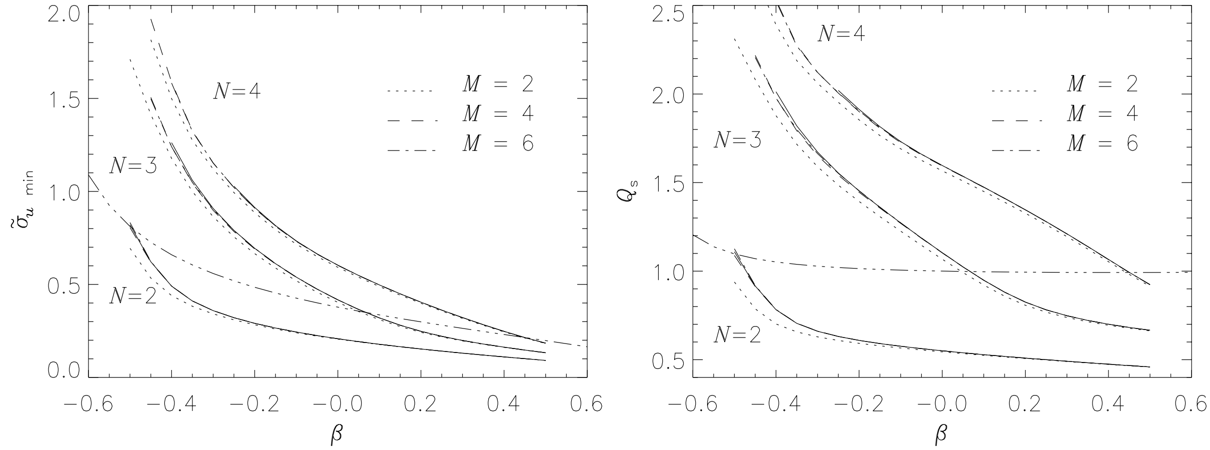


Figure 7.26: Minimum temperature for bisymmetric stability plotted against β, for disks with various cut-out functions. The left-hand plot shows the minimum velocity dispersion $\tilde{\sigma}_u$; the right-hand one the same data presented in terms of the stability parameter $Q_{\rm s}$. The solid lines indicate the results for inner cut-out disks with $N = 1, 2, 3, 4$. The broken lines indicate the corresponding doubly cut-out disks with $M = 2, 4, 6$ and $\tilde{R}_c = 10$. The remaining line shows the temperature necessary for global axisymmetric stability in the self-consistent disk. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = -30$, $l_{\max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2.0$, $\epsilon_\lambda = 10^{-6}$.]

Curves with different N lie well apart; the inner cut-out function has a profound effect on the stability of the disk. The more sharply the core is cut out, the more unstable the disk becomes. For each N, the curves with different M lie very close together. Even with this very low truncation radius, the outer cut-out function has made very little difference to the temperature necessary for stability. With a larger $\tilde{R}_c$, say 100, the doubly cut-out curves are scarcely distinguishable from those with only an inner cut-out (data not shown).

We can use fig. 7.26 to compare the stability of the various cut-out disks to $m = 0$ and $m = 2$ perturbations. The $N = 2$ disk is much less prone to bisymmetric than to ring-like perturbations over almost all β. For $N = 3$, the relative tendency to $m = 2$ and $m = 0$ disturbances depends on β. Roughly speaking, $N = 3$ disks with rising rotation curves are more prone to $m = 2$ disturbances, and disks with falling rotation curves are more prone to $m = 0$. Conversely, disks with $N = 4$ disturbances are much more prone to bisymmetric disturbances over almost all β; only when the rotation curve is steeply falling does the $m = 0$ mode become harder to stabilise. It is perhaps surprising that it is the lowest outer cut-out index, $M = 2$, which shows the largest departure from the inner cut-out curves. It seems that tapering the disk gently at large radius has more of an effect than truncating it abruptly. This is presumably because for lower M, the effect of the outer cut-out is felt further in towards the centre of the disk. For higher M, the outer cut-out is sharp, but occurs entirely outside the outer Lindblad resonance. As for the axisymmetric case (section 6.6), disks with a gentle outer cut-out are more stable than those where the cut-out is sharper.

The critical pattern speed obtained for these marginal modes is shown in fig. 7.27. We see that for all cut-out functions and β, the pattern speed is confined to a fairly narrow range, roughly 0.3 to 0.6. Over a wide range in β, the pattern speed varies roughly linearly with β.

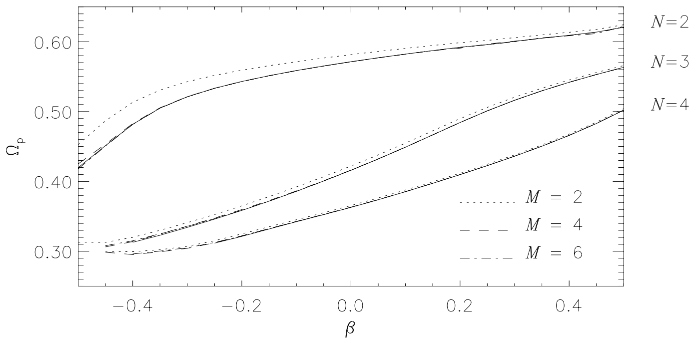
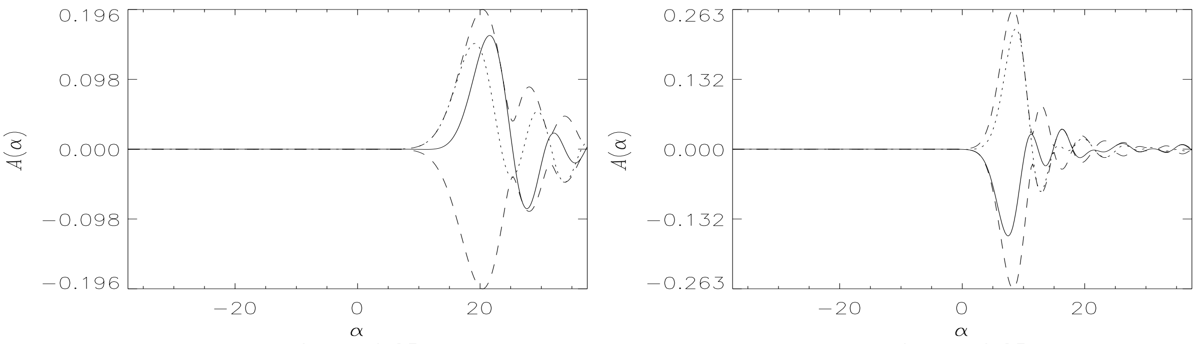


Figure 7.27: Critical pattern speed $\Omega_{\rm p}$ plotted against β, for an $m = 2$ perturbation in disks with various cut-out functions. The solid lines are labelled with the appropriate inner cut-out index N. For each N, different outer cut-out indices M are plotted. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$, $\epsilon_\lambda = 10^{-6}$.]

Let us look in more detail at the marginal modes. Fig. 7.28 shows the density transforms $A(\alpha)$ for modes in inner cut-out disks with $N = 2$, $\beta = \pm 0.25$.



β = 0.25

β = −0.25

Figure 7.28: Density transform $A(\alpha)$ of the marginal modes in inner cut-out disks with $N = 2$, $\beta = \pm 0.25$. The solid line is $\mathrm{Re}(A(\alpha))$, the dotted is $\mathrm{Im}(A(\alpha))$ and the dashed line shows the envelope $\pm|A(\alpha)|$. For $\beta = 0.25$, the marginal mode has $Q_{\rm s} = 0.500$, $\tilde{\sigma}_u = 0.142$, $\gamma = 48.4$, $\Omega_{\rm p} = 0.596$. For $\beta = -0.25$, the marginal mode has $Q_{\rm s} = 0.631$, $\tilde{\sigma}_u = 0.321$, $\gamma = 9.19$, $\Omega_{\rm p} = 0.533$. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.25$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

The contour plots in figs. 7.29 - 7.31 show the form of the marginal modes for $\beta = \pm 0.25$ in inner cut-out disks with $N = 2$, $N = 3$ and $N = 4$.

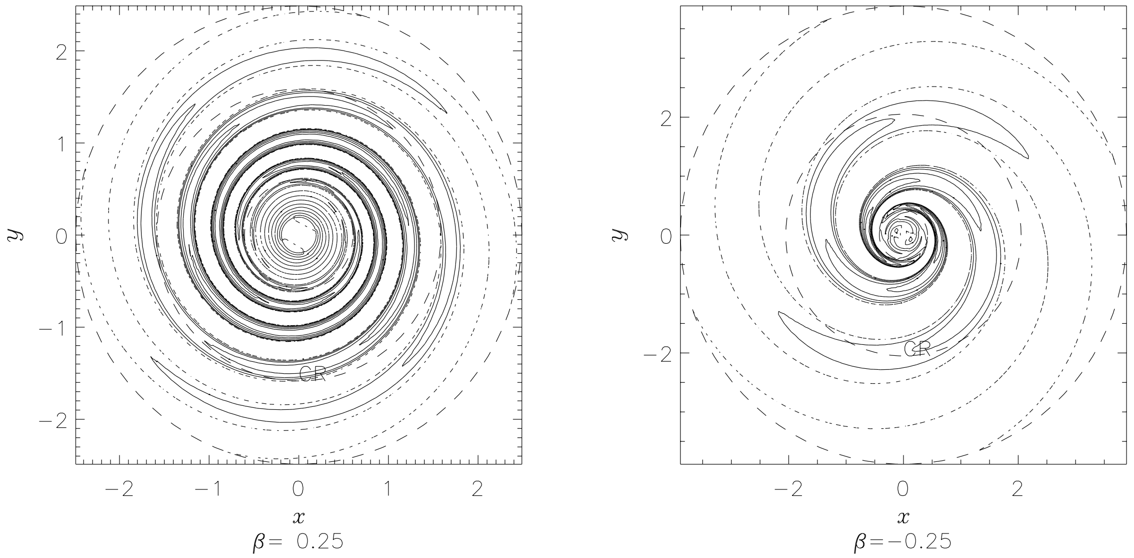


Figure 7.29: Density contour plot of a marginally stable mode in inner cut-out disks with $N = 2$. The left-hand plot is for $\beta = 0.25$, for which the marginally stable mode has $Q_{\rm s} = 0.500$, $\gamma = 48.4$, $\tilde{\sigma}_u = 0.142$, $\Omega_{\rm p} = 0.596$, $\alpha_{\rm u} \approx 20.4$. The right-hand plot is for $\beta = -0.25$, for which the marginally stable mode has $Q_{\rm s} = 0.631$, $\gamma = 9.19$, $\tilde{\sigma}_u = 0.321$, $\Omega_{\rm p} = 0.533$, $\alpha_{\rm u} \approx 8.4$. In each plot, the density is calculated between $R = 0.3R_{\rm ILR}$ and $R = R_{\rm OLR}$. The solid lines mark 10%, 20%, 40%, 60% and 80% of the maximum density within this range; the dotted lines show the nodes, where the density perturbation is zero. The dashed circles indicate the Lindblad resonances and the co-rotation radius. For $\beta = 0.25$, $R_{\rm ILR} = 0.605$, $R_{\rm CR} = 1.58$, $R_{\rm OLR} = 2.49$; for $\beta = -0.25$, $R_{\rm ILR} = 0.421$, $R_{\rm CR} = 2.05$, $R_{\rm OLR} = 3.89$, in units where $R_0 = 1$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

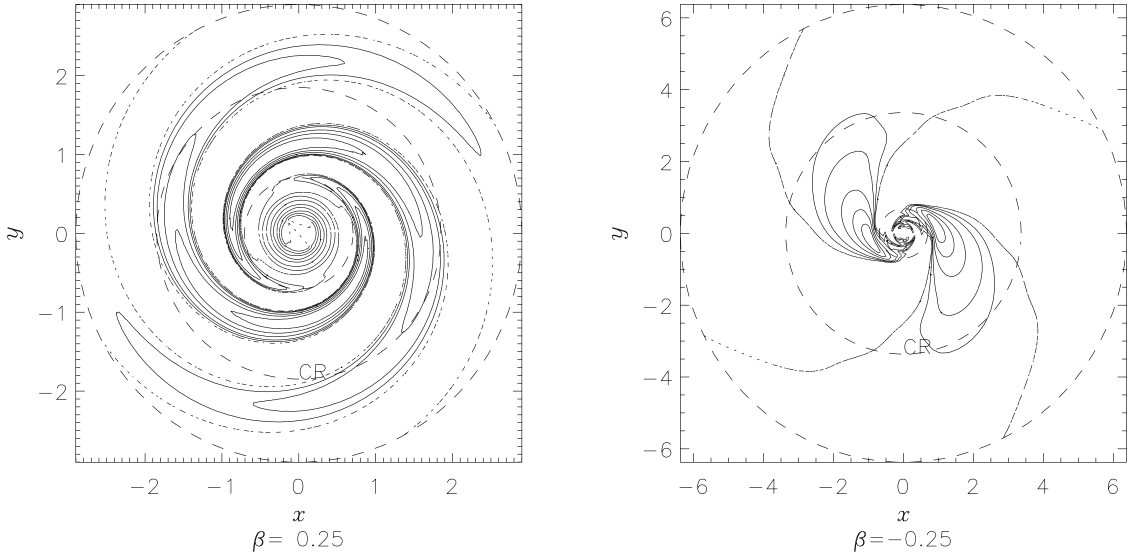


Figure 7.30: Density contour plot of a marginally stable mode in inner cut-out disks with $N = 3$. The left-hand plot is for $\beta = 0.25$, for which the marginally stable mode has $Q_{\rm s} = 0.782$, $\gamma = 18.9$, $\tilde{\sigma}_u = 0.221$, $\Omega_{\rm p} = 0.501$, $\alpha_{\rm u} \approx 10.0$. The right-hand plot is for $\beta = -0.25$, for which the marginally stable mode has $Q_{\rm s} = 1.55$, $\gamma = 1.10$, $\tilde{\sigma}_u = 0.791$, $\Omega_{\rm p} = 0.346$ $\alpha_{\rm u} \approx 2.2$. In each plot, the density is calculated between $R = 0.3R_{\rm ILR}$ and $R = R_{\rm OLR}$. The solid lines mark 10%, 20%, 40%, 60% and 80% of the maximum density within this range; the dotted lines show the nodes, where the density perturbation is zero. The dashed circles indicate the Lindblad resonances and the co-rotation radius. For $\beta = 0.25$, $R_{\rm ILR} = 0.706$, $R_{\rm CR} = 1.85$, $R_{\rm OLR} = 2.90$; for $\beta = -0.25$, $R_{\rm ILR} = 0.690$, $R_{\rm CR} = 3.36$, $R_{\rm OLR} = 6.38$, in units where $R_0 = 1$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

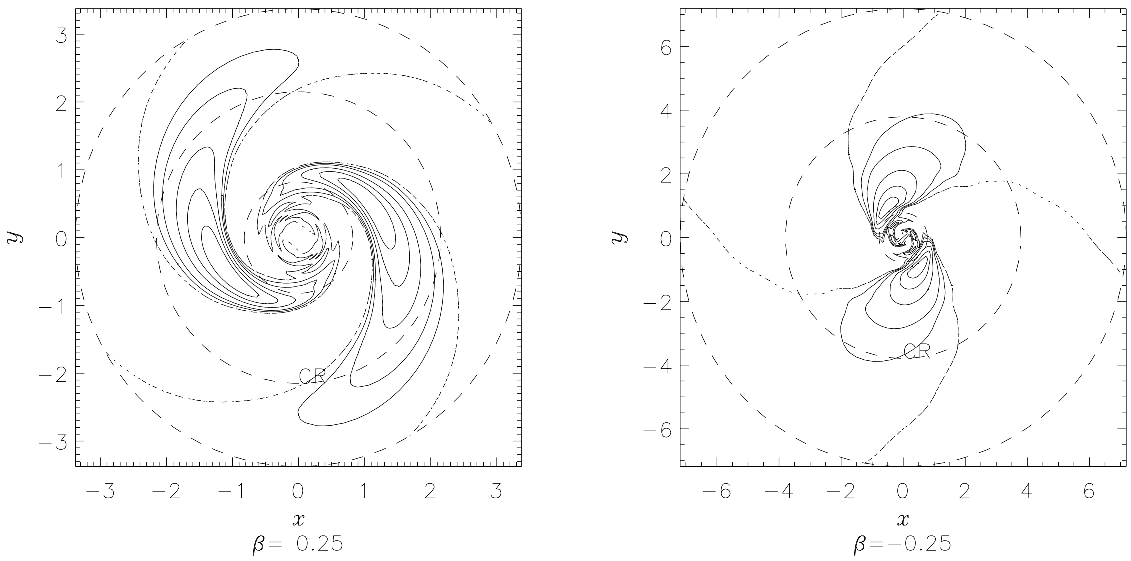


Figure 7.31: Density contour plot of a marginally stable mode in inner cut-out disks with $N = 4$. The left-hand plot is for $\beta = 0.25$, for which the marginally stable mode has $Q_\mathrm{s} = 1.28$, $\gamma = 6.13$, $\tilde{\sigma}_u = 0.362$, $\Omega_\mathrm{p} = 0.423$ $\alpha_\mathrm{u} \approx 4.4$. The right-hand plot is for $\beta = -0.25$, for which the marginally stable mode has $Q_\mathrm{s} = 2.02$, $\gamma = 0.446$, $\tilde{\sigma}_u = 1.03$, $\Omega_\mathrm{p} = 0.312$ $\alpha_\mathrm{u} \approx 1.6$. In each plot, the density is calculated between $R = 0.3R_\mathrm{ILR}$ and $R = R_\mathrm{OLR}$. The solid lines mark 10%, 20%, 40%, 60% and 80% of the maximum density within this range; the dotted lines show the nodes, where the density perturbation is zero. The dashed circles indicate the Lindblad resonances and the co-rotation radius. For $\beta = 0.25$, $R_\mathrm{ILR} = 0.820$, $R_\mathrm{CR} = 2.15$, $R_\mathrm{OLR} = 3.38$; for $\beta = -0.25$, $R_\mathrm{ILR} = 0.776$, $R_\mathrm{CR} = 3.79$, $R_\mathrm{OLR} = 7.19$, in units where $R_0 = 1$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -20$, $l_\mathrm{max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 2$.]

For falling rotation curves, the modes are much more tightly wound than for rising rotation curves. They extend beyond co-rotation, up to the outer Lindblad resonance, whereas in disks with rising rotation curves the modes are concentrated between the inner Lindblad resonance and the co-rotation radius. The modes are also more tightly wound for disks where the cut-out function is gentler.

These correlations are related to the temperature of the disks. Disks with rising rotation curves are less stable than disks with falling rotation curves; similarly, disks where the cut-out is sharp are less stable than disks with gentle inner cut-outs. In both cases, higher temperatures are needed for marginal modes. The observation that disks with negative β have looser, less extensive marginal modes reflects the higher temperature rather than an innate tendency for spirals to be more loosely wound in disks with rising rotation curves. For comparison, we now investigate growing modes in disks with the same velocity dispersion. We consider $N = 2$ disks with $\beta = \pm 0.25$, with in each case $\tilde{\sigma}_u = 0.1$. This corresponds to $Q_\mathrm{s} = 0.353$ in the disk with $\beta = 0.25$, and $Q_\mathrm{s} = 0.197$ in the disk with $\beta = -0.25$. The disk with $\beta = -0.25$ is less stable, so at this velocity dispersion it admits much faster-growing modes than the disk with $\beta = 0.25$. The fastest-growing modes in each case are shown in fig. 7.32. They are equally tightly wound (in each case the density transform $A(\alpha)$ peaks close to $\alpha = 24$), confirming that it is velocity dispersion, rather than rotation curve, which principally affects the winding.

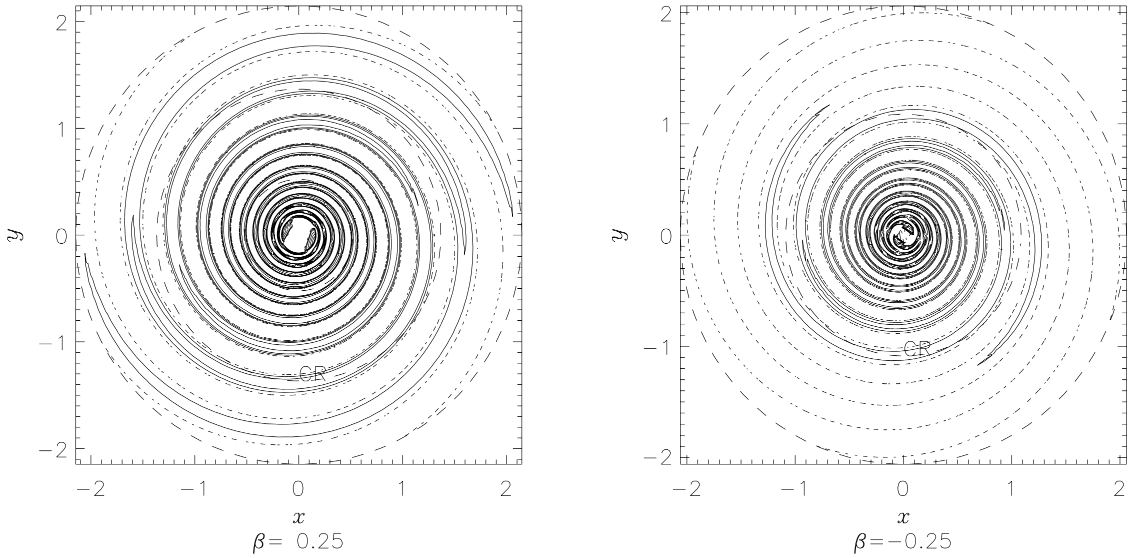


Figure 7.32: Density contour plot of growing modes in inner cut-out disks with $N = 2$, $\tilde{\sigma}_u = 0.1$. The left-hand plot is for $\beta = 0.25$, for which the fastest-growing mode has $s = 0.422$, $\Omega_{\rm p} = 0.704$. The right-hand plot is for $\beta = -0.25$, for which the fastest-growing mode has $s = 1.74$, $\Omega_{\rm p} = 0.930$. In each plot, the density is calculated between $R = 0.3R_{\rm ILR}$ and $R = R_{\rm OLR}$. The solid lines mark 10%, 20%, 40%, 60% and 80% of the maximum density within this range; the dotted lines show the nodes, where the density perturbation is zero. The dashed circles indicate the Lindblad resonances and the co-rotation radius. For $\beta = 0.25$, $R_{\rm ILR} = 0.522$, $R_{\rm CR} = 1.37$, $R_{\rm OLR} = 2.15$; for $\beta = -0.25$, $R_{\rm ILR} = 0.223$, $R_{\rm CR} = 1.09$, $R_{\rm OLR} = 2.06$, in units where $R_0 = 1$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

We can see physically why the winding should be related to the velocity dispersion. If a growing mode exists, then the stars cannot have enough random motion to move entirely out of a density enhancement within one growth time (see the discussion following fig. 6.6). As the temperature increases, the stars have more random motion, and the spiral modes can therefore be more loosely wound.

These points are illustrated in fig. 7.33, which plots the most unstable wavenumber $\alpha_{\rm u}$ against β and $\tilde{\sigma}_u$ for inner cut-out disks with $N = 2$, 3 and 4. As in Chapter 6, we define the most unstable wavenumber to be the dominant wavenumber of the marginal mode. This relies on the fact that the density transforms $A(\alpha)$ tend to be fairly strongly peaked about the dominant wavenumber (see figs. 7.19, 7.28). Our method of solving the integral equation discretises the wavenumber, so the values of $\alpha_{\rm u}$ shown below are correct to the nearest $\Delta\alpha$. This accounts for the slightly wobbly appearance of the curves.

The left-hand plot of fig. 7.33 confirms that modes are more tightly wound in disks with falling rotation curves and disks where the cut-out is gentler. The right-hand plot illustrates that modes are more tightly wound in cooler disks. The curves for different cut-out indices are much closer together than in the left-hand plot, indicating that most of the difference in the winding of marginal modes in disks with different cut-out indices can be ascribed to the different critical temperatures.

Modes in the cooler disks shown in fig. 7.32 also extend slightly further out towards the outer Lindblad resonance than the marginal modes shown in fig. 7.29. However, the mode in the disk with falling rotation curve still extends further than the mode in the disk with rising rotation curve (both in actual space,

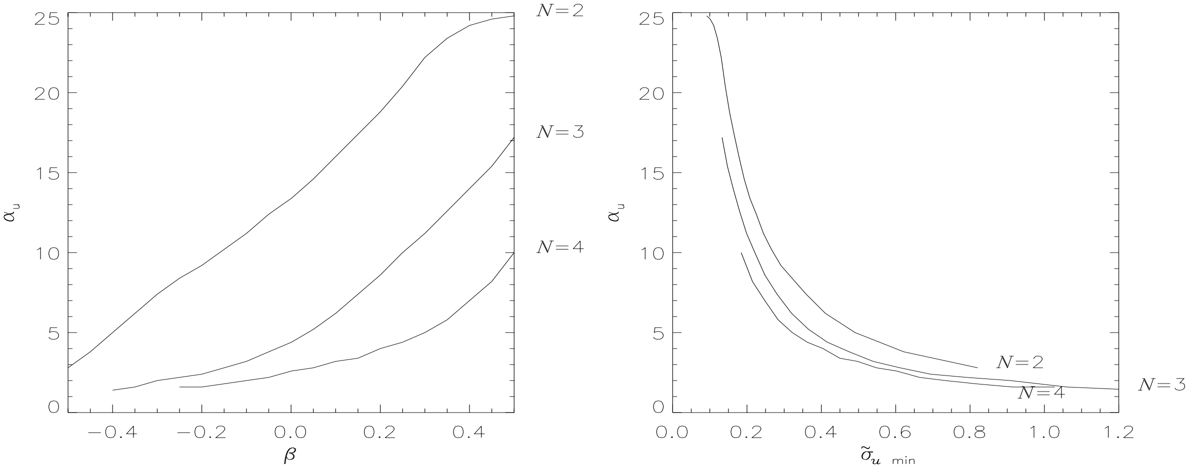


Figure 7.33: Dependence of most unstable wavenumber α_{u} on β and velocity dispersion for inner cut-out disks with $N = 2, 3, 4$. The most unstable wavenumber α_{u} is defined to be that wavenumber at which the amplitude of the marginal density transform, $|A(\alpha)|$, is maximum. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = -20$, $l_{\max} = +30$, $n_{GL} = 9$, $f_{\sigma} = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 2$.]

and in terms of the co-rotation radius). This suggests that rotation curve, as well as velocity dispersion, influences the extent of the mode.

The positions of the Lindblad resonances and the co-rotation radius are determined by the pattern speed Ω_{p} of the mode. We saw in fig. 7.27 that the pattern speed of modes increases with β and is greater for disks with lower cut-out indices. According to fig. 7.25, then, the pattern should be concentrated within a smaller radius for $\beta = 0.25$ than for $\beta = -0.25$, and similarly for the lower cut-out indices. This is indeed what we observe in figs. 7.29 - 7.31. Fig. 7.34 shows the position of the Lindblad and co-rotation resonances for marginal modes.

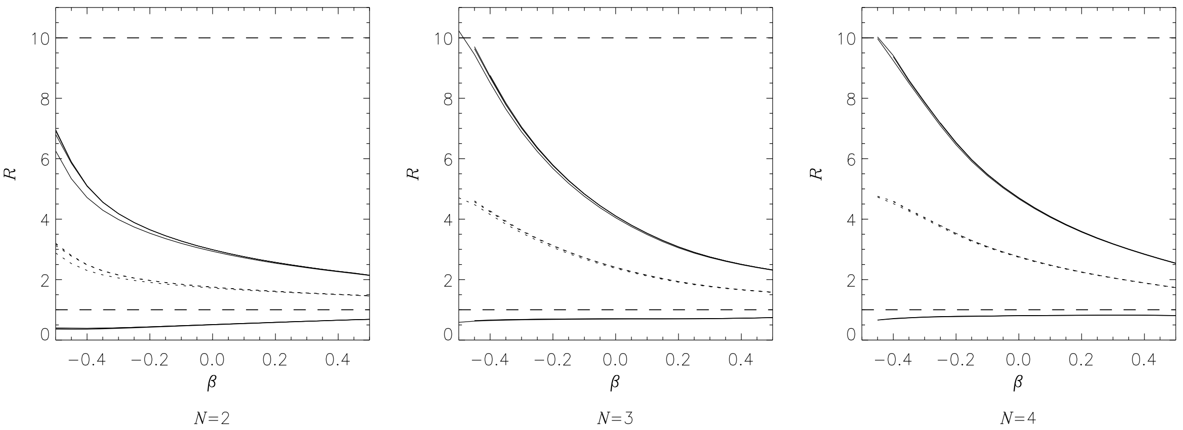


Figure 7.34: Position of resonances for marginal modes in cut-out disks with different β. The dashed horizontal lines show the position of the inner and outer cut-out radii. The pattern speed corresponding to the marginal mode in each disk (fig. 7.27) is used to calculate the position of the resonances in the disk. The solid lines mark the Lindblad resonances; the dotted line marks co-rotation. The three plots show results for $N = 2$, 3 and 4; each plot shows data for doubly cut-out disks with $M = 2, 4, 6$, $\tilde{R}_{\mathrm{c}} = 10.$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = -20$, $l_{\max} = +30$, $n_{GL} = 9$, $f_{\sigma} = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 2$.]

It is significant that for every value of β the pattern speed is such that the inner Lindblad resonance falls just inside the inner cut-out. It appears that the disk will not admit modes unless the pattern speed is large enough to push the inner Lindblad resonance safely within the inner cut-out. The resonance is allowed nearer the cut-out when the cut-out is sharper, and hence a more efficient reflector. Similarly, if we examine modes in an unstable disk (Appendix F), we find that the fastest-growing modes have the largest pattern speeds. As the pattern speed decreases, allowing the inner Lindblad resonance to move out closer to the cut-out radius, the growth rate slows. The pattern speed is also related to the winding of the spiral mode. In fig. 7.35, the dominant wavenumber is plotted against pattern speed for modes in $N = 2$ disks with various β. Modes which rotate quickly are more tightly wound.

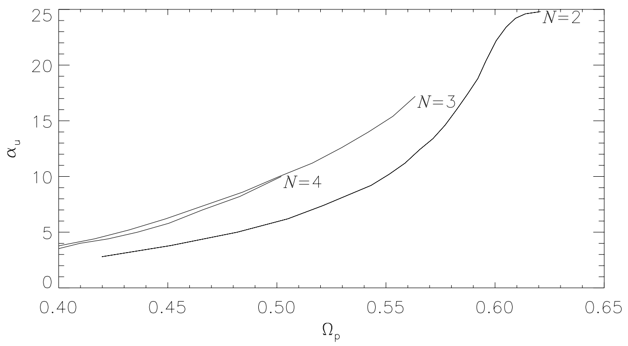


Figure 7.35: Dependence of dominant wavenumber $\alpha_{\rm u}$ on pattern speed $\Omega_{\rm p}$ for modes in inner cut-out disks with $N = 2$, 3, 4. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

The Ostriker-Peebles criterion

Ostriker and Peebles (1973) suggested that disk galaxies are stable to bar-like modes only when the ratio of total rotational energy T to total gravitational energy $|W|$ is less than 0.14. The energies W and T for the self-consistent disk are given by eq. (2.76) and eqs. (2.78) - (2.80). We thus recover the expressions given by Evans (1994):

$$\frac{T}{|W|} = \frac{\Gamma^2\left[1+\frac{\gamma}{2}\right]\Gamma^2\left[2+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]}{\beta\Gamma^2\left[\frac{1}{2}+\frac{\gamma}{2}\right]\Gamma^2\left[\frac{5}{2}+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]}, \qquad \beta > 0, \qquad (7.12)$$

$$\frac{T}{|W|} = \frac{\Gamma^2\left[1+\frac{\gamma}{2}\right]\Gamma^2\left[-\frac{3}{2}-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}{-\beta\Gamma^2\left[\frac{1}{2}+\frac{\gamma}{2}\right]\Gamma^2\left[-1-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}, \qquad \beta < 0, \qquad (7.13)$$

$$\frac{T}{|W|} = \frac{1}{1+\gamma}\frac{\Gamma^2\left[1+\frac{\gamma}{2}\right]}{\Gamma^2\left[\frac{1}{2}+\frac{\gamma}{2}\right]}, \qquad \beta = 0. \qquad (7.14)$$

Note that the $\beta = 0$ result can be derived from the values for $\beta \neq 0$ using a property of the gamma function (Gradshteyn & Ryzhik 1978, 8.328.2)

$$\lim_{|z|\to\infty} \frac{\Gamma[z+a]}{z^a\Gamma[z]} = 1. \qquad (7.15)$$

In the expression for negative β, we have

$$\lim_{\beta\to 0^-}\left(-1-\frac{1}{\beta}-\frac{\gamma}{\beta}\right)\frac{\Gamma^2\left[-\frac{3}{2}-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}{\Gamma^2\left[-1-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}=1,\qquad \lim_{\beta\to 0^-}-\frac{1}{\beta}\frac{\Gamma^2\left[-\frac{3}{2}-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}{\Gamma^2\left[-1-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}=\frac{1}{1+\gamma}. \tag{7.16}$$

Similarly in the expression for positive β, we have

$$\lim_{\beta\to 0^+}\left(\frac{5}{2}+\frac{1}{\beta}+\frac{\gamma}{\beta}\right)\frac{\Gamma^2\left[2+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]}{\Gamma^2\left[\frac{5}{2}+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]}=1,\qquad \lim_{\beta\to 0^+}\frac{1}{\beta}\frac{\Gamma^2\left[2+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]}{\Gamma^2\left[\frac{5}{2}+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]}=\frac{1}{1+\gamma}. \tag{7.17}$$

Fig. 7.36 shows the ratio $T/|W|$ calculated for the self-consistent disk (eqs. (7.12) and (7.13)). The anisotropy parameter used at each β is $\gamma=\gamma_{\max}$, i.e. the temperature needed for marginal stability in the cut-out disk. These values of γ are tabulated in Appendix E.

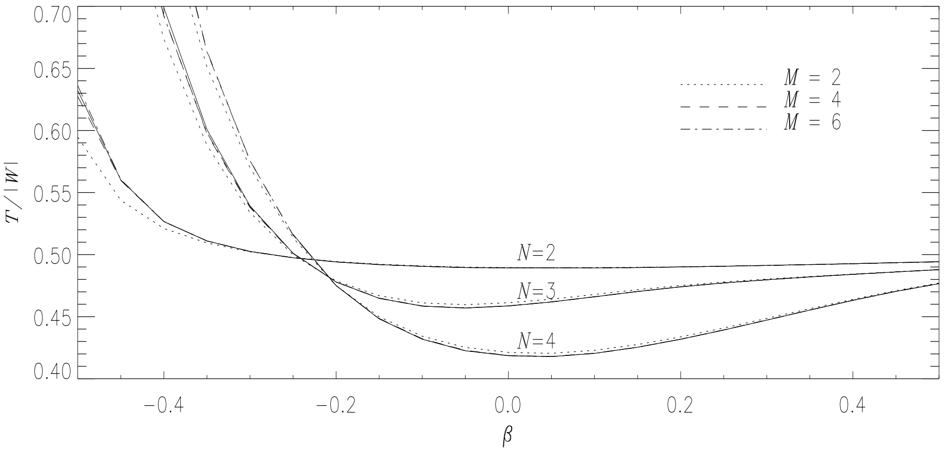


Figure 7.36: Comparison of the results for the cut-out disks with the Ostriker-Peebles criterion. The solid lines are labelled with the appropriate inner cut-out index N. For each N, different outer cut-out indices M are plotted, as shown in the key.

For all the disks, the ratio $T/|W|$ is much greater than the value of 0.14 suggested by Ostriker and Peebles. (Note that $T/|W|$ can exceed 0.5, since the standard form of the virial theorem does not apply to the power-law disks; see eq. (2.77).) This suggests that the Ostriker-Peebles criterion is not useful in predicting the stability of the power-law disks. However, the temperatures used in fig. 7.36 were derived for the cut-out disks, and we have already seen that the self-consistent and cut-out disks have very different stability properties. We should therefore consider the form of $T/|W|$ for the cut-out disks. In this case, the density Σ (2.43) and streaming velocity $\langle v\rangle$ (2.62) must be calculated numerically using the appropriate cut-out distribution function. The integrations over radius in the expressions for T and W (eqs. (2.73), (2.74)) must then also be performed numerically.

We used a Gaussian quadrature over eccentric velocity to calculate Σ and $\langle v\rangle$, and then found the energies by a midpoint integration over radius. The success of this approach was tested by comparing the numerical results for the self-consistent disk with those calculated analytically. This indicated that the accuracy was only 1 or 2 s.f., the error stemming from the Gaussian quadrature. This is sufficient for our purposes, which are to examine whether the Ostriker-Peebles criterion is a good rule of thumb.

Since the cut-out breaks the self-similarity, the ratio $T/|W|$ now depends on radius, as shown in fig. 7.37.

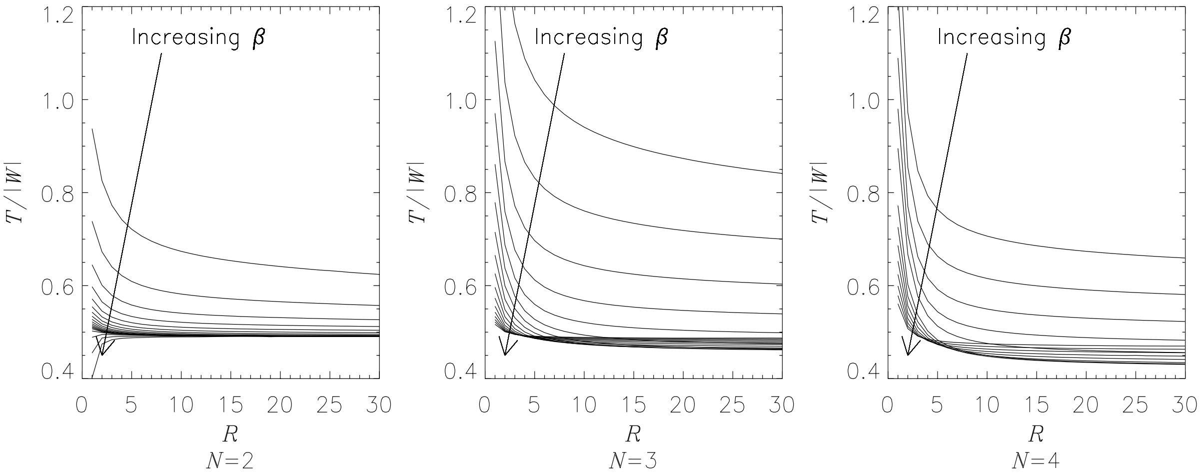


Figure 7.37: The ratio of the rotational to gravitational energy contained within a radius R, for inner cut-out disks. The solid lines are labelled with the appropriate inner cut-out index N. The arrows indicate the direction of increasing β, from $\beta = -0.5$ to $\beta = 0.4$.

For positive β, the $T/|W|$ curves have levelled off to a constant value at large radii. The convergence is slower for negative β. Eq. (2.8) shows that the cumulative mass rises much more slowly at large radii for disks with falling rotation curves than for disks with rising rotation curves. Thus we expect the total kinetic and gravitational energies within a radius R to converge faster with R when β is positive than when β is negative.

Fig. 7.38 plots the value of $T/|W|$ at $R = 30$, for a range of β.

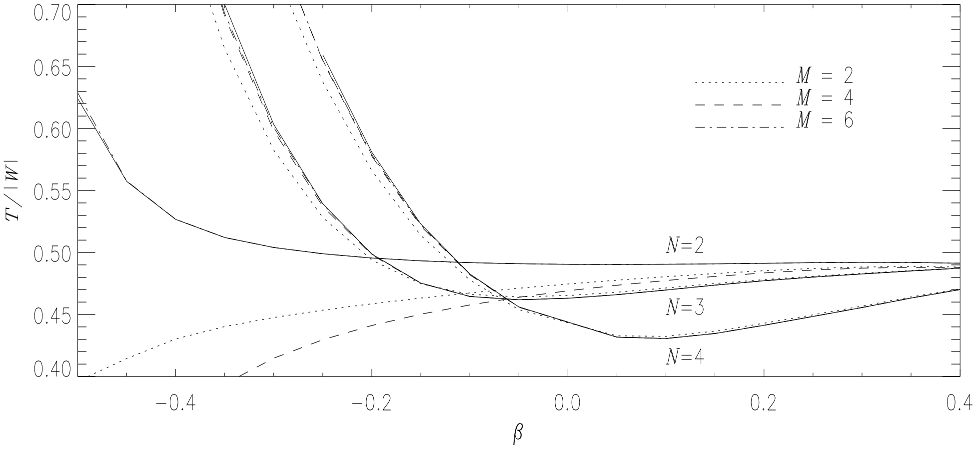


Figure 7.38: The ratio of the rotational to gravitational energy plotted against β, for doubly cut-out disks. The solid lines are labelled with the appropriate inner cut-out index N. For each N, different outer cut-out indices M are plotted, as shown in the key. The doubly cut-out disks all have $\tilde{R}_{\rm c} = 10$.

The curves are similar in form to those in fig. 7.36. The major difference concerns the doubly cut-out disks with $N = 2$, which depart from the curves for the $N = 2$ inner cut-out disk. (This may well be a numerical artefact. T and W are not evaluated very accurately, so an error of ≈ 0.1 in their ratio is entirely possible.) However, the main conclusion holds, namely that for the power-law disks the Ostriker-Peebles criterion greatly overestimates the amount of energy which must be in the form of random motion in order to achieve bisymmetric stability.

7.4 Neutral modes in the self-consistent disk

We close this chapter with a brief examination of the bisymmetric neutral modes admitted by the self-consistent disk. The response function $\mathcal{R}$ for these modes was derived in Chapter 4, eq. (4.12). Fig. 7.39 shows the dependence of the response function on the wavenumber α for different temperatures. As discussed in Chapter 5, it is difficult to obtain an accurate value for $\mathcal{R}$ when both temperature and wavenumber are high. In fig. 7.39, the curve for each temperature has been truncated when the inaccuracy becomes severe enough to be visible on the graph.

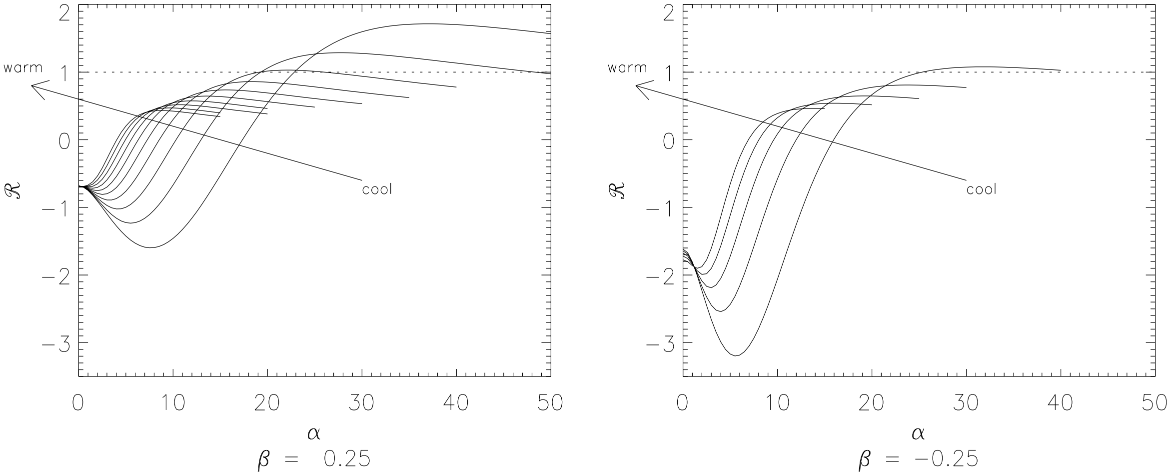


Figure 7.39: The response function for the self-consistent disks with $\beta = \pm 0.25$. For $\beta = 0.25$, curves are shown for ten values of Q_s in steps of 0.1 from 0.3 to 1.2. For $\beta = -0.25$, curves are shown for seven values of Q_s in steps of 0.1 from 0.3 to 0.9. [Numerical accuracy parameters: $l_\mathrm{min} = -20$, $l_\mathrm{max} = +30$, $n_{GL} = 12$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 50$, $b_\mathrm{acc} = 2.5$.]

The response is qualitatively similar to that for cut-out disks. When the temperature is low enough, the disk admits neutral $m = 2$ modes. The modes are more tightly wound in disks with lower temperatures. As the disk is heated, the $\mathcal{R}$-α curve drops, until a critical temperature is reached at which $\mathcal{R} = 1$ only at some wavenumber α_u. Above this temperature, the response remains less than unity for all wavenumbers, and the disk does not admit neutral modes.

It is apparent from fig. 7.39 that disks with falling rotation curves admit neutral modes at higher values of the stability criterion Q_s than disks with rising curves. For $\beta = 0.25$, the curve with $Q_\mathrm{s} = 0.4$ intersects the line $\mathcal{R} = 1$, whereas for $\beta = -0.25$, this curve never reaches $\mathcal{R} = 1$. This is the opposite of the behaviour of the critical temperature for marginally-stable modes in the cut-out disks. Fig. 7.26 shows that cut-out disks with falling rotation curves require lower values of Q_{s} for stability. We investigated the dependency of the critical temperature on β by finding the maximum of each $\mathcal{R}$-α curve at a given Q, and then adjusting Q until this maximum was unity. The results are shown in fig. 7.40 in terms of the velocity dispersion $\tilde{\sigma}_u$ and the stability criterion Q_{s}.

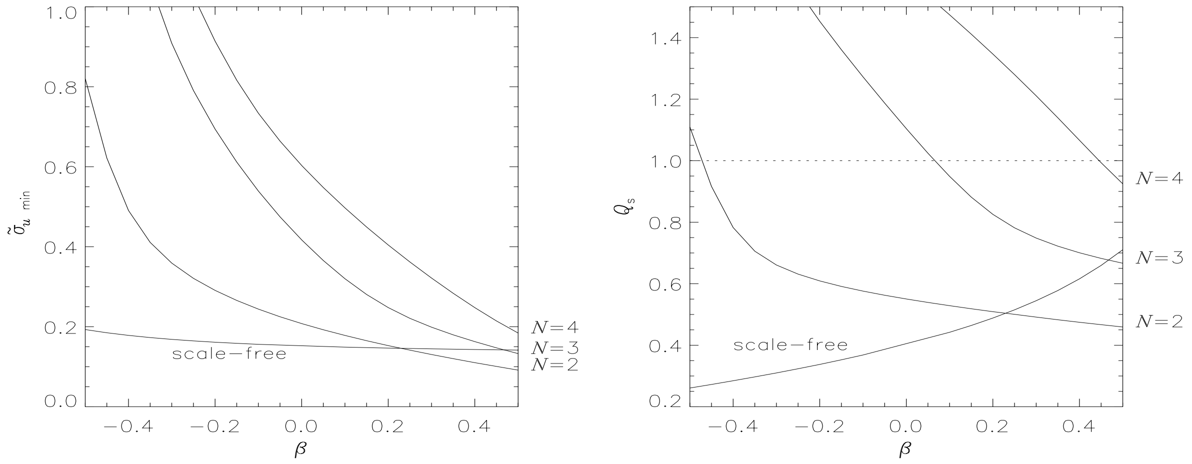


Figure 7.40: Velocity dispersion $\tilde{\sigma}_u$ and stability parameter Q_{s} plotted against β, for an $m = 2$ perturbation in self-consistent and cut-out disks. The curves labelled with the inner cut-out index N show the $\tilde{\sigma}_{u,\mathrm{min}}$ and Q_{s} at which the inner cut-out disks are marginally stable to bisymmetric disturbances. The curve labelled "scale-free" shows the $\tilde{\sigma}_{u,\mathrm{min}}$ and Q_{s} at which the self-consistent disk just fails to admit any bisymmetric neutral modes. [Numerical accuracy parameters: $l_{\mathrm{min}} = -20$, $l_{\mathrm{max}} = +30$, $n_{GL} = 12$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 50$, $b_{\mathrm{acc}} = 2.5$.]

In the limit of vanishing pattern speed, the inner Lindblad resonance moves far beyond the inner cut-out radius. We then expect the cut-out disk to respond like the self-consistent disk. If the self-consistent disk admits neutral modes, the cut-out disk should too. In our examination of the cut-out disks in section 7.3, we found no neutral modes. In the limit $\Omega_{\mathrm{p}} \rightarrow 0$, the eigenvalues approach the negative real axis (figs. 7.16 and 7.17). Thus neutral bisymmetric modes are not possible in cut-out disks at these temperatures. We can see why this is so by considering fig. 7.40. Figs. 7.16 and 7.17 were plotted for disks with $Q_{\mathrm{s}} = 1$. Fig. 7.40 shows that this temperature is high enough to abolish neutral modes in the self-consistent disk. Very low temperatures are required before neutral bisymmetric modes become possible (fig. 7.40 shows we require $\tilde{\sigma}_u \lesssim 0.15$, virtually independently of β). At these low temperatures, in most cut-out disks the neutral modes are already swamped by growing modes. However, fig. 7.40 indicates that when the rotation curve is falling steeply, a higher temperature is required to banish neutral modes from the self-consistent disk than to stabilise the $N = 2$ cut-out disk. We therefore investigated the eigenvalues of the $N = 2$ disk with $\beta = 0.4$ and $Q_{\mathrm{s}} = 0.5$. If we are correct in our supposition that the cut-out disk mimics the self-consistent disk in its response to perturbations with vanishingly small pattern speed, this disk should admit neutral modes even when it is stable to growing modes. This is indeed what is found. For example, when $\tilde{s} = 10^{-7}$ and $\tilde{\Omega}_{\mathrm{p}} = 10^{-4}$, the 39th eigenvalue is $1.03 - 0.00$.

As demonstrated in fig. 7.33, modes tend to be more tightly wound in cooler disks. Since the neutral modes set in only at very low temperatures, we expect them to be very tightly wound. This is illustrated in fig. 7.41. In the left-hand plot, the wavenumber of the neutral mode at the critical temperature is plotted against β. For comparison, the most unstable wavenumbers of the cut-out disks are also shown. In the right-hand plot of fig. 7.41, the dominant wavenumbers of the modes are plotted against the velocity dispersion at which they set in. (For the self-similar disk, this curve is close to vertical, since the neutral modes set in at almost the same velocity dispersion in all disks.) This figure indicates that the neutral modes are about as tightly-wound as would be expected from their low temperature.

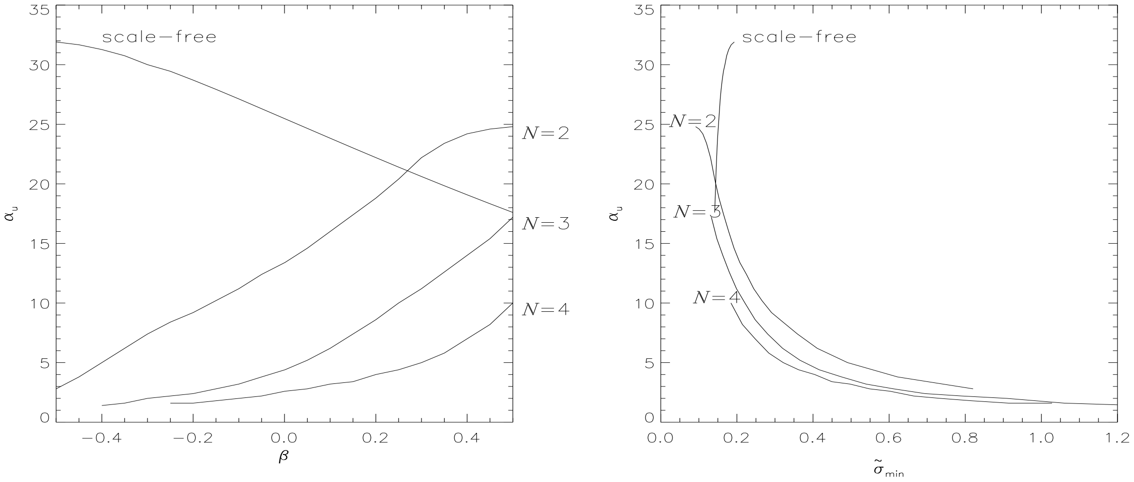


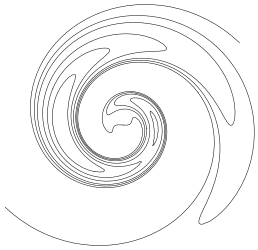
Figure 7.41: Dependence of most unstable wavenumber on β and velocity dispersion, for the self-consistent disk and for inner cut-out disks with $N = 2$, 3, 4. The curves labelled with the inner cut-out index N show the dominant wavenumber of the marginally-stable mode which sets in as the temperature is lowered. The curve labelled "scale-free" shows the wavenumber of the first neutrally-stable mode to set in as the temperature is lowered. In the left-hand plot, $\alpha_{\rm u}$ is plotted against β, and in the right-hand plot against $\tilde{\sigma}_{u,\min}$. [Numerical accuracy parameters: $l_{\min} = -20$, $l_{\max} = +30$, $n_{GL} = 12$, $f_\sigma = 0.8$, $a_{\rm acc} = 50$, $b_{\rm acc} = 2.5$.]

7.5 Summary

We began this Chapter by reviewing the results of local theory. We studied how waves could propagate through our disks at different temperatures, and sought to understand qualitatively how instabilities could arise. We then studied the global stability of the cut-out disks. As Toomre and Zang (1976) found for the disk with flat rotation curve, the power-law disks are remarkably stable to $m = 2$ perturbations. Gently cut-out disks are more prone to axisymmetric than to bisymmetric disturbances. Finally, we looked at the self-consistent disk. This admits neutral modes only at very low temperatures, at which most cut-out disks already admit growing modes.

Chapter 8

Global One-armed Modes



8.1 Introduction

The stability of the power-law disks to one-armed disturbances differs in several important respects from the bisymmetric stability. It is intuitively clear that one-armed disturbances are capable of moving the barycentre from the origin towards the overdense spiral arm. This cannot occur with modes for $m \neq 1$, since there are then matching overdense arms to balance the force. For an isolated system, acceleration of the barycentre is clearly not physical. The possibility occurs here because our analysis holds the equilibrium disk fixed while imposing a perturbation on it. If a one-armed disturbance were to occur in a physical system, the equilibrium disk would be displaced in such a way as to balance the perturbation, so that the barycentre remained fixed. Our analysis thus reproduces a famous “mistake”. Maxwell, in his celebrated Adams Prize essay “On the stability of motion of Saturn’s rings” (Maxwell 1859), examined the infinitesimal disturbances of a ring of small masses rotating in centrifugal equilibrium around a *fixed* central mass. A complete investigation of the $m = 1$ modes would require our analysis to be modified to permit the barycentre to move. Instead, let us investigate the stability of the disk to “naïve” one-armed disturbances, and check *a posteriori* that the barycentre moves only slightly.

8.2 Global stability of the cut-out disks

Once again, we are concerned to understand the position of the mathematical eigenvalue λ in the complex plane. We first investigate how the largest mathematical eigenvalue depends on pattern speed and growth rate for a particular temperature. We choose $Q_{\mathrm{s}} = 1$, i.e. the temperature at which the self-consistent disk is locally just stable to axisymmetric disturbances, and examine cut-out disks with inner cut-out index $N = 2$. As in figs. 7.9 - 7.11 in Chapter 7, the diamonds indicate the data points actually calculated. The smooth curves joining the data points were obtained using a spline fitting procedure.

For growing perturbations ($s \gtrsim 0.05$), these curves are qualitatively similar to those obtained for bisym-

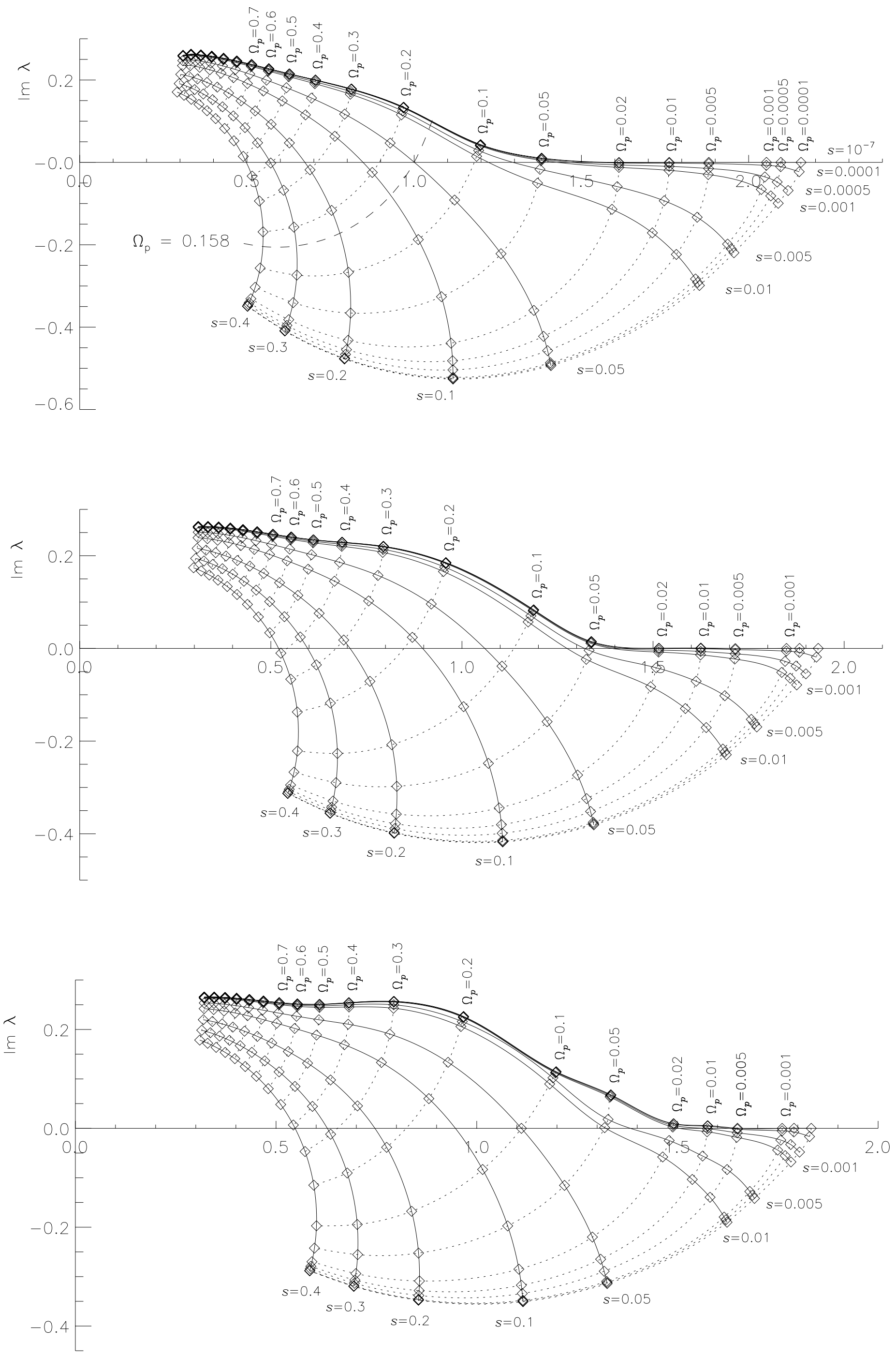


Figure 8.1: The dependence of the largest eigenvalue on growth rate and pattern speed, for inner cut-out disks with $N = 2$ and $Q_{\rm s} = 1$. In the uppermost plot, $\beta = 0.25$ ($\tilde{\sigma}_u = 0.283$ and $\gamma = 11.0$). In the middle plot, $\beta = 0.00$ ($\tilde{\sigma}_u = 0.378$ and $\gamma = 6.00$). In the lowest plot, $\beta = -0.25$ ($\tilde{\sigma}_u = 0.509$ and $\gamma = 3.36$). The solid lines show curves of constant s at 11 values of s: $s = 10^{-7}$, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4. The dotted lines show curves of constant $\Omega_{\rm p}$ at 20 values of $\Omega_{\rm p}$: $\Omega_{\rm p} = 0.0001$, 0.0005, 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3 … 1.2, 1.3. The dashed line in the top plot indicates the pattern speed of the fastest-growing mode in the $\beta = 0.25$ disk. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$.]

metric perturbations (figs. 7.9 - 7.11). Curves of a given s cross the real axis at higher values of these $m = 1$ plots than for $m = 2$, indicating that the disk is more prone to one-armed than two-armed disturbances. For instance, we saw in fig. 7.9 that $Q_{\mathrm{s}} = 1$ was already sufficient to stabilise the $\beta = 0.25$ disk to $m = 2$ disturbances. But fig. 8.1 indicates that this disk remains susceptible to $m = 1$ disturbances. One such mode is indicated by the dashed line in the top plot of fig. 8.1. The corresponding density transform $A(\alpha)$ is shown in fig. 8.2.

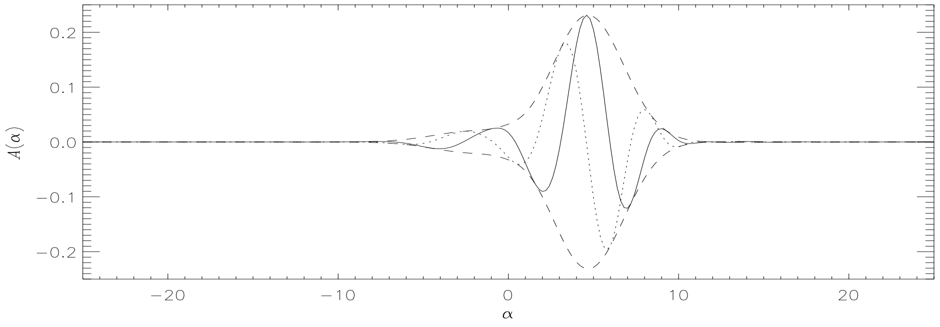


Figure 8.2: Density transform $A(\alpha)$ of a one-armed mode in an inner cut-out disk with $N = 2$, $\beta = 0.25$. The mode has pattern speed $\Omega_{\mathrm{p}} = 0.158$ and growth rate $s = 0.0502$. This is for $Q_{\mathrm{s}} = 1$, i.e. $\tilde{\sigma}_u = 0.283$ and $\gamma = 11.0$. The solid line is $\mathrm{Re}(A(\alpha))$, the dotted is $\mathrm{Im}(A(\alpha))$ and the dashed line shows the envelope $\pm|A(\alpha)|$. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\mathrm{min}} = -20$, $l_{\mathrm{max}} = +30$, $n_{GL} = 9$, $f_{\sigma} = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 2$.]

The density transform peaks at $\alpha = 4.6$. Fig. 8.3 shows a contour plot of the density corresponding to the mode, and for comparison the log-spiral with $\alpha = 4.6$. As for the $m = 2$ mode shown in fig. 7.20, the log-spiral accurately reflects the winding of the spiral arms, but not the density distribution within the arms.

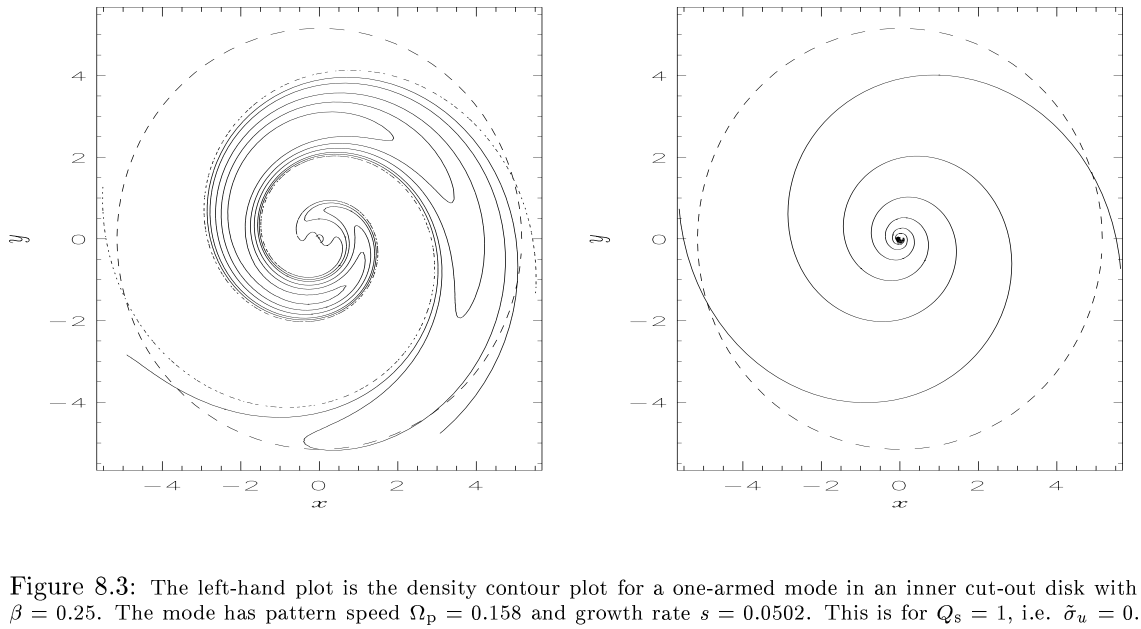


Figure 8.3: The left-hand plot is the density contour plot for a one-armed mode in an inner cut-out disk with $N = 2$, $\beta = 0.25$. The mode has pattern speed $\Omega_{\mathrm{p}} = 0.158$ and growth rate $s = 0.0502$. This is for $Q_{\mathrm{s}} = 1$, i.e. $\tilde{\sigma}_u = 0.283$ and $\gamma = 11.0$. In each plot, the density is calculated between $R = 0.03$ and $R = 1.1 R_{\mathrm{CR}}$. The solid lines plot 10%, 20%, 40%, 60% and 80% of the maximum density within this range; in the left-hand plot, the dotted lines show the nodes, where the density perturbation is zero. The dashed circle indicates the co-rotation radius, at 7.11. Units of $R_0 = 1$ are used. The right-hand plot shows the log-spiral with $m = 2$, $\alpha = 4.6$. Again, contours are plotted at 10%, 20%, 40%, 60% and 80% of the maximum density. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\mathrm{min}} = -30$, $l_{\mathrm{max}} = +40$, $n_{GL} = 9$, $f_{\sigma} = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 2$.]

However, the most striking difference between the present situation and that of Chapter 7 occurs as the growth rate is reduced to zero. It appears that in the limit of vanishing growth rate, the marginal eigenvalue curves never cross the real axis. Eigenvalues for small but non-zero growth rate (e.g. $s = 10^{-4}, 10^{-3}$) follow the marginal curve for high pattern speeds. As the pattern speed is reduced, one by one they peel away from the marginal curve and cross the real axis. This has dire consquences for the stability of the disks to one-armed disturbances. Ordinarily we expect to be able to stabilise disks by raising their temperature sufficiently. Fig. 8.4 shows the effect of increasing temperature. The largest mathematical eigenvalue curves are plotted for the same values of growth rate and pattern speed as in fig. 8.1, for $Q_\mathrm{s} = 1.0$, $Q_\mathrm{s} = 1.5$ and $Q_\mathrm{s} = 2.0$. The high-temperature results look very strange, and it is probable that the code has been pushed too far in the limits of very low growth rate, very low pattern speed and very high temperature, so that individual eigenvalues may be significantly in error. But it is clear that even these very high temperatures are not sufficient to stabilise the disk to modes with very small growth rates. As the temperature is increased, the curves with higher growth rate shrink towards the origin, crossing the real axis at successively lower values. But the curves with very low growth rate continue to stretch out along the real axis. It appears that, no matter how high the temperature, a mode exists for sufficiently low growth rate and pattern speed. Unlike for $m = 2$, there is no clear critical temperature distinguishing stable and unstable disks. We can find a temperature sufficient to stabilise the disk down to $|\omega| = 10^{-3}, 10^{-4}$ or whatever. But the numerical difficulties of calculating eigenvalues in the limit of vanishing growth rate and pattern speed suggest that there is little point in attempting to locate a critical temperature necessary for $m = 1$ stability.

In the latter two plots of fig. 8.4, the eigenvalue curves with vanishing or very small growth rate develop strange kinks in the vicinity of $\Omega_\mathrm{p} = 0.4$. This behaviour is investigated further in fig. 8.5, which shows the marginal curves for the first two eigenvalues in the $N = 2$, $\beta = 0.25$ disk, for several different temperatures.

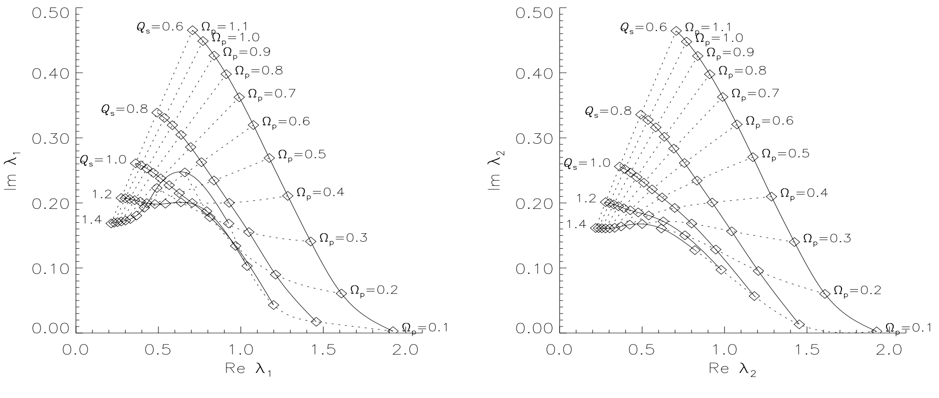


Figure 8.5: The dependence of the two largest mathematical eigenvalues on temperature and pattern speed for $\beta = 0.25$. The left-hand plot shows the largest eigenvalue λ_1, and the right-hand plot shows the second-largest eigenvalue λ_2. Curves are plotted for an $s = 10^{-7}$ perturbation in an inner cut-out disk with $N = 2$. The temperature used for each curve is expressed in terms of Q_s: $\tilde{\sigma}_u = 0.283 Q_\mathrm{s}$. The solid lines mark curves of constant Q_s for eleven values of Ω_p from 0.1 to 1.1. The dotted lines mark curves of constant Ω_p for five values of Q_s from 0.6 to 1.4. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -20$, $l_\mathrm{max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 1.5$.]

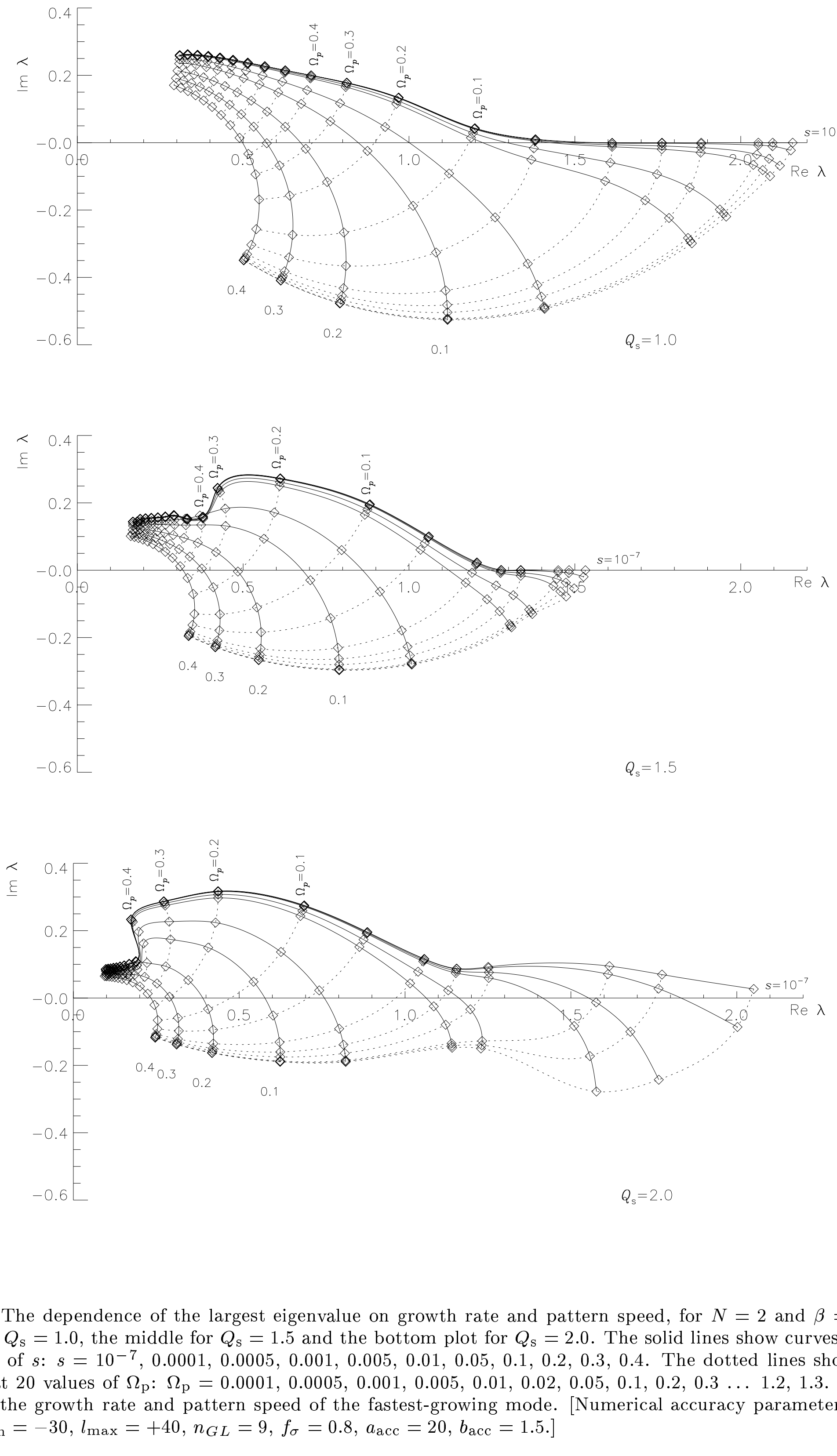


Figure 8.4: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = 0.25$. The top plot is for $Q_{\rm s} = 1.0$, the middle for $Q_{\rm s} = 1.5$ and the bottom plot for $Q_{\rm s} = 2.0$. The solid lines show curves of constant s at 11 values of s: $s = 10^{-7}$, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4. The dotted lines show curves of constant $\Omega_{\rm p}$ at 20 values of $\Omega_{\rm p}$: $\Omega_{\rm p} = 0.0001$, 0.0005, 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3 … 1.2, 1.3. The dashed lines indicate the growth rate and pattern speed of the fastest-growing mode. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$.]

At higher temperatures, the marginal eigenvalue curves develop a kink, which in the case of the largest eigenvalue means that the curve actually crosses that for lower temperatures. (Of course, the mapping from Q_s and Ω_p to λ is not conformal, and therefore the curves of constant Q_s and Ω_p are allowed to cross.) In the corresponding diagrams for $m = 2$ (figs. 7.21 - 7.23), the curves never crossed. It was then obvious that the effect of raising temperature, at a given pattern speed, was always to reduce the modulus of the eigenvalue, thus stabilising the disk. In fact, although this is not clear from fig. 8.5, the modulus of the eigenvalue at a given pattern speed always decreases as the temperature is raised. Thus raising temperature does still have a stabilising effect. It is unclear why, at high temperatures, the eigenvalue should change so abruptly with pattern speed.

We have seen (fig. 8.4) that for suitable growth rate and pattern speed, the largest eigenvalue can be made equal to unity, even at very high temperatures. Thus there is at least one mode. The existence of secondary, tertiary etc. modes depends on the behaviour of the sub-dominant eigenvalues. We find that the sub-dominant eigenvalues behave in a similar way to the dominant eigenvalue as the growth rate and pattern speed are varied. That is, for growing disturbances they move in arcs around the origin as the pattern speed decreases, whereas for disturbances of vanishing growth rate, they do not cross the real axis. Fig. 8.6 shows the dependence of the eight largest mathematical eigenvalues on pattern speed for $s = 0.1$.

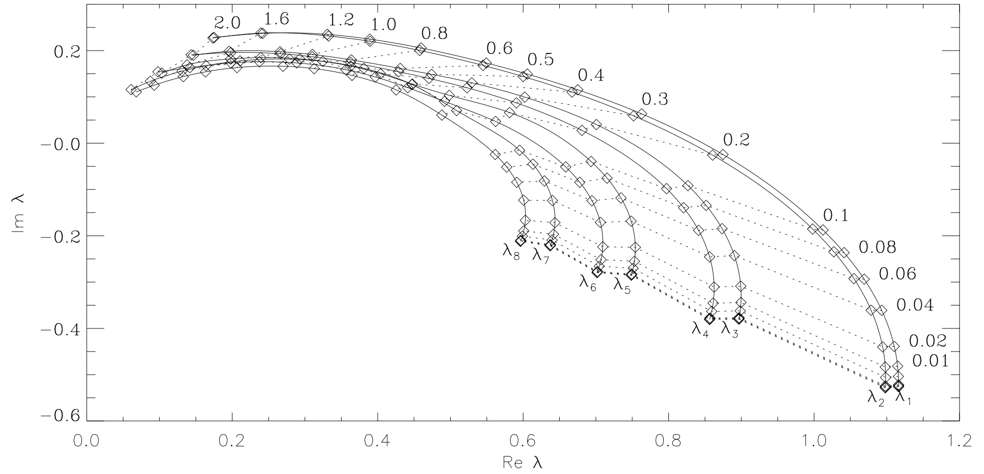


Figure 8.6: The dependence of the eight largest mathematical eigenvalues on pattern speed, for $s = 0.1$, $N = 2$ and $\beta = 0.25$. This is for $Q_\mathrm{s} = 1$, i.e. $\tilde{\sigma}_u = 0.283$ and $\gamma = 11.0$. The solid lines show marginal eigenvalue curves for the eight largest eigenvalues. The dotted lines show curves of constant Ω_p. 20 values of Ω_p are plotted: $\Omega_\mathrm{p} = 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.02, 0.04, 0.06, 0.08, 0.1, 0.2, 0.3 \ldots 2.0$. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_\mathrm{min} = -30$, $l_\mathrm{max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 1.5$.]

As in fig. 7.16, to a first approximation the pattern speed controls the phase of each eigenvalue. At $\Omega_\mathrm{p} = 0$, the phase of each eigenvalue λ_i is roughly $-25°$. As the pattern speed increases, the eigenvalues move anticlockwise. This plot also illustrates the tendency for $m = 1$ eigenvalues to occur in pairs. The density transforms for pairs of $m = 1$ modes (not shown) are very similar.

The behaviour for vanishing growth rate is very different. Fig. 8.7 traces the four largest marginal eigenvalue curves as the pattern speed is brought to zero. At high pattern speeds, the eigenvalues again occur in pairs. As the pattern speed is reduced, the eigenvalue curves draw together. This accounts for the ragged appearance of the curves, as eigenvalues swap positions, first one and then another becoming dominant. All four eigenvalues show the same reluctance to cross the real axis.

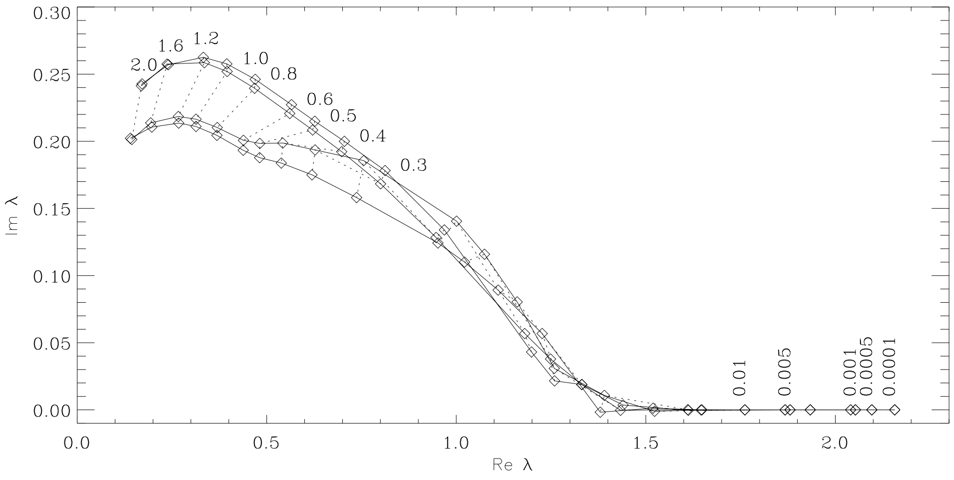


Figure 8.7: The dependence of the four largest mathematical eigenvalues on pattern speed, for $s = 10^{-7}$, $N = 2$ and $\beta = 0.25$. This is for $Q_{\rm s} = 1$, i.e. $\tilde{\sigma}_u = 0.283$ and $\gamma = 11.0$. The solid lines show marginal eigenvalue curves for the four largest eigenvalues. The dotted lines show curves of constant $\Omega_{\rm p}$, as labelled. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$.]

Finally, we study the complete eigenvalue spectrum in the limit $s \to 0$ and $\Omega_{\rm p} \to 0$. Fig. 8.8 shows eigenvalue spectra for various growth rates and pattern speeds in an inner cut-out disk with $N = 2$, $\beta = 0.25$, and $Q_{\rm s} = 1.0$. We immediately note that all the eigenvalues have positive real parts, in contrast to the situation for $m = 2$ eigenvalues (fig. 7.15). In the limit $s \to 0$, the largest eigenvalue becomes real. As $\Omega_{\rm p} \to 0$, it is joined on the real axis by increasingly many sub-dominant eigenvalues. The top-left plot, in particular, contains eigenvalues close to (1,0), supporting the idea that even a hot disk will admit modes with sufficiently low growth rates and pattern speeds. The hoop-like structure that develops in the limit of vanishing growth rate is probably a numerical artefact. A similar hoop develops for $m = 0$ modes in this limit, even though we know analytically that all the eigenvalues are in fact real. Symmetry arguments certainly suggest that all eigenvalues should be real in the limit $s \to 0$ and $\Omega_{\rm p} \to 0$ (see sections 4.2 and 4.3).

Having thoroughly investigated the behaviour of the eigenvalues, we now seek to understand it in physical terms. We have just noted that the real part of the $m = 1$ eigenvalues is always positive. Expressing this physically, the response of the power-law disks to a one-armed perturbation is *in phase* with the disturbance, whereas the response to a neutral two-armed perturbation is in *antiphase*. We can relate this to the experience of a star on an approximately circular orbit in the perturbation potential (fig. 8.9). The star is given a “push” outwards by the potential m times on each orbit. Its angular velocity is approximately Ω_0, so the frequency with which it is forced by the perturbation is $m\Omega_0$. Its natural radial frequency is κ_0. It is well known that harmonic oscillators respond in phase when driven at less than their natural frequency, and out of phase when driven at greater than their natural frequency.

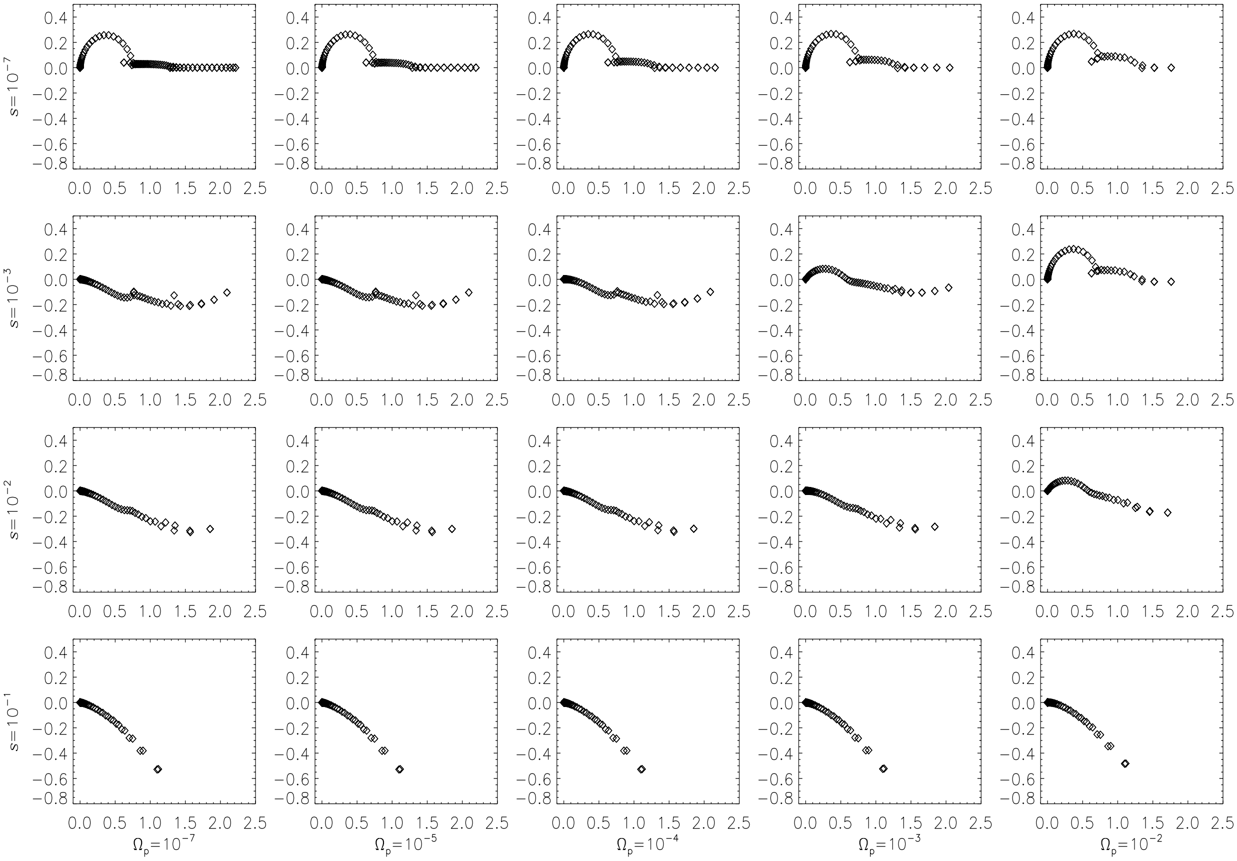
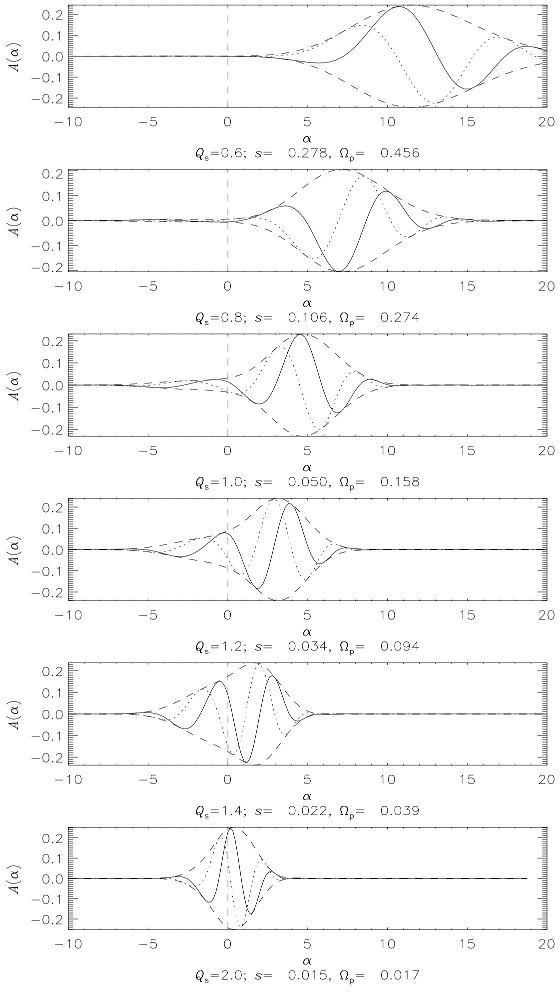


Figure 8.8: The eigenvalue spectra for various growth rates and pattern speeds in an inner cut-out disk with $N = 2$, $\beta = 0.25$, $Q_{\rm s} = 1.0$. The three rows of plots show eigenvalue spectra for increasing growth rates: $s = 10^{-7}$, 10^{-3}, 10^{-2}, 10^{-1}. The four columns show spectra for increasing pattern speeds: $\Omega_{\rm p} = 10^{-7}$, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2.0$.]

Figure 8.9: A star orbiting in a perturbation potential. The lighter line represents the star's original, circular orbit, and the heavier line its modified orbit (exaggerated!) due to the perturbation represented by the shaded region.

For the power-law disks, $\tilde{\Omega}_0 = 1$ and $\tilde{\kappa}_0 = \sqrt{2-\beta}$. For $m = 1$, therefore, the forcing frequency is always less than the natural frequency, and so the response is in phase. For $m = 2$, the forcing frequency is greater than the natural frequency, and the response is out of phase.

A further distinguishing feature of one-armed perturbations is that they are the only non-axisymmetric disturbances to have no inner Lindblad resonance (see eq. (7.3)). The inner Lindblad resonance acts as a damping mechanism by absorbing incident waves (section 7.2). In the absence of an inner Lindblad resonance, incoming trailing waves are reflected by the inner cut-out as leading waves, which are then swing-amplified. For $m = 2$, many of the incoming waves are absorbed by the Lindblad resonance and never return from the central regions. This point is illustrated in figs. 8.10 and 8.11. These show the density transforms of, respectively, $m = 1$ and $m = 2$ modes in disks of various temperatures. The disks are chosen so as to obtain modes of broadly similar growth rates. The top plot in each figure shows a quickly-growing mode in a cool disk. Both transforms are dominated by a large trailing component. Heating the disk reveals differences between the $m = 1$ and $m = 2$ modes. For $m = 1$, the presence of a large leading component – previously swamped by swing-amplified trailing waves – becomes apparent. Close to marginal stability, in the bottom plot, the leading and trailing components are of almost equal amplitude. However, for $m = 2$ modes, no such leading component exists. This indicates that few leading waves are returning from the central region to fuel the amplifier. This is presumably due to absorption by the Lindblad resonance, even though it lies within the inner cut-out radius.



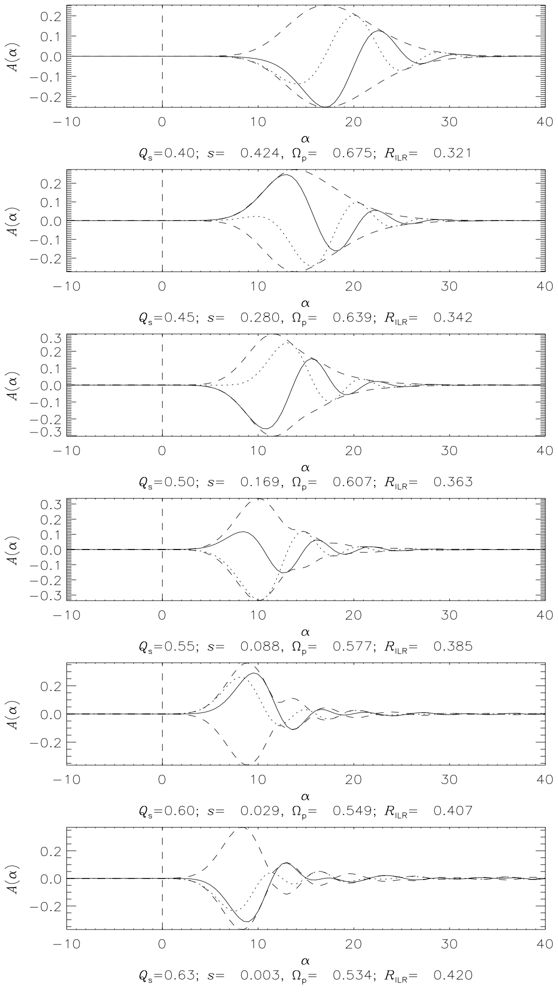


Figure 8.10: Density transforms for $m = 1$ modes in disks of various temperatures. The captions to each plot record the value of $Q_{\rm s}$, and the growth rate and pattern speed of the mode. In each plot, the solid curve shows $\mathrm{Re}(A(\alpha))$, the dotted line $\mathrm{Im}(A(\alpha))$, and the dashed line the envelope $\pm|A(\alpha)|$. The dashed vertical line separates leading from trailing waves. [$\beta = 0.25$, $N = 2$. Numerical accuracy parameters: $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$. Various different n and $\Delta\alpha$ were used.]

Figure 8.11: Density transforms for $m = 2$ modes in disks of various temperatures. The captions to each plot record the value of $Q_{\rm s}$, the growth rate and pattern speed of the mode, and the position of the inner Lindblad radius in units of R_0. In each plot, the solid curve shows $\mathrm{Re}(A(\alpha))$, the dotted line $\mathrm{Im}(A(\alpha))$, and the dashed line the envelope $\pm|A(\alpha)|$. The dashed vertical line separates leading from trailing waves. [$\beta = -0.25$, $N = 2$. Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.5$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

Growing modes

We are interested in how the rotation curve and inner cut-out affect the stability of the disks to $m = 1$ disturbances. We thus need a way of comparing the stability of different disks. For $m = 2$ perturbations, a convenient method was to find the critical temperature at which the disk is just stable to bisymmetric disturbances. The higher the temperature needed, the more unstable the disk. However, in the $m = 1$ case we are unsure whether such a critical temperature exists. Therefore we instead use the growth rate of the fastest-growing mode in disks with the same temperature as a measure of stability. More unstable disks admit faster-growing modes.

There is an ambiguity in defining what is meant by disks "of the same temperature". In this thesis, we have used three measures of temperature: the velocity dispersion $\tilde{\sigma}_u$, the anisotropy parameter γ and the stability parameter Q_s. These are interchangeable for a given disk, but differ when we compare disks with different β. To begin with, we examine disks with the same velocity dispersion. Fig. 8.12 shows the growth rates and pattern speeds obtained for the four fastest-growing modes in the $N = 2$ inner cut-out disks with $\tilde{\sigma}_u = 0.378$. This is the velocity dispersion at which the disk with $\beta = 0$ is just stable to axisymmetric disturbances.

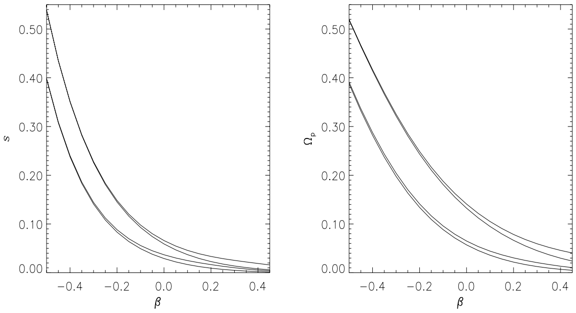


Figure 8.12: Growth rates and pattern speeds of the four fastest-growing modes in $N = 2$ inner cut-out disks with $\tilde{\sigma}_u = 0.378$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -20$, $l_\mathrm{max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 1.5$, $\epsilon_\lambda = 10^{-8}$.]

The growth rates rise steeply as β becomes more negative. Thus, as for bisymmetric perturbations, disks with rising rotation curves are less stable than those with falling curves. We also note that the modes come in pairs with similar growth rate and pattern speed. (We saw in fig. 8.6 that the $m = 1$ eigenvalues tend to come in pairs.)

An alternative way of comparing the stability of different disks is to compare disks with the same Q_s, rather than the same velocity dispersion $\tilde{\sigma}_u$. In fig. 8.12, we examined disks with $\tilde{\sigma}_u = 0.378$. At this velocity dispersion, disks with rising rotation curves are still unstable to axisymmetric modes. Thus although their $m = 1$ modes grow faster than in disks with falling rotation curves, they might actually be less important, relative to $m = 0$, than in disks with falling rotation curves. By examining disks with $Q_\mathrm{s} = 1$, we ensure that all the disks are stable to axisymmetric perturbations.

Fig. 8.13 shows the growth rates obtained for the four fastest-growing modes in four different inner cut-out disks with $Q_s = 1$. Fig. 8.14 shows the corresponding pattern speeds. The data used in these plots are recorded in Appendix F.

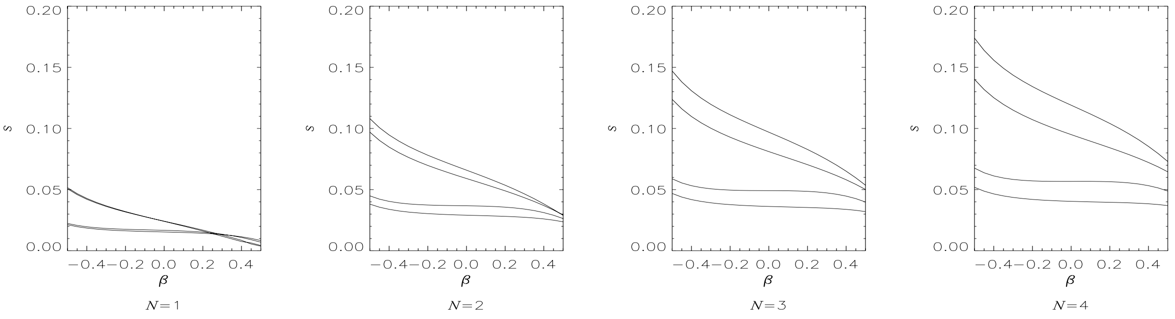


Figure 8.13: Growth rates of the four fastest-growing modes for inner cut-out disks with $Q_s = 1.0$. The four plots show results for $N = 1, 2, 3, 4$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = -20$, $l_{\max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 1.5$, $\epsilon_\lambda = 10^{-8}$.]

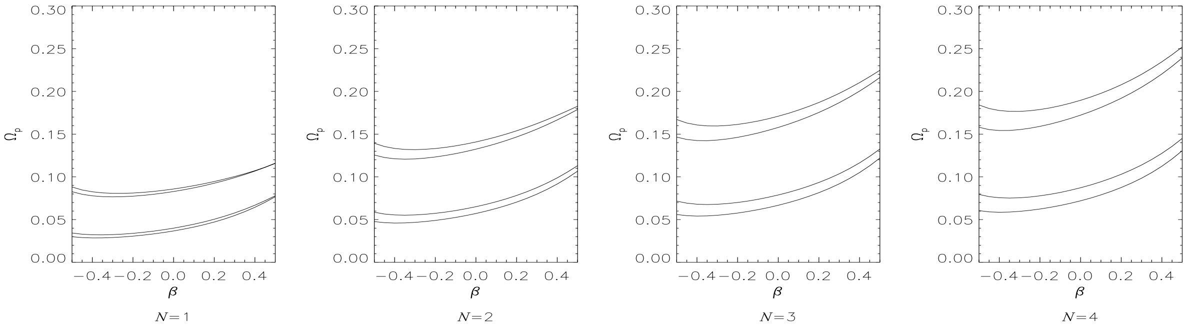


Figure 8.14: Pattern speeds of the four fastest-growing modes for inner cut-out disks with $Q_s = 1.0$. The four plots show results for $N = 1, 2, 3, 4$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = -20$, $l_{\max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 1.5$, $\epsilon_\lambda = 10^{-8}$.]

Even though the condition $Q_s = 1$ corresponds to larger velocity dispersions for negative β, modes still grow faster in disks with rising rotation curves. Thus disks with rising rotation curves are more unstable to both one-armed and axisymmetric modes than are disks with falling rotation curves, and they are even more unstable to one-armed modes than to axisymmetric modes.

As for $m = 2$ modes, disks with sharper cut-outs are more unstable than those where the cut-out is gentle. In section 7.3, we argued that this was because sharper cut-outs are more effective at reflecting the wave away from the damping inner Lindblad resonance. Where the cut-out was too gentle, as for $N = 1$, the wave was severely damped by the resonance, and modes were basically impossible. For $m = 1$ modes, there is no inner Lindblad resonance, so even disks with $N = 1$ admit growing modes. However, it appears that a sharper cut-out still destabilises the disk, presumably because the increased reflection makes a more efficient feedback circuit for swing amplification.

Fig. 8.15 examines the effect of lowering the temperature. It shows the growth rates and pattern speeds for inner cut-out disks with $N = 2$. The dotted lines are the results for $Q_\mathrm{s} = 1.0$, shown in figs. 8.13 and 8.14. The solid lines are for cooler disks, with $Q_\mathrm{s} = 0.8$. As expected, lowering the temperature destabilises the disk, so its modes grow and rotate more quickly.

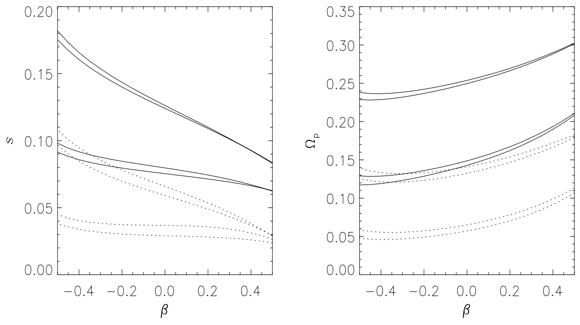
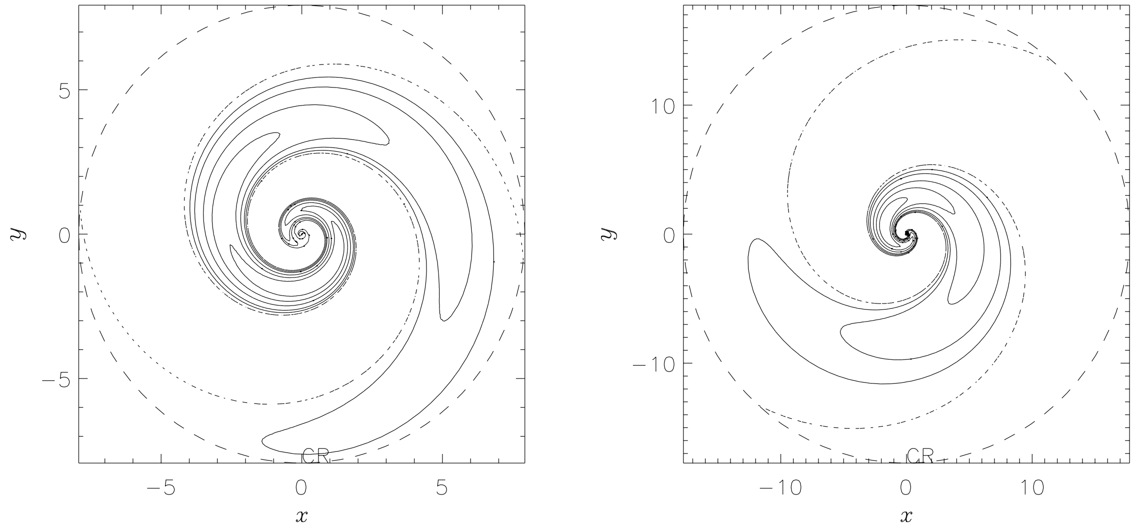


Figure 8.15: Comparison of growth rates and pattern speeds for $N = 2$ disks with $Q_\mathrm{s} = 1.0$ (dotted lines) and disks with $Q_\mathrm{s} = 0.8$ (solid lines). The left-hand plot shows the growth rates of the four fastest-growing modes, and the right-hand plot their pattern speeds. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -20$, $l_\mathrm{max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 1.5$.]

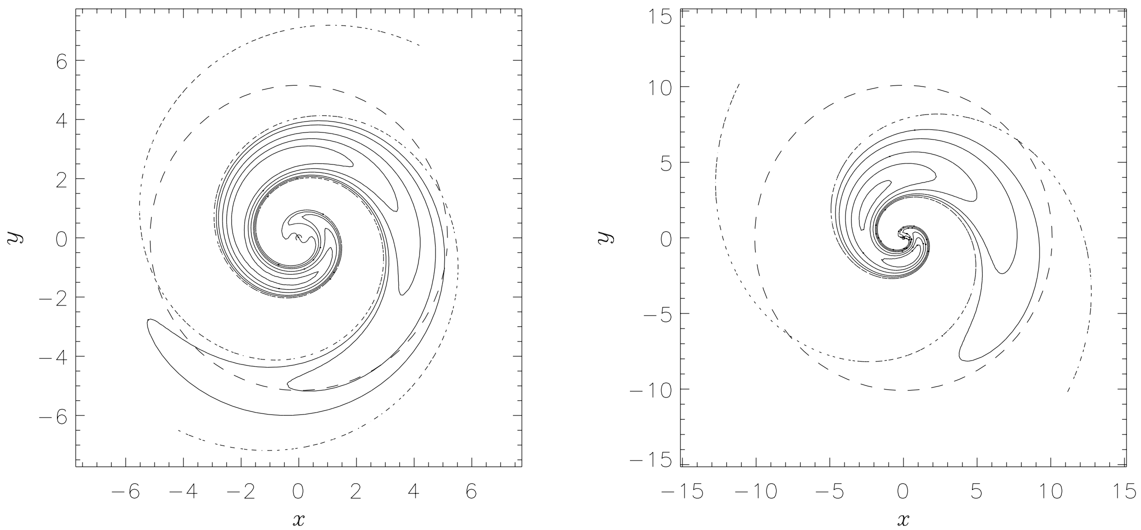
Figs. 8.16 - 8.19 show contour plots of the fastest-growing modes in inner cut-out disks with $Q_\mathrm{s} = 1$. α_u in the caption of each figure reports the dominant wavenumber of the modes, that is, the wavenumber at which the magnitude of the density transform is greatest.



$\beta =$ 0.25

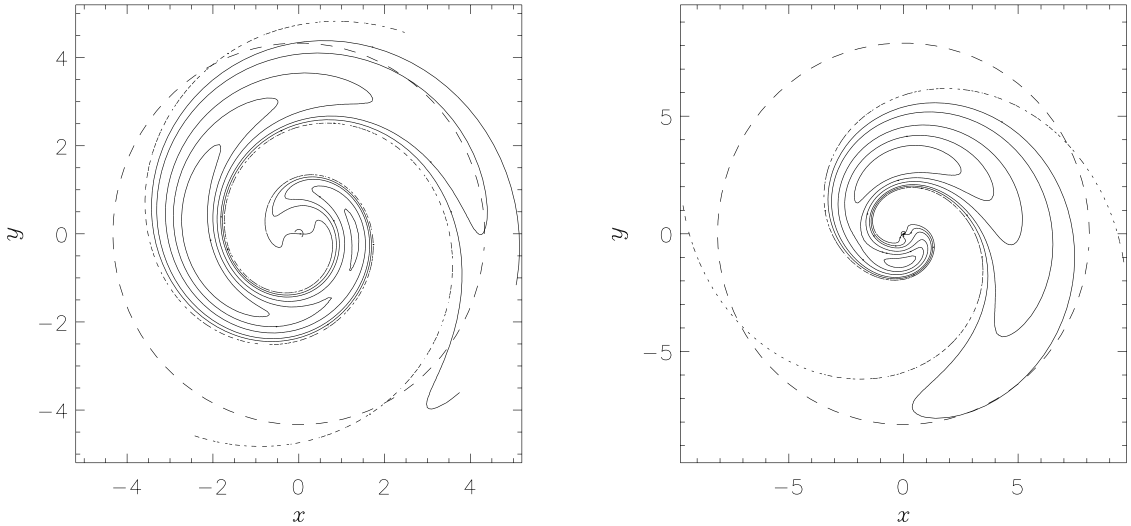
$\beta = -0.25$

Figure 8.16: Density contour plot of the fastest-growing $Q_\mathrm{s} = 1$ mode in inner cut-out disks with $N = 1$. The left-hand plot is for $\beta = 0.25$, for which the fastest-growing mode has $\Omega_\mathrm{p} = 0.0974$, $s = 0.0140$, $\alpha_\mathrm{u} \approx 4.2$. The right-hand plot is for $\beta = -0.25$, for which the fastest-growing mode has $\Omega_\mathrm{p} = 0.0807$, $s = 0.0343$, $\alpha_\mathrm{u} \approx 2.8$. In each plot, the density was calculated between $R = 0.02$ and R_CR. The solid lines mark 10%, 20%, 40%, 60% and 80% of the maximum density in this range; the dotted lines show the nodes, where the density perturbation is zero. The dashed circle indicates the co-rotation radius. For $\beta = 0.25$, $R_\mathrm{CR} = 7.93$, $R_\mathrm{OLR} = 16.8$; for $\beta = -0.25$, $R_\mathrm{CR} = 17.7$, $R_\mathrm{OLR} = 50.6$, in units where $R_0 = 1$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -30$, $l_\mathrm{max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 2$.]



β= 0.25 β=−0.25

Figure 8.17: Density contour plot of the fastest-growing $Q_\mathrm{s} = 1$ mode in inner cut-out disks with $N = 2$. The left-hand plot is for $\beta = 0.25$, for which the fastest-growing mode has $\Omega_\mathrm{p} = 0.158$, $s = 0.0502$, $\alpha_\mathrm{u} \approx 4.6$. The right-hand plot is for $\beta = -0.25$, for which the fastest-growing mode has $\Omega_\mathrm{p} = 0.132$, $s = 0.0816$, $\alpha_\mathrm{u} \approx 3.0$. In each plot, the density was calculated between $R = 0.02$ and $1.5R_\mathrm{CR}$. The solid lines mark 10%, 20%, 40%, 60% and 80% of the maximum density in this range; the dotted lines show the nodes, where the density perturbation is zero. The dashed circle indicates the co-rotation radius. For $\beta = 0.25$, $R_\mathrm{CR} = 5.16$, $R_\mathrm{OLR} = 10.9$; for $\beta = -0.25$, $R_\mathrm{CR} = 10.1$, $R_\mathrm{OLR} = 28.8$, in units where $R_0 = 1$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -30$, $l_\mathrm{max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 2$.]



β= 0.25 β=−0.25

Figure 8.18: Density contour plot of the fastest-growing $Q_\mathrm{s} = 1$ mode in inner cut-out disks with $N = 3$. The left-hand plot is for $\beta = 0.25$, for which the fastest-growing mode has $\Omega_\mathrm{p} = 0.192$, $s = 0.0789$, $\alpha_\mathrm{u} \approx 4.6$. The right-hand plot is for $\beta = -0.25$, for which the fastest-growing mode has $\Omega_\mathrm{p} = 0.160$, $s = 0.115$, $\alpha_\mathrm{u} \approx 2.8$. In each plot, the density was calculated between $R = 0.03$ and $1.2R_\mathrm{CR}$. The solid lines mark 10%, 20%, 40%, 60% and 80% of the maximum density in this range; the dotted lines show the nodes, where the density perturbation is zero. The dashed circle indicates the co-rotation radius. For $\beta = 0.25$, $R_\mathrm{CR} = 4.33$, $R_\mathrm{OLR} = 9.16$; for $\beta = -0.25$, $R_\mathrm{CR} = 8.11$, $R_\mathrm{OLR} = 23.1$, in units where $R_0 = 1$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = -30$, $l_\mathrm{max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 2$.]

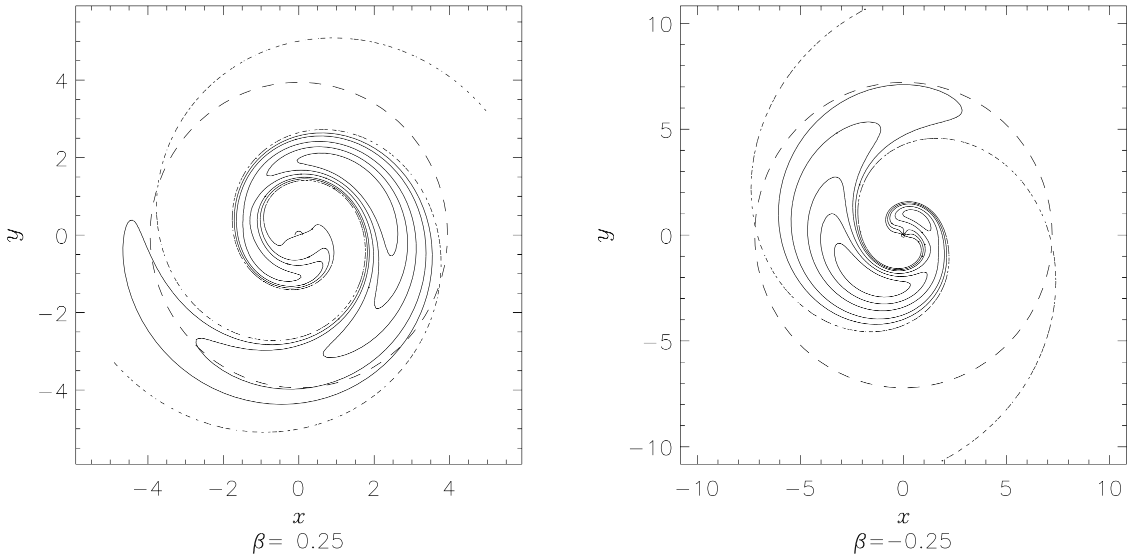


Figure 8.19: Density contour plot of the fastest-growing $Q_{\rm s} = 1$ mode in inner cut-out disks with $N = 4$. The left-hand plot is for $\beta = 0.25$, for which the fastest-growing mode has $\Omega_{\rm p} = 0.214$, $s = 0.100$, $\alpha_{\rm u} \approx 4.6$. The right-hand plot is for $\beta = -0.25$, for which the fastest-growing mode has $\Omega_{\rm p} = 0.177$, $s = 0.139$, $\alpha_{\rm u} \approx 2.8$. In each plot, the density was calculated between $R = 0.03$ and $1.5R_{\rm CR}$. The solid lines mark 10%, 20%, 40%, 60% and 80% of the maximum density in this range; the dotted lines show the nodes, where the density perturbation is zero. The dashed circle indicates the co-rotation radius. For $\beta = 0.25$, $R_{\rm CR} = 3.94$, $R_{\rm OLR} = 8.34$; for $\beta = -0.25$, $R_{\rm CR} = 7.22$,, $R_{\rm OLR} = 20.6$, in units where $R_0 = 1$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

As for the marginally stable bisymmetric modes plotted in figs. 7.29 - 7.31, the one-armed spirals shown here are more tightly wound in disks with falling rotation curves than disks with rising rotation curves. However, as in figs. 7.29 - 7.31, this reflects the dependence of winding on velocity dispersion rather than on rotation curve. The modes shown here are for $Q_{\rm s} = 1$, which corresponds to higher velocity dispersions for $\beta = -0.25$ ($\tilde{\sigma}_u = 0.509$) than for $\beta = 0.25$ ($\tilde{\sigma}_u = 0.283$). We saw in section 7.3 that spirals tend to be more tightly-wound in disks with lower velocity dispersions.

In accordance with the lack of an inner Lindblad radius, these one-armed modes extend right to the centre of the disk. Modes in disks with rising rotation curves have more density concentrated close to the centre than those in disks with falling rotation curves. When discussing the distance to which these modes extend out in the disk, it is useful to distinguish between the distance in physical space, and the distance in terms of the co-rotation radius. Modes in disks with low velocity dispersions generally extend further beyond the co-rotation radius than modes in warmer disks. However, modes in cooler disks also rotate more quickly, so their co-rotation and outer Lindblad radii are smaller. Thus the actual extent in space of a mode is generally smaller in cool disks than in warm disks. This is illustrated in fig. 8.20, which shows the fastest-growing modes in two disks with $Q_{\rm s} = 0.5$. When we compare this with the modes shown in fig. 8.17, we see that the pattern speeds are considerably larger in the cooler disks (for $\beta = 0.25$, $\Omega_{\rm p} = 0.564$ in the disk with $Q_{\rm s} = 0.5$, as against $\Omega_{\rm p} = 0.158$ in the disk with $Q_{\rm s} = 1.0$).

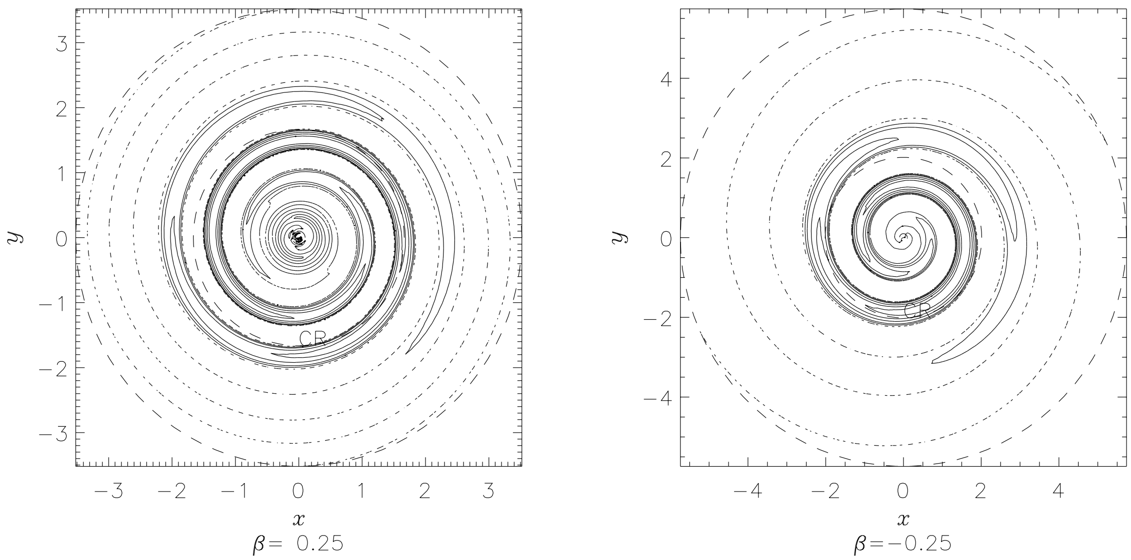


Figure 8.20: Density contour plot of the fastest-growing $Q_{\mathrm{s}} = 0.5$ mode in inner cut-out disks with $N = 2$. The left-hand plot is for $\beta = 0.25$, for which the fastest-growing mode has $\Omega_{\mathrm{p}} = 0.564$, $s = 0.444$, $\alpha_{\mathrm{u}} \approx 15$. The right-hand plot is for $\beta = -0.25$, for which the fastest-growing mode has $\Omega_{\mathrm{p}} = 0.542$, $s = 0.523$, $\alpha_{\mathrm{u}} \approx 9.2$. In each plot, the density was calculated between $R = 0.02$ and R_{OLR}. The solid lines mark 10%, 20%, 40%, 60% and 80% of the maximum density in this range; the dotted lines show the nodes, where the density perturbation is zero. The dashed circles indicates the co-rotation and outer Lindblad radius. For $\beta = 0.25$, $R_{\mathrm{CR}} = 1.66$, $R_{\mathrm{OLR}} = 3.52$; for $\beta = -0.25$, $R_{\mathrm{CR}} = 2.01$, $R_{\mathrm{OLR}} = 5.74$, in units where $R_0 = 1$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\mathrm{min}} = -30$, $l_{\mathrm{max}} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 2$.]

Bearing these temperature effects in mind, there is little difference in terms of spatial extent between one- and two-armed modes. The one-armed modes shown in figs. 8.17 - 8.19 extend further in space, but less far beyond the co-rotation radius, than the two-armed modes plotted in figs. 7.29 - 7.31. This reflects the higher temperature of the disks in which the $m = 1$ modes are plotted. The left-hand plot of fig. 8.20, showing $\beta = 0.25$, may be directly compared with the left-hand plot of fig. 7.29, since both these have $Q_{\mathrm{s}} = 0.5$. The extent of the one-armed and two-armed modes are then very similar, both in actual space and in terms of their co-rotation radius. The two-armed mode is slightly more tightly-wound ($\alpha_{\mathrm{u}} \approx 20$) than the one-armed mode ($\alpha_{\mathrm{u}} \approx 15$).

The effect of velocity dispersion on the form of the mode also explains why figs. 8.12 and 8.15 show different relationships between pattern speed and growth rate. In fig. 8.12, faster-growing modes rotate more rapidly; in fig. 8.15 the opposite is true. This happens because fig. 8.12 compares disks with the same velocity dispersion, whereas fig. 8.15 compares disks with the same stability parameter Q_{s}. Thus for disks with rising rotation curve, the velocity dispersion used for the dotted ($Q_{\mathrm{s}} = 1$) curves in fig. 8.15 is greater than that in fig. 8.12. For a given value of β, modes rotate more slowly in disks with larger velocity dispersions. This means that the pattern speeds of negative-β modes in fig. 8.15 must be smaller than their relatives in fig. 8.12. Conversely, for modes with falling rotation curves, the velocity dispersions of the modes in fig. 8.15 are lower than those in fig. 8.12. The modes therefore rotate more quickly, meaning that the pattern speeds of positive-β modes in fig. 8.15 are larger than their relatives in fig. 8.12. This explains why the pattern speed can fall with β in fig. 8.12, yet rise with β in fig. 8.15.

The position of the barycentre

One-armed modes are unique in that they are capable of displacing the barycentre of the disk away from the origin. However, visual inspection of figs. 8.17 - 8.19 suggests that in these modes the barycentre is close to the origin. The spiral arm wraps around the origin, with more density in the parts of the arm closer to the origin. It is difficult to guess by eye even the direction of any displacement of the barycentre from the origin. Intriguingly, the density distribution in many cases has two separate peaks, one on either side of the origin. This further suggests a balancing of the density on either side of the origin.

In general, the coordinates of the barycentre of the part of the disk within a radius L are given by

$$M(L)\bar{x} = \int_0^L dR \int_0^{2\pi} d\theta \Sigma(R,\theta) R^2 \cos\theta, \qquad M(L)\bar{y} = \int_0^L dR \int_0^{2\pi} d\theta \Sigma(R,\theta) R^2 \sin\theta, \tag{8.1}$$

where $M(L)$ is the mass contained within the radius L. Our analysis divides the density into two components: the axisymmetric equilibrium density $\Sigma_{\rm eq}$ (2.1), and the m-symmetric perturbation density $\Sigma_{\rm imp}$. From eqs. (3.25) and (3.26), we find that the perturbation density is given by

$$\Sigma_{\rm imp}(R,\theta) = \Sigma_{\rm p} e^{st} e^{im(\theta - \Omega_{\rm p} t)} \int_{-\infty}^{+\infty} d\alpha A(\alpha) \left(\frac{R_0}{R}\right)^{3/2 - i\alpha}, \tag{8.2}$$

where we take the real part to obtain the physical density. Each log-spiral component makes equal positive and negative contributions at every radius, so the mass $M(L)$ contained within a radius L is just the integral of the equilibrium density, eq. (2.8). Conversely, since the equilibrium density is axisymmetric, it does not shift the barycentre from the origin. Thus the expression (8.1) for the x-coordinate of the barycentre becomes

$$\begin{aligned} \frac{4\pi}{1-\beta}\left(\frac{L}{R_0}\right)^{1-\beta} \frac{\Sigma_0}{\Sigma_{\rm p}} \bar{x} = \frac{e^{(s - im\Omega_{\rm p})t}}{R_0^{1/2+i\alpha}} \int_{-\infty}^{+\infty} d\alpha A(\alpha) \int_0^L dR\, R^{1/2+i\alpha} \\ \times \int_0^{2\pi} d\theta \left\{ e^{i(m+1)\theta} + e^{i(m-1)\theta} \right\} \end{aligned} \tag{8.3}$$

where we must take the real part to obtain the physical co-ordinates. For $m \neq 1$, the integrals over θ are zero and we obtain $\bar{x} = \bar{y} = 0$. For $m = 1$, the terms with $\exp i(m+1)\theta$ integrate to zero, but the terms with $\exp i(m-1)\theta$ do not. We then obtain

$$\frac{2}{1-\beta}\left(\frac{L}{R_0}\right)^{1-\beta} \frac{\Sigma_0}{\Sigma_{\rm p}} \bar{x} = e^{(s - i\Omega_{\rm p})t} \int_{-\infty}^{+\infty} d\alpha A(\alpha) \int_0^L dR \left(\frac{R}{R_0}\right)^{1/2+i\alpha}, \tag{8.4}$$

$$\frac{2}{1-\beta}\left(\frac{L}{R_0}\right)^{1-\beta} \frac{\Sigma_0}{\Sigma_{\rm p}} \bar{x} = R_0 e^{(s - i\Omega_{\rm p})t} \int_{-\infty}^{+\infty} d\alpha A(\alpha) \left(\frac{L}{R_0}\right)^{3/2+i\alpha}. \tag{8.5}$$

Taking the real part, we obtain

$$\bar{x} = \frac{1-\beta}{2}\frac{\Sigma_{\rm p}}{\Sigma_0}R_0 e^{st}\left(\frac{L}{R_0}\right)^{1/2+\beta} \mathrm{Re}\left\{e^{-i\Omega_{\rm p}t}\int_{-\infty}^{+\infty} d\alpha A(\alpha)\left(\frac{L}{R_0}\right)^{i\alpha}\right\}. \tag{8.6}$$

Going through a similar process for the y-coordinate, we obtain

$$\bar{y} = -\frac{1-\beta}{2}\frac{\Sigma_{\rm p}}{\Sigma_0}R_0 e^{st}\left(\frac{L}{R_0}\right)^{1/2+\beta} \mathrm{Im}\left\{e^{-i\Omega_{\rm p}t}\int_{-\infty}^{+\infty} d\alpha A(\alpha)\left(\frac{L}{R_0}\right)^{i\alpha}\right\}. \tag{8.7}$$

If the integral over α is non-zero, the barycentre is displaced from the origin. As L is taken to infinity, the coordinates of the barycentre tend to infinity for $\beta > -\frac{1}{2}$. If we consider a large but finite value of L, eqs. (8.6) and (8.7) show that the barycentre starts at a finite distance from the origin, and gradually spirals outwards. The distance of the barycentre from the origin is given by

$$\sqrt{\bar{x}^2+\bar{y}^2} = \frac{1-\beta}{2}\frac{\Sigma_{\rm p}}{\Sigma_0}R_0 e^{st}\left(\frac{L}{R_0}\right)^{1/2+\beta} \left|\int_{-\infty}^{+\infty} d\alpha A(\alpha)\left(\frac{L}{R_0}\right)^{i\alpha}\right|. \tag{8.8}$$

We calculated this quantity for a range of L for the modes shown in figs. 8.17 - 8.19. We used $t = 0$ and $\Sigma_{\rm p} = \Sigma_0$. The results are shown in figs. 8.21 - 8.24. These consider the part of the disk within a radius L. For a range of L, they plot the quantity $\sqrt{\bar{x}^2+\bar{y}^2}/R_{\rm CR}$, i.e. the distance from the origin of the barycentre of the part of the disk within a radius L, in units of the co-rotation radius. The co-rotation radius roughly delimits the edge of the spiral mode, so $\sqrt{\bar{x}^2+\bar{y}^2}/R_{\rm CR}$ describes how far the barycentre is displaced from the origin, relative to the overall size of the mode. The dashed vertical lines in figs. 8.21 - 8.24 indicate the positions of the co-rotation and outer Lindblad resonances.

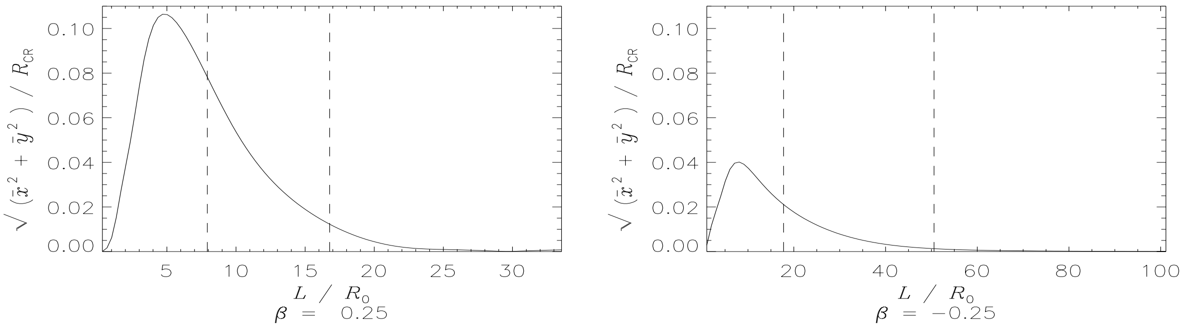


Figure 8.21: The displacement of the barycentre from the origin, for $N = 1$ disks with $\beta = \pm 0.25$. The displacement of the barycentre is plotted in units of the co-rotation radius. The dashed vertical lines mark the position of the co-rotation and outer Lindblad radii. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2.0$.]

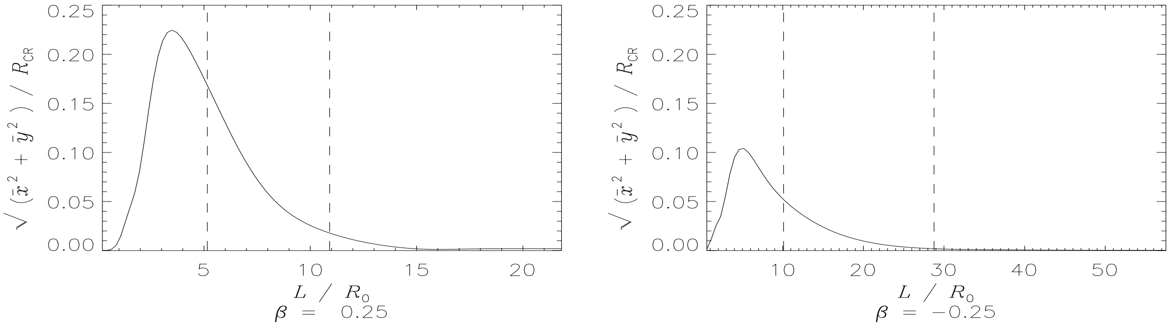


Figure 8.22: The displacement of the barycentre from the origin, for $N = 2$ disks with $\beta = \pm 0.25$. The displacement of the barycentre is plotted in units of the co-rotation radius. The dashed vertical lines mark the position of the co-rotation and outer Lindblad radii. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = -30$, $l_{\max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 2.0$.]

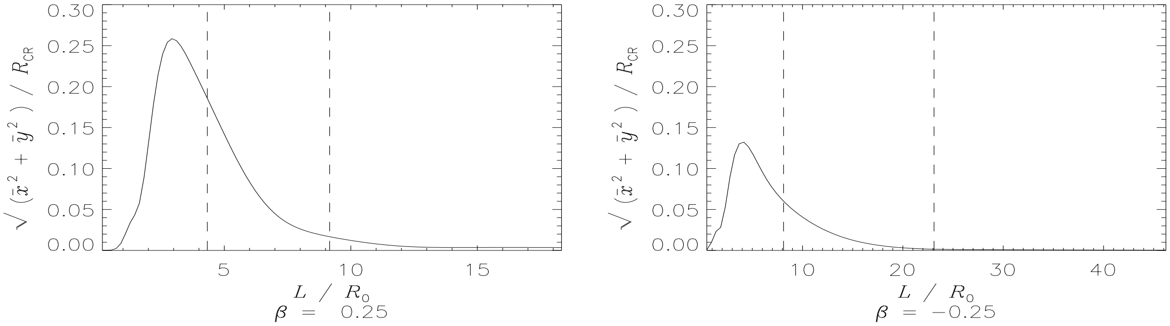


Figure 8.23: The displacement of the barycentre from the origin, for $N = 3$ disks with $\beta = \pm 0.25$. The displacement of the barycentre is plotted in units of the co-rotation radius. The dashed vertical lines mark the position of the co-rotation and outer Lindblad radii. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = -30$, $l_{\max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\ textacc} = 2.0$.]

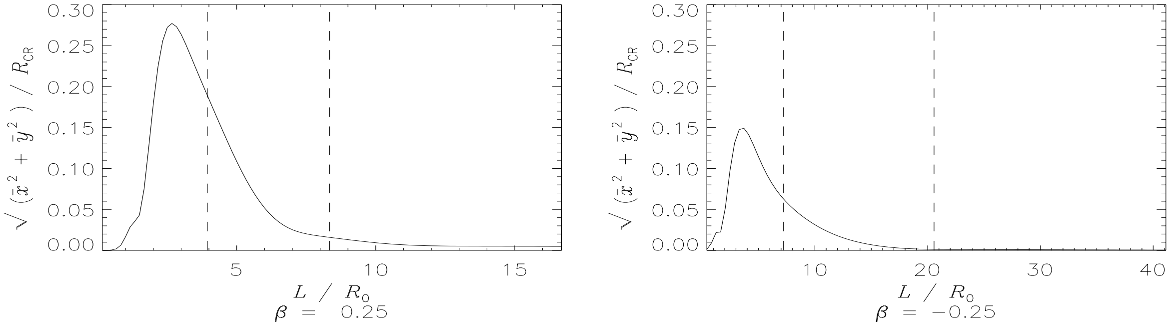


Figure 8.24: The displacement of the barycentre from the origin, for $N = 4$ disks with $\beta = \pm 0.25$. The displacement of the barycentre is plotted in units of the co-rotation radius. The dashed vertical lines mark the position of the co -rotation and outer Lindblad radii. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = -30$, $l_{\max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 2.0$.]

For small values of L, the displacement of the barycentre is approximately linear. This is as expected – if we consider only the very inner region of the disk, the emerging spiral arm is predominantly on one side of the origin. But as we move to considering larger regions of the disk, the contributions of the spiral arm from opposite sides of the origin begin to cancel out. For values of L beyond the outer Lindblad resonance, the part of the disk we are considering is now large enough to include the entire spiral mode. The forces pulling the barycentre away from the origin in different directions cancel almost exactly, so that the barycentre of the whole disk remains at the origin. This cancellation is especially remarkable in view of the factor $L^{1/2+\beta}$ in eq. (8.8). It implies that the integral $\int d\alpha A(\alpha) L^{i\alpha}$ decreases faster than $L^{-1/2-\beta}$.

One potential problem in evaluating the integral is aliasing. As L becomes very large, $\exp(i\alpha \ln L)$ varies very rapidly. The grid-spacing $\Delta\alpha$ must be very small for the integral to be evaluated accurately. As a rule of thumb, the integration is reliable where $\Delta\alpha \ln L \lesssim \pi/4$. This implies that for $\Delta\alpha = 0.2$, the integration is accurate up to $L \approx 50$. Experimentation with different grids of α confirms that figs. 8.21 - 8.24 accurately reflect the shape of the curve. The height of the maximum reached is affected by narrower grid spacing, but the main conclusion, that the displacement of the barycentre falls to zero at values of L beyond the outer Lindblad resonance, is unchanged.

In fact the fast oscillation of the term $\exp(i\alpha \ln L)$ is probably the reason for the cancellation. For typical modes, $A(\alpha)$ is a smooth function of α which tends to zero as $\alpha \to \pm\infty$. As L is increased, the term $\exp(i\alpha \ln L)$ must eventually vary much faster with α than does $A(\alpha)$ (*unless* $A(\alpha)$ is a delta-function or some other strange entity – see section 8.3). Thus the integration over wavenumber in eq. (8.8) adds up many positive and negative contributions of equal magnitude. These cancel, ensuring that the integral vanishes. Thus it appears that self-consistent modes do not shift the barycentre of the disk, at least in the linear regime.

8.3 Neutral modes in the self-consistent disk

In this section we briefly examine the neutral modes admitted by the self-consistent disk. The response function $\mathcal{R}$ for these modes was derived in Chapter 4, eq. (4.12). Fig. 8.25 shows the dependence of the response function on the wavenumber α for different temperatures. As discussed in section 5.4, it is difficult to obtain an accurate value for $\mathcal{R}$ when both temperature and wavenumber are high. In fig. 8.25, the curve for each temperature has been truncated when the inaccuracy becomes severe enough to be visible on the graph. These graphs suggest that, no matter how high the temperature, the disk admits neutral modes. Raising the temperature reduces the wavenumber of the mode – spirals are looser in hotter disks – but there is no indication that even very high temperatures will suffice to abolish neutral modes, since it appears that for all temperatures $\mathcal{R} > 1$ at $\alpha = 0$, and falls below unity as $\alpha \to \infty$.

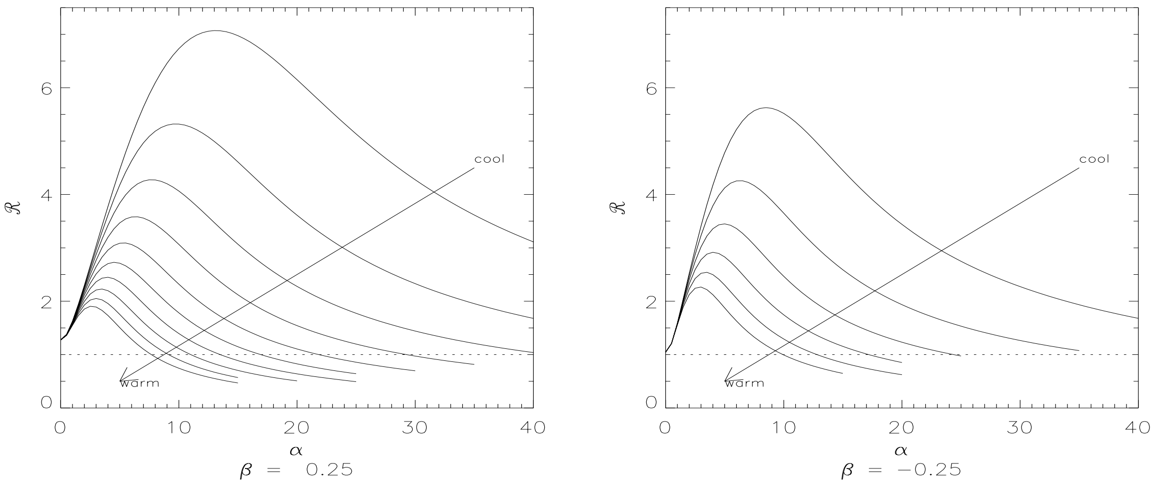


Figure 8.25: The response function for the self-consistent disks with $\beta = \pm 0.25$. For $\beta = 0.25$, ten values of $Q_{\rm s}$ in steps of 0.1 from 0.3 to 1.2. For $\beta = -0.25$, seven values of $Q_{\rm s}$ in steps of 0.1 from 0.3 to 0.9. [Numerical accuracy parameters: $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 12$, $f_\sigma = 0.8$, $a_{\rm acc} = 50$, $b_{\rm acc} = 2.5$.]

We have seen that the growing, rotating self-consistent modes in figs. 8.17 - 8.19 do not shift the barycentre away from the origin. However, if $A(\alpha)$ is a delta-function $\delta(\alpha - \alpha_0)$, eqs. (8.6) and (8.7) become

$$\bar{x} = \frac{1-\beta}{2}\frac{\Sigma_{\rm p}}{\Sigma_0} R_0 e^{st} \left(\frac{L}{R_0}\right)^{1/2+\beta} \cos\left(\alpha_0 \ln\frac{L}{R_0} - \Omega_{\rm p} t\right), \tag{8.9}$$

and

$$\bar{y} = -\frac{1-\beta}{2}\frac{\Sigma_{\rm p}}{\Sigma_0} R_0 e^{st} \left(\frac{L}{R_0}\right)^{1/2+\beta} \sin\left(\alpha_0 \ln\frac{L}{R_0} - \Omega_{\rm p} t\right). \tag{8.10}$$

The direction of the displacement of the barycentre from the origin depends on the wavenumber of the mode, but its magnitude is independent of wavenumber. For modes where $A(\alpha)$ is non-zero over a range of wavenumbers, therefore, the different wavenumbers pull the barycentre equally in all directions at once. Overall, they cancel out, and the barycentre remains at the origin. For a single log-spiral, no such cancellation occurs.

For these neutral modes, $s = \Omega_{\rm p} = 0$. The barycentre is therefore stationary. The magnitude of its displacement from the origin, in terms of the radius L, is

$$\frac{\sqrt{\bar{x}^2 + \bar{y}^2}}{L} = \frac{1-\beta}{2}\frac{\Sigma_{\rm p}}{\Sigma_0}\left(\frac{R_0}{L}\right)^{1/2-\beta}. \tag{8.11}$$

Single $m = 1$ log-spirals always tend to move the barycentre. However, in disks with $\beta < \frac{1}{2}$, as $L \to \infty$ the displacement of the barycentre becomes insignificant compared to the radius of the disk. This is because the perturbation surface density falls off as $R^{-3/2}$, whereas the surface density of the equilibrium disk falls off as $R^{-1-\beta}$. For values of $\beta < \frac{1}{2}$, therefore, the perturbation density falls off more rapidly than the equilibrium disk. As we go to large radii, the axisymmetric equilibrium disk will always “win”. When $\beta > \frac{1}{2}$, the neutral mode can actually move the barycentre, in seeming violation of Newton’s Laws.

This happens simply because the perturbation density falls off more slowly than the equilibrium density, a situation of course in contradiction with the assumptions of our linear analysis.

8.4 Summary

In this Chapter, we have studied the stability of the power-law disks to one-armed modes. As Toomre and Zang (1976) found for the disk with flat rotation curve, the power-law disks are very prone to one-armed disturbances. Growing $m = 1$ modes remain at $Q_{\mathrm{s}} = 1$ and beyond. Neutral modes remain possible, in both the cut-out and the self-consistent disks, up to indefinitely high temperatures. One possible concern about our analysis is that single $m = 1$ log-spirals can in principle move the barycentre of the disk. However, we have shown (analytically and numerically) that modes built of many log-spirals do not move the barycentre substantially.

Chapter 9

The Other Angular Harmonics

9.1 Introduction

In this chapter, we briefly examine modes with $m = 3$ and 4. We dub these respectively triskele and tetraskele modes, after the three- and four-legged devices common in early European art. In real galaxies, three- or four-armed spirals are much less prevalent than spirals with one or two arms. However, Fourier decomposition of the spiral patterns of real galaxies indicate significant contributions from $m = 3$, $m = 4$ (Elmegreen & Elmegreen 1985). Swing amplification also suggests that multi-armed patterns could be favoured.

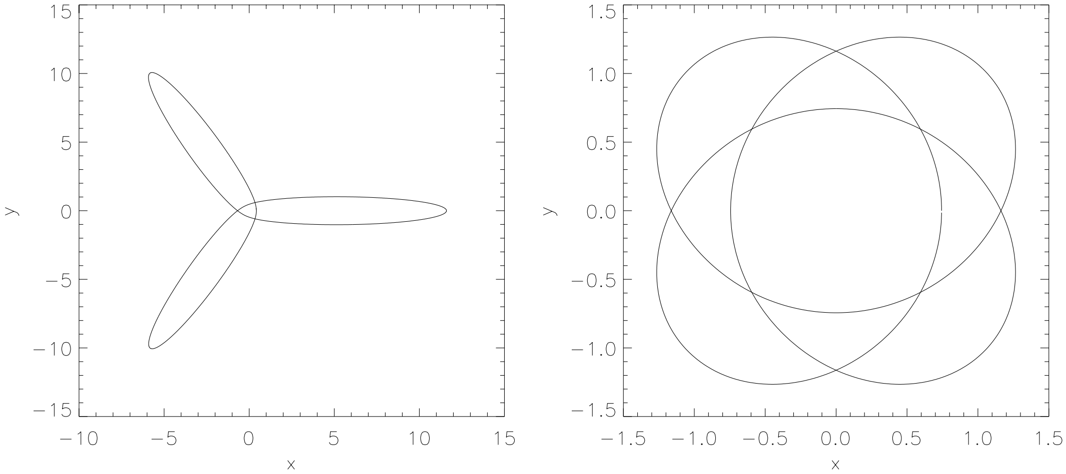
9.2 Numerical difficulties

Modes with $m \geq 2$ present an additional numerical problem. The expressions for the angular momentum function derived in Appendix C involve division by the term $l\tilde{\kappa} + m\tilde{\Omega}$. They therefore fail when $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes, although the integral still exists in a principal value sense. In the modes we have examined so far, this problem has not arisen. For $m = 0$, $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes at the $l = 0$ radial harmonic; but we have seen that at this harmonic the angular momentum function F_{00} should be set to zero for both the self-consistent and cut-out disks. For $m = 1$ and $m = 2$, $l\tilde{\kappa} + m\tilde{\Omega}$ never vanishes. But for higher angular harmonics, such as $m = 3$ and $m = 4$, the term $l\tilde{\kappa} + m\tilde{\Omega}$ may vanish at some eccentric velocity.

What does it mean physically when $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes? The time the star takes to perform l complete revolutions, $2\pi|l|/\Omega$, is then identical to the time it takes to perform m radial oscillations, $2\pi m/\kappa$. Thus the star's orbit closes in an inertial frame. Two examples are shown in fig. 9.1.

Closed orbits such as these do not occur in all disks. For $l\tilde{\kappa} + m\tilde{\Omega}$ to vanish, then in terms of the auxiliary integral defined in (2.27)

$$\mathcal{J}_2(\tilde{U}) = -2\pi\frac{l}{m}.$$



$\tilde{U}$=1.635 $\tilde{U}$=0.439

Figure 9.1: Closed orbits in a $\beta = 0.25$ disk. The left-hand plot shows the orbit of a star with $\tilde{U} = 1.64$, $R_{\rm H} = 1$. The orbit closes after 3 radial oscillations and 2 complete revolutions. The right-hand plot shows the orbit of a star with $\tilde{U} = 0.439$, $R_{\rm H} = 1$. The orbit closes after 4 radial oscillations and 3 complete revolutions.

In section 2.2 we saw that $\mathcal{J}_2(\tilde{U})$ declines monotonically from $2\pi/\sqrt{2-\beta}$ at $\tilde{U} = 0$ to π at $\tilde{U} = \infty$. Thus $l\tilde{\kappa} + m\tilde{\Omega}$ will vanish for some $\tilde{U}$ if there is an integer l which satisfies

$$\frac{1}{2} < -\frac{l}{m} \leq \frac{1}{\sqrt{2-\beta}}. \tag{9.1}$$

Whether or not this occurs depends on the values of m and β. For $m = 3$, according to eq. 9.1, $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes at $l = -2$ for $\beta \geq -0.25$, while for $\beta < -0.25$ it never vanishes.For $m = 4$, $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes at $l = -3$ for $\beta \geq 2/9$, while for $\beta < 2/9$ it never vanishes. This is illustrated in figs. 9.2 and 9.3, which show the dependence of $l\tilde{\kappa} + m\tilde{\Omega}$ on $\tilde{U}$. Curves are shown for 21 values of β. The line $l\tilde{\kappa} + m\tilde{\Omega} = 0$ is also drawn. In fig. 9.2, $m = 3$ and $l = -2$; in fig. 9.3, $m = 4$ and $l = -3$.

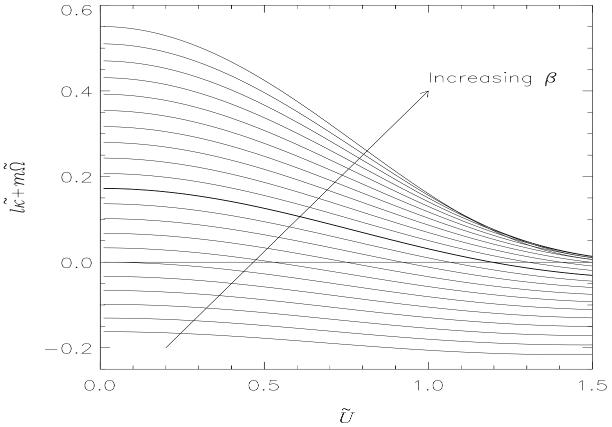


Figure 9.2: The dependence of $l\tilde{\kappa} + m\tilde{\Omega}$ on $\tilde{U}$ and β, for $m = 3$, $l = -2$. Curves of $l\tilde{\kappa} + m\tilde{\Omega}$ against $\tilde{U}$ are shown for 21 values of β, from $\beta = -0.5$ to $\beta = 0.5$ in steps of 0.05. $\beta = 0.0$ is drawn with a bolder line. The line $l\tilde{\kappa} + m\tilde{\Omega} = 0$ is also shown.

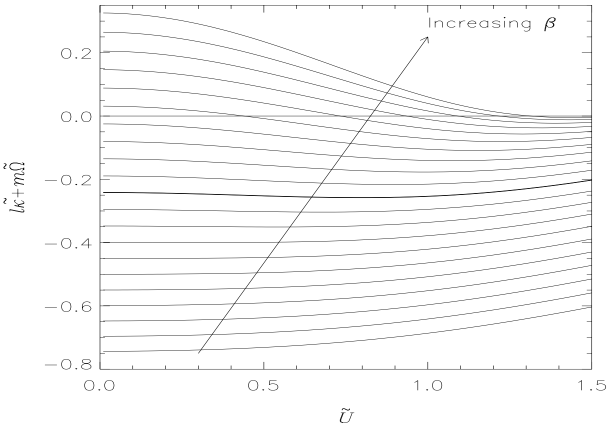
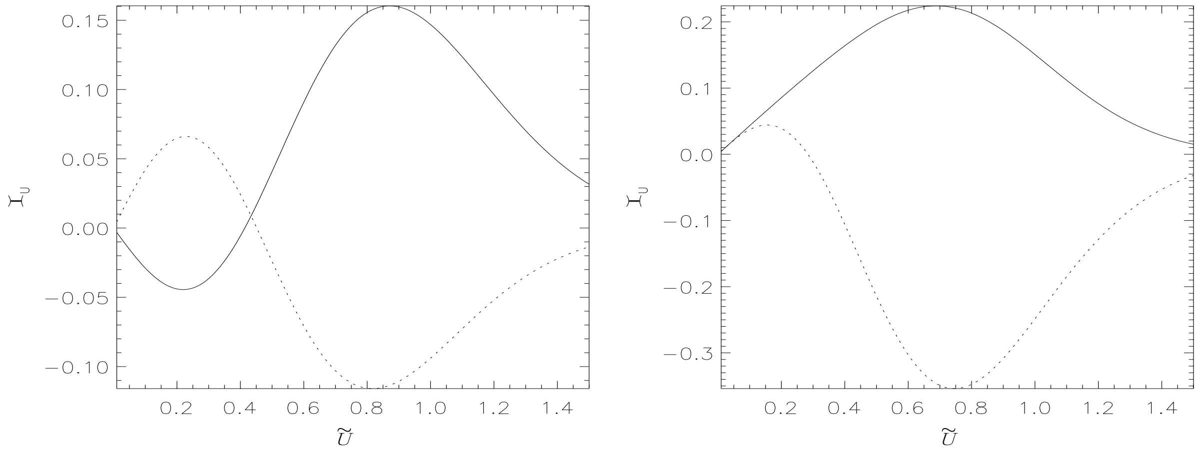


Figure 9.3: The dependence of $l\tilde{\kappa} + m\tilde{\Omega}$ on $\tilde{U}$ and β, for $m = 4$, $l = -3$. Curves of $l\tilde{\kappa} + m\tilde{\Omega}$ against $\tilde{U}$ are shown for 21 values of β, from $\beta = -0.5$ to $\beta = 0.5$ in steps of 0.05. $\beta = 0.0$ is drawn with a bolder line. The line $l\tilde{\kappa} + m\tilde{\Omega} = 0$ is also shown.

Can we still evaluate the mathematical eigenvalue accurately, even when $l\tilde{\kappa}+m\tilde{\Omega}$ vanishes? Fig. 9.4 shows the integrand $\mathcal{I}_U$ of the transfer function $\mathcal{S}_m(\alpha, \alpha')$ (see section 5.3) plotted against eccentric velocity, for two different pairs of wavenumbers. This is for $\beta = -0.2$, for which $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes at $\tilde{U} = 0.527$. There is no discontinuity at this eccentric velocity. This suggests that our integration technique is likely to succeed even when $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes for one of the radial harmonics.



α = −1.00, α' = −1.00

α = −1.00, α' = −2.00

Figure 9.4: The integrand $\mathcal{I}_U$ of the transfer function $\mathcal{S}_m(\alpha, \alpha')$ plotted against eccentric velocity for $m = 3$, $\beta = -0.2$. The left-hand plot shows $\mathcal{I}_U$ for $\alpha = \alpha' = -1$; the right-hand plot shows $\mathcal{I}_U$ for $\alpha = \alpha' = -2$. The integrand is not evaluated at the eccentric velocity at which $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes, but is evaluated at points close on either side. [$\beta = -0.2$, $N = 2$; $m = 3$, $\Omega_{\rm p} = 0.50$, $s = 0.30$, $Q_{\rm s} = 1.0$]

As explained in Chapter 5, we perform the integration over eccentric velocity using Gauss-Laguerre quadrature. The transfer function is evaluated at $n_{\rm GL}$ discrete values of the eccentric velocity $\tilde{U}$. We now investigate the effect on the mathematical eigenvalue of changing the number and position of these evaluation points. The following tables show results for three different values of β, for which $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes at different eccentric velocities. In each case, we consider an $m = 3$ perturbation with pattern speed $\Omega_{\rm p} = 0.7$ and growth rate $s = 0.1$ in an inner cut-out disk with $N = 2$ and $Q_{\rm s} = 1.0$. We use numerical accuracy parameters $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$. To begin with, we consider a disk with $\beta = 0.25$.

$\beta = 0.25$: $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes at $\tilde{U} = 1.635$.			
Gauss-Laguerre integration:			Mathematical eigenvalue
$n_{\rm GL}$	f_σ	Eccentric velocities $\tilde{U}$	
4	0.8	0.182, 0.423, 0.682, 0.981	$0.257303 + 0.036400i$
6	0.8	0.151, 0.349, 0.554,0.769, 1.00, 1.28	$0.257042 + 0.036180i$
8	0.8	0.132, 0.304, 0.480, 0.661, 0.850, 1.05,1.27,1.53	$0.257033 + 0.036164i$
10	0.8	0.119, 0.273, 0.430, 0.590, 0.754, 0.924, 1.10, 1.29, 1.50, 1.75	$0.257030 + 0.036162i$

When $n_{\rm GL} \leq 8$, the transfer function is evaluated at eccentric velocities below that at which $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes. When $n_{\rm GL} = 10$, the uppermost point falls above the critical value. No change in the mathematical eigenvalue is observed to within 5 s.f. This suggests that the algorithm is successfully interpolating across the troublesome eccentric velocity. For lower values of β, the results are similar:

$\beta = 0.00$: $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes at $\tilde{U} = 1.192$.

		Gauss-Laguerre integration:	
n_{GL}	f_σ	Eccentric velocities $\tilde{U}$	Mathematical eigenvalue
4	0.8	0.243, 0.565, 0.911, 1.31	$0.270107 + 0.036339i$
6	0.8	0.202, 0.466, 0.740, 1.03, 1.34, 1.71	$0.269823 + 0.036217i$
8	0.8	0.176, 0.406, 0.641, 0.883, 1.135, 1.402, 1.696, 2.044	$0.269799 + 0.036217i$
10	0.8	0.159, 0.365, 0.575, 0.789, 1.01, 1.23, 1.47, 1.73, 2.01, 2.34	$0.269801 + 0.036221i$

Similar mathematical eigenvalues are returned no matter how many points fall on either side of the eccentric velocity for which $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes. This is convincing evidence that the vanishing is not affecting the integration. However, the integrand is usually largest at small values of the eccentric velocity. For the two values of β examined so far, $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes at a relatively high value of $\tilde{U}$. Thus it is possible that the good convergence observed so far is because the integrand is disturbed only where it is already very small. We therefore next test two values of β where $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes at a low values of $\tilde{U}$.

$\beta = -0.20$: $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes at $\tilde{U} = 0.527$.

		Gauss-Laguerre integration:	
n_{GL}	f_σ	Eccentric velocities $\tilde{U}$	Mathematical eigenvalue
4	0.8	0.327, 0.761, 1.23, 1.77	$0.292522 - 0.028130i$
6	0.8	0.272, 0.628, 0.996, 1.38, 1.81, 2.30	$0.292532 - 0.028178i$
8	0.8	0.238, 0.547, 0.864, 1.19, 1.53, 1.89, 2.28, 2.75	$0.292529 - 0.028172i$
10	0.8	0.214, 0.492, 0.774, 1.06, 1.36, 1.66, 1.98, 2.32, 2.70, 3.149	$0.292530 - 0.028171i$

$\beta = -0.25$: $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes at $\tilde{U} = 0.000$.

		Gauss-Laguerre integration:	
n_{GL}	f_σ	Eccentric velocities $\tilde{U}$	Mathematical eigenvalue
4	0.8	0.327, 0.761, 1.23, 1.77	$0.304394 - 0.032953i$
6	0.8	0.272, 0.628, 0.996, 1.38, 1.81, 2.30	$0.304469 - 0.032995i$
8	0.8	0.238, 0.547, 0.864, 1.19, 1.53, 1.89, 2.28, 2.75	$0.304473 - 0.032991i$
10	0.8	0.214, 0.492, 0.774, 1.06, 1.36, 1.66, 1.98, 2.32, 2.70, 3.149	$0.304474 - 0.032992i$

The convergence is as good as ever. We conclude that the mathematical eigenvalue is still being evaluated successfully, despite the vanishing of $l\tilde{\kappa} + m\tilde{\Omega}$. Thus reassured, we proceed to examine the behaviour of the mathematical eigenvalue for $m = 3$ and $m = 4$.

9.3 Triskele modes

Global stability of the cut-out disks

Figs. 9.5 - 9.7 show the dependence of the largest eigenvalue on growth rate and pattern speed, for β = 0.25, 0.00 and -0.25. The left-hand plot of each pair shows the results for $Q_{\mathrm{s}} = 1.0$. Since the power-law disks are very stable to $m = 3$ perturbations, the results for a considerably cooler disk, with $Q_{\mathrm{s}} = 0.5$, are shown in the right-hand plot of each pair. As usual in these plots, the diamond symbols mark eigenvalues actually calculated. The linking curves were drawn using a spline interpolation procedure.

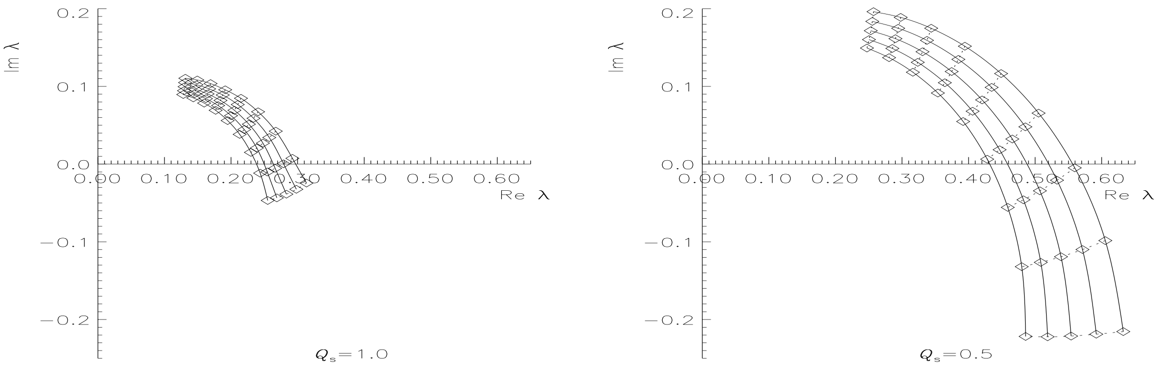


Figure 9.5: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = 0.25$. The left-hand plot is for $Q_{\mathrm{s}} = 1.0$, i.e. $\tilde{\sigma}_u = 0.283$ and $\gamma = 11.0$. The right-hand plot is for $Q_{\mathrm{s}} = 0.5$, i.e. $\tilde{\sigma}_u = 0.141$, $\gamma = 48.5$. The solid lines show curves of constant s at intervals of 0.1 from $s = 0^+$ (right-most curve) to $s = 0.4$ (left-most curve). The dotted lines show curves of constant Ω_{p} at intervals of 0.1 from $\Omega_{\mathrm{p}} = 0.5$ (lowest curve) to $\Omega_{\mathrm{p}} = 1.3$ (highest curve). [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\min} = -30$, $l_{\max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 1.5$.]

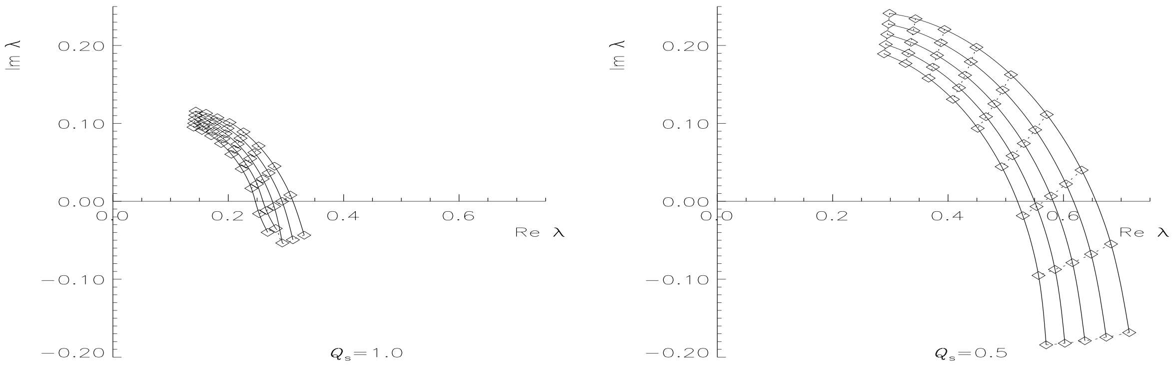


Figure 9.6: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = 0.00$. The left-hand plot is for $Q_{\mathrm{s}} = 1.0$, i.e. $\tilde{\sigma}_u = 0.378$ and $\gamma = 6.00$. The right-hand plot is for $Q_{\mathrm{s}} = 0.5$, i.e. $\tilde{\sigma}_u = 0.189$ and $\gamma = 27.0$. The solid lines show curves of constant s at intervals of 0.1 from $s = 0^+$ (right-most curve) to $s = 0.4$ (left-most curve). The dotted lines show curves of constant Ω_{p} at intervals of 0.1 from $\Omega_{\mathrm{p}} = 0.5$ (lowest curve) to $\Omega_{\mathrm{p}} = 1.3$ (highest curve). [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\min} = -30$, $l_{\max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 1.5$.]

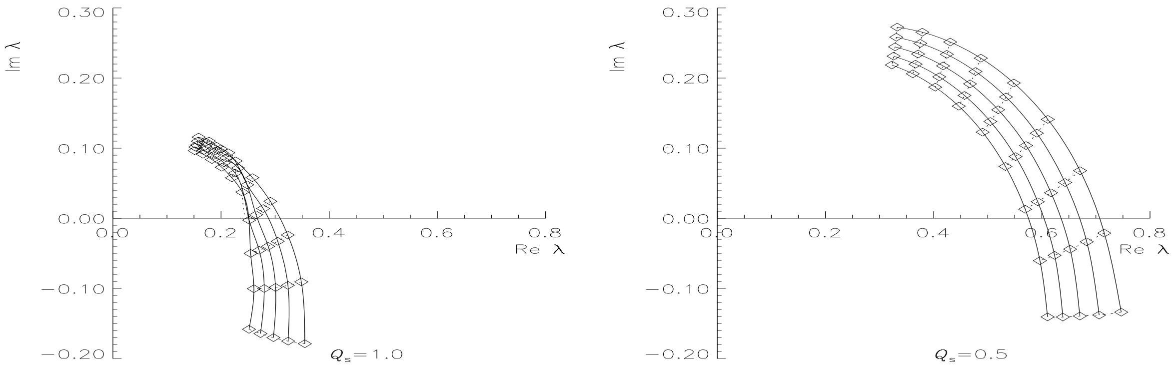


Figure 9.7: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = -0.25$. The left-hand plot is for $Q_{\mathrm{s}} = 1.0$, i.e. $\tilde{\sigma}_u = 0.509$ and $\gamma = 3.36$. The right-hand plot is for $Q_{\mathrm{s}} = 0.5$, i.e. $\tilde{\sigma}_u = 0.254$ and $\gamma = 14.9$. The solid lines show curves of constant s at intervals of 0.1 from $s = 0^+$ (right-most curve) to $s = 0.4$ (left-most curve). The dotted lines show curves of constant Ω_{p} at intervals of 0.1 from $\Omega_{\mathrm{p}} = 0.5$ (lowest curve) to $\Omega_{\mathrm{p}} = 1.3$ (highest curve). [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\min} = -30$, $l_{\max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 1.5$.]

These plots suggest that disks with β in the range of interest (principally $|\beta| < 0.5$) are very stable to $m = 3$ perturbations. Even when the temperature is as low as $Q_{\rm s} = 0.5$, the eigenvalue curves intersect the real axis at values considerably less than unity. As for $m = 2$, we find that disks with lower β are less securely stable, in that the eigenvalue curves cross the real axis at higher values. We have found that for disks with $N = 2$, $Q_{\rm s}$ has to be lower than 0.4 before even the $\beta = -0.5$ disk becomes unstable to $m = 3$ perturbations. For disks with falling rotation curves, still lower temperatures are necessary.

Before we present graphs showing these findings, a word of caution. At these very low temperatures, the convergence is poor. This is mainly because, as we saw in section 7.3, modes are more tightly wound at low temperatures. Thus the density transform has a peak at a high wavenumber. The density transforms also tend to be highly oscillatory (see fig. 9.9 below). Thus very large, very finely-meshed grids are required, which are numerically impractical. We cannot claim to locate the eigenvalue to better than 2 s.f. As an example, the following tables investigate the convergence of the largest mathematical eigenvalue for a perturbation with $m = 3$, $\Omega_{\rm p} = 0.668$ and $s = 10^{-7}$, in a disk with $\beta = -0.5$, $N = 2$, $Q_{\rm s} = 0.354$. We set $l_{\rm min} = -20$, $l_{\rm max} = +30$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$ and $b_{\rm acc} = 2.0$, and investigate the effect of changing n, $\Delta\alpha$ and $n_{\rm GL}$.

To begin with, we choose a grid covering a range of 25 in wavenumber.

$n_{\rm GL}$	Grid-spacing $\Delta\alpha$	Number of grid points n	Largest mathematical eigenvalue
7	1.042	25	$0.969493 - 0.0366762i$
	0.500	51	$0.961756 - 0.0440979i$
	0.403	63	$0.960370 - 0.0452787i$
	0.298	85	$0.959657 - 0.0458892i$
	0.202	125	$0.958707 - 0.0466800i$
	0.100	251	$0.956902 - 0.0481349i$
9	1.042	25	$0.969493 - 0.0366763i$
	0.500	51	$0.961756 - 0.0440979i$
	0.403	63	$0.960370 - 0.0452787i$
	0.298	85	$0.959657 - 0.0458891i$
	0.202	125	$0.958707 - 0.0466799i$
	0.100	251	$0.956902 - 0.0481349i$

The results are insensitive to the number of Gauss-Laguerre points $n_{\rm GL}$. The size and spacing of the grid is much more important. We now increase the range of the grid to 150 in α.

$n_{\rm GL}$	Grid-spacing $\Delta\alpha$	Number of grid points n	Largest mathematical eigenvalue
9	1.000	151	$1.01719 + 0.0551624i$
	0.500	301	$1.01223 + 0.0539355i$
	0.401	375	$1.00930 + 0.0539236i$
	0.300	501	$1.00737 + 0.0469115i$

and finally to 200.

$n_{\rm GL}$	Grid-spacing $\Delta\alpha$	Number of grid points n	Largest mathematical eigenvalue
9	1.000	201	$1.01054 + 0.0566810i$
	0.500	401	$1.00671 + 0.0497124i$
	0.400	501	$1.01320 + 0.0599816i$

Even with these high accuracy parameters, the largest mathematical eigenvalue has converged to only 2 or 3 s.f. In general, the convergence was worse for colder disks and for higher β. The $N = 2$ disk is extremely stable to $m = 3$ perturbations; marginal modes thus require very low temperatures, so the convergence is extremely poor. The $N = 2$ results presented below are probably not trustworthy to 2 s.f., but are included since they give at least an indication of the very low temperatures required. For $N = 3$ and $N = 4$, the critical temperature is higher and hence the convergence is better. For $N = 3$, we ran the mode-finding programs with two different grids and ranges in l ($n = 501$, $\Delta\alpha = 0.3$, $l_{\min} = -20$, $l_{\max} = 30$ and $n = 601$, $\Delta\alpha = 0.2$, $l_{\min} = -30$, $l_{\max} = 40$). The values of Q_{s} obtained agreed to 4 s.f. The pattern speed could be obtained less reliably, but was around 0.6 throughout. Fig. 9.8 presents this information in terms of the velocity dispersion $\tilde{\sigma}_u$ and anisotropy parameter γ.

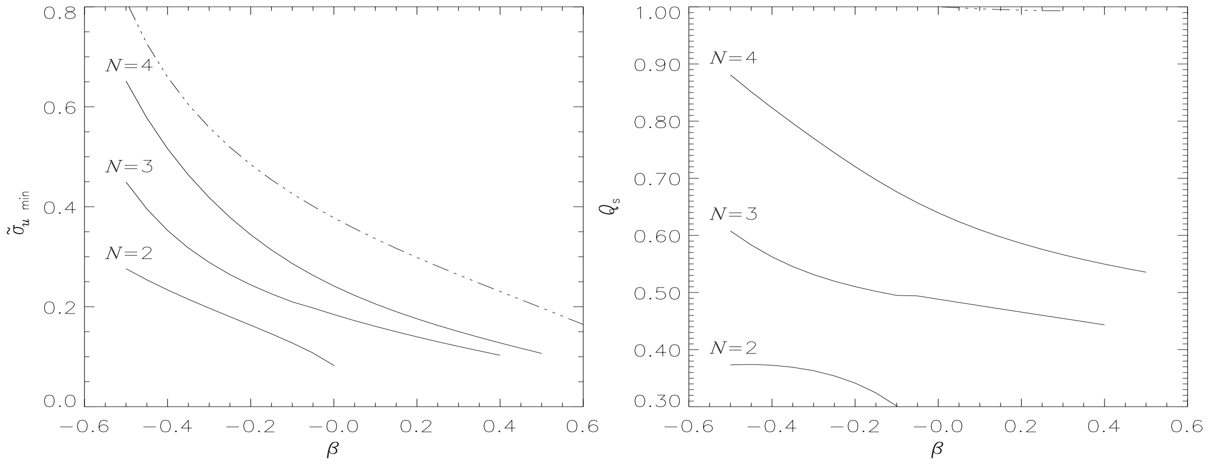


Figure 9.8: Minimum temperature for triskele stability plotted against β, for disks with various cut-out functions. The left-hand plot shows the minimum velocity dispersion $\tilde{\sigma}_u$; the right-hand one the same data presented in terms of the stability parameter Q_{s}. The solid lines indicate the results for inner cut-out disks with $N = 1, 2, 3, 4$. The broken line shows the temperature necessary for global axisymmetric stability in the self-consistent disk. [Numerical accuracy parameters: $n = 601$, $\Delta\alpha = 0.2$, $l_{\min} = -20$, $l_{\max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 2.0$, $\epsilon_\lambda = 10^{-4}$.]

The rotation curve and inner cut-out affect the stability to triskele perturbations in the same way as they affect the $m = 2$ stability. Disks where the centre has been cut out more sharply are more unstable than those where it has been removed relatively gently. Disks with rising rotation curves are more unstable than those with falling rotation curves, both in terms of $\tilde{\sigma}_{u,\min}$, the absolute amount of velocity dispersion required, and in terms of Q_{s}, the ratio of the velocity dispersion needed relative to that necessary for axisymmetric stability.

Fig. 9.9 shows the density transforms of the marginal modes in inner cut-out disks with $N = 3$ and $\beta = \pm 0.25$. Once again, the transforms are strongly peaked at a dominant wavenumber α_{u}, but there is a considerable contribution from wavenumbers more trailing than α_{u}.

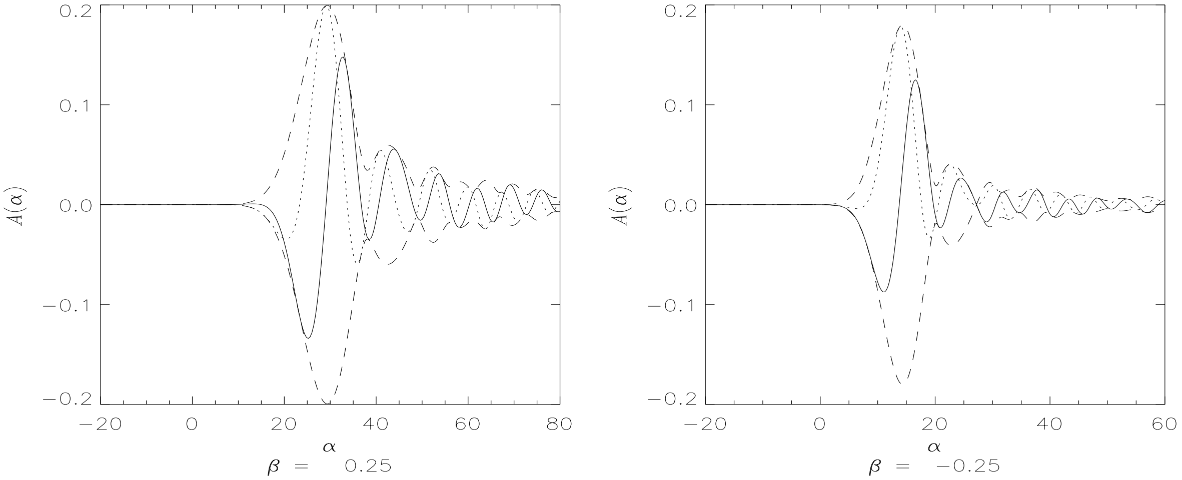


Figure 9.9: Density transform $A(\alpha)$ of the marginal modes in inner cut-out disks with $N = 3$, $\beta = \pm 0.25$. The solid line is $\mathrm{Re}(A(\alpha))$, the dotted is $\mathrm{Im}(A(\alpha))$ and the dashed line shows the envelope $\pm|A(\alpha)|$. The left-hand plot is for $\beta = 0.25$, for which the marginally stable mode has $Q_\mathrm{s} = 0.460$, $\gamma = 57.5$, $\tilde{\sigma}_u = 0.130$, $\Omega_\mathrm{p} = 0.635$, $\alpha_\mathrm{u} \approx 29.2$. The right-hand plot is for $\beta = -0.25$, for which the marginally stable mode has $Q_\mathrm{s} = 0.520$, $\gamma = 13.8$, $\tilde{\sigma}_u = 0.265$, $\Omega_\mathrm{p} = 0.636$, $\alpha_\mathrm{u} \approx 14.4$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.3$, $l_\mathrm{min} = -20$, $l_\mathrm{max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 2$.]

Figs. 9.10 and 9.11 show the density contour plots of the marginal modes in inner cut-out disks with $\beta = \pm 0.25$.

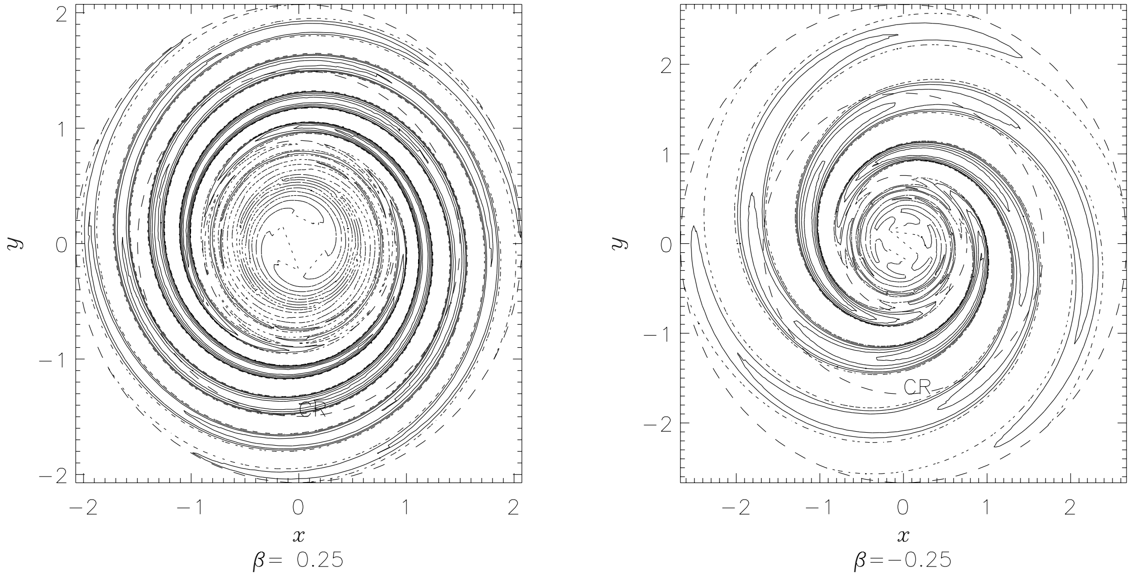
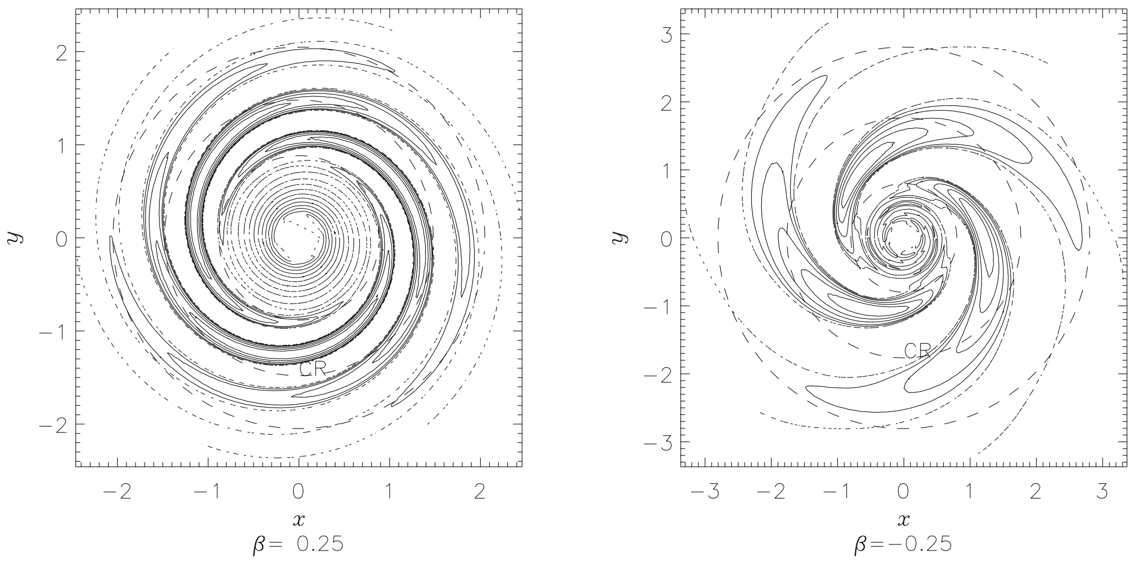


Figure 9.10: Density contour plot of a marginally stable mode in inner cut-out disks with $N = 3$. The left-hand plot is for $\beta = 0.25$, for which the marginally stable mode has $Q_\mathrm{s} = 0.460$, $\gamma = 57.5$, $\tilde{\sigma}_u = 0.130$, $\Omega_\mathrm{p} = 0.635$, $\alpha_\mathrm{u} \approx 29.2$. The right-hand plot is for $\beta = -0.25$, for which the marginally stable mode has $Q_\mathrm{s} = 0.520$, $\gamma = 13.8$, $\tilde{\sigma}_u = 0.265$, $\Omega_\mathrm{p} = 0.636$, $\alpha_\mathrm{u} \approx 14.4$. In each plot, the density is calculated between $R = 0.3 R_\mathrm{ILR}$ and $R = R_\mathrm{OLR}$. The solid lines mark 10%, 20%, 40%, 60% and 80% of the maximum density within this range; the dotted lines show the nodes, where the density perturbation is zero. The dashed circles indicate the Lindblad resonances and the co-rotation radius. For $\beta = 0.25$, $R_\mathrm{ILR} = 0.893$, $R_\mathrm{CR} = 1.50$, $R_\mathrm{OLR} = 2.07$; for $\beta = -0.25$, $R_\mathrm{ILR} = 0.760$, $R_\mathrm{CR} = 1.68$, $R_\mathrm{OLR} = 2.67$, in units where $R_0 = 1$. [Numerical accuracy parameters: for $\beta = 0.25$, $n = 401$, $\Delta\alpha = 0.4$; for $\beta = 0.25$, $n = 601$, $\Delta\alpha = 0.2$; $l_\mathrm{min} = -20$, $l_\mathrm{max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 20$, $b_\mathrm{acc} = 2$.]



$\beta= 0.25$ $\beta=-0.25$

Figure 9.11: Density contour plot of a marginally stable mode in inner cut-out disks with $N = 4$. The left-hand plot is for $\beta = 0.25$, for which the marginally stable mode has $Q_{\rm s} = 0.576$, $\gamma = 36.2$, $\tilde{\sigma}_u = 0.163$, $\Omega_{\rm p} = 0.643$, $\alpha_{\rm u} \approx 19.8$. The right-hand plot is for $\beta = -0.25$, for which the marginally stable mode has $Q_{\rm s} = 0.745$, $\gamma = 6.46$, $\tilde{\sigma}_u = 0.379$, $\Omega_{\rm p} = 0.608$, $\alpha_{\rm u} \approx 7.6$. In each plot, the density is calculated between $R = 0.3 R_{\rm ILR}$ and $R = 1.2 R_{\rm OLR}$. The solid lines mark 10%, 20%, 40%, 60% and 80% of the maximum densitywithin this range; the dotted lines show the nodes, where the density perturbation is zero. The dashed circles indicate the Lindblad resonances and the co-rotation radius. For $\beta = 0.25$, $R_{\rm ILR} = 0.883$, $R_{\rm CR} = 1.48$, $R_{\rm OLR} = 2.05$; for $\beta = -0.25$, $R_{\rm ILR} = 0.800$, $R_{\rm CR} = 1.77$, $R_{\rm OLR} = 2.81$, in units where $R_0 = 1$. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.25$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

As for $m = 2$ and $m = 1$, the modes start at the inner Lindblad radius. Once again the pattern speed is such as to place the inner Lindblad radius safely inside the inner cut-out. They extend significantly further than the lower-order modes, reaching out to the outer Lindblad radius even though these disks are fairly warm (in general, modes extend less far in hotter disks).

Neutral modes

So far, it appears as if the stability of the power-law disks to $m = 3$ perturbations is qualitatively exactly like that to $m = 2$ perturbations, the only difference being that even lower temperatures are required before growing modes can be excited. We were initially led to this conclusion by the eigenvalue curves plotted in figs. 9.5 - 9.7. The very low values at which the eigenvalue curves crossed the real axis suggested that the disks are very stable to $m = 3$ perturbations. However, a very different picture emerges when we pursue the mathematical eigenvalues down to vanishingly low growth rates and pattern speeds. Figs. 9.12 - 9.14 show the dominant eigenvalue curves as s and $\Omega_{\rm p}$ tend to zero, for $Q_{\rm s} = 1$ and $\beta = 0.25$, 0.00 and -0.25.

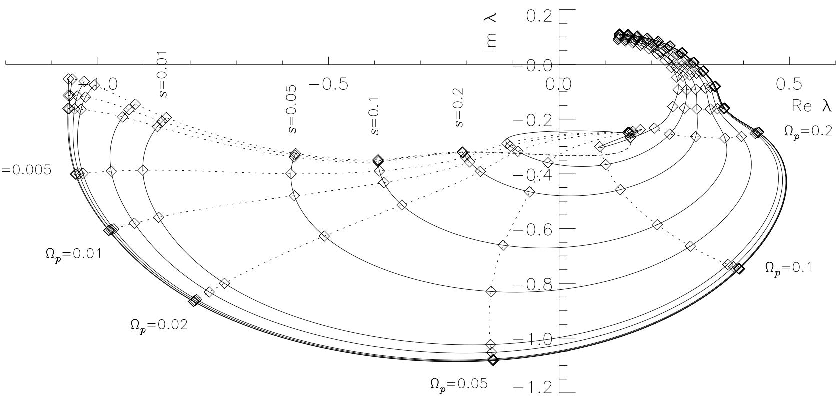


Figure 9.12: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = 0.25$. This is for $Q_{\rm s} = 1$, i.e. $\tilde{\sigma}_u = 0.283$ and $\gamma = 11.0$. The solid lines show curves of constant s at 11 values of s: $s = 10^{-7}$, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4. The dotted lines show curves of constant $\Omega_{\rm p}$ at 20 values of $\Omega_{\rm p}$: $\Omega_{\rm p} = 0.0001$, 0.0005, 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3 … 1.2, 1.3. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$.]

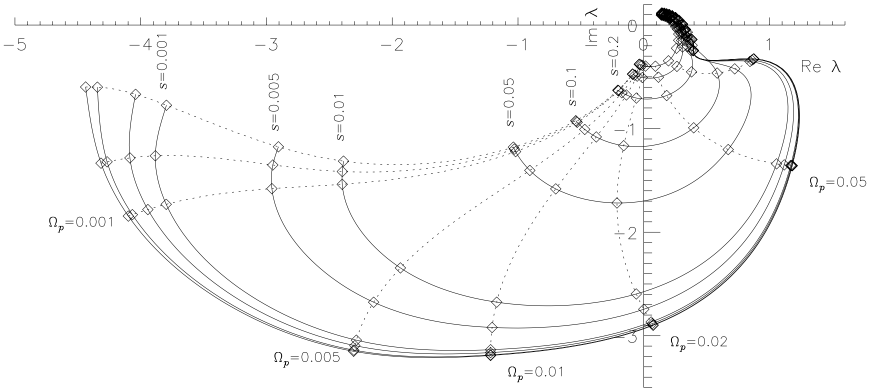


Figure 9.13: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = 0.00$. This is for $Q_{\rm s} = 1$, i.e. $\tilde{\sigma}_u = 0.378$ and $\gamma = 6.00$. The solid lines show curves of constant s at 11 values of s: $s = 10^{-7}$, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4. The dotted lines show curves of constant $\Omega_{\rm p}$ at 20 values of $\Omega_{\rm p}$: $\Omega_{\rm p} = 0.0001$, 0.0005, 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3 … 1.2, 1.3. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$.]

For $\beta = 0.25$ and $\beta = 0.00$, the behaviour is similar to that observed for $m = 2$ perturbations (e.g. fig. 7.16). As the pattern speed is decreased, the eigenvalue curves continue to arc clockwise around the origin, tending to a a constant value (for given growth rate) in the limit $\Omega_{\rm p} \to 0$. They appear to be tending towards the negative real axis in the twin limits $s \to 0$, $\Omega_{\rm p} \to 0$. However, for $\beta = -0.25$ the behaviour is strikingly different.

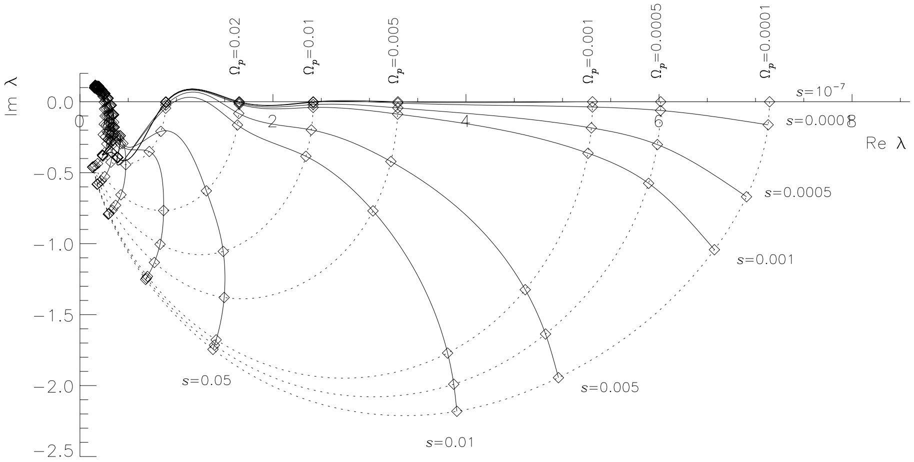


Figure 9.14: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = -0.25$. This is for $Q_{\rm s} = 1$, i.e. $\tilde{\sigma}_u = 0.509$ and $\gamma = 3.36$. The solid lines show curves of constant s at 11 values of s: $s = 10^{-7}$, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05, 0.1, 0.2, 0.3, 0.4. The dotted lines show curves of constant $\Omega_{\rm p}$ at 20 values of $\Omega_{\rm p}$: $\Omega_{\rm p}$ = 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3 ... 1.2, 1.3. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$.]

Fig. 9.14 is much closer to the $m = 1$ eigenvalue curves, fig. 8.1. It indicates that modes with vanishing pattern speed and growth rate are possible even at the high temperature of $Q_{\rm s} = 1.0$. It appears that there is a dramatic change in the behaviour of the eigenvalue curves between $\beta = 0.00$ and $\beta = -0.25$. However, before drawing this conclusion we should investigate further. First of all, we have established that the behaviour is not a numerical artefact due to the very low growth rate and pattern speed. We tested the convergence of the mathematical eigenvalue with different numbers of Gauss-Laguerre points, and confirmed that the Gaussian quadrature copes well with the vanishing of $l\tilde{\kappa} + m\tilde{\Omega}$ even in the limit of vanishing $\tilde{\omega}$. A more subtle explanation of the apparent change in behaviour might be that the underlying eigenvalue spectra for the three β are in fact similar, with eigenvalues reaching out along the positive and negative real axes as s and $\Omega_{\rm p}$ tend to zero. For $\beta = 0.25$, the largest eigenvalue would happen to be in the branch along the negative real axis, whereas for $\beta = -0.25$, the largest eigenvalue would happen to lie in the branch reaching along the positive real axis. Thus the very different answers obtained for the largest eigenvalue would mask a similar underlying eigenvalue spectrum. Figs. 9.15 and 9.16 compare the complete eigenvalue spectra of disks with $\beta = \pm 0.25$ as the limits $s \to 0$, $\Omega_{\rm p} \to 0$ are approached. We see that in fact the spectra are radically different throughout this range of s and $\Omega_{\rm p}$ (note the differing scales). As s and $\Omega_{\rm p}$ both approach zero, the $\beta = -0.25$ spectrum develops a long "spur" out along the positive real axis. The $\beta = 0.25$ spectrum has no such real spur. It shows a spiral/lambda pattern similar to that observed for $m = 2$ eigenvalues (fig. 7.15).

To track down the β at which this change of behaviour occurs, we calculated the marginal eigenvalue curves for several β. We saw the form of these curves for bisymmetric disturbances in fig. 7.17. For $\beta \geq 0.2$, the behaviour is similar for $m = 3$ perturbations. The marginal eigenvalue curves for $\beta \geq 0.2$ are drawn in the left-hand plot of fig. 9.17. The eigenvalues arc clockwise around the origin, ending up close to the negative real axis. The right-hand plot shows the marginal eigenvalue curves for $\beta \leq -0.25$. Here the behaviour is strikingly different. As in fig. 9.14, the eigenvalue curves begin by moving clockwise around the origin, but around $\Omega_{\rm p} \approx 0.1$ they change direction. As $\Omega_{\rm p} \to 0$, they end up on the positive real axis.

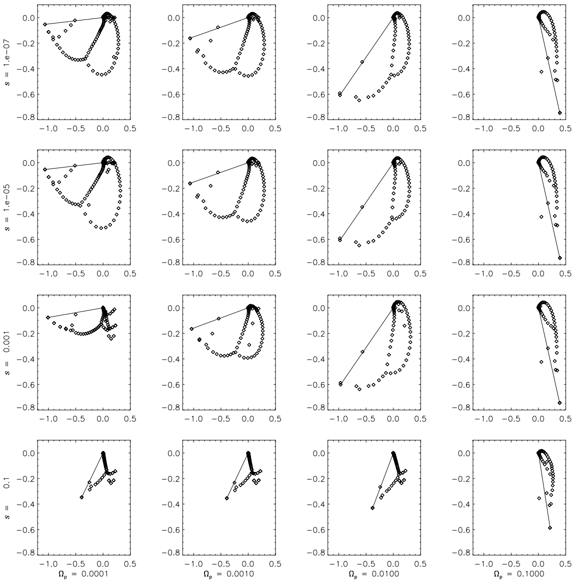


Figure 9.15: Complete eigenvalue spectrum for perturbations in a disk with $\beta = 0.25$, $N = 2$, $Q_{\rm s} = 1$. The four rows show spectra for increasing growth rate: $s = 10^{-7}$, 10^{-5}, 10^{-3}, 10^{-1}. The four columns show spectra for increasing pattern speed: $\Omega_{\rm p} = 10^{-4}$, 10^{-3}, 10^{-2}, 10^{-1}. In each case, a line is drawn from the origin to the largest eigenvalue. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2.0$.]

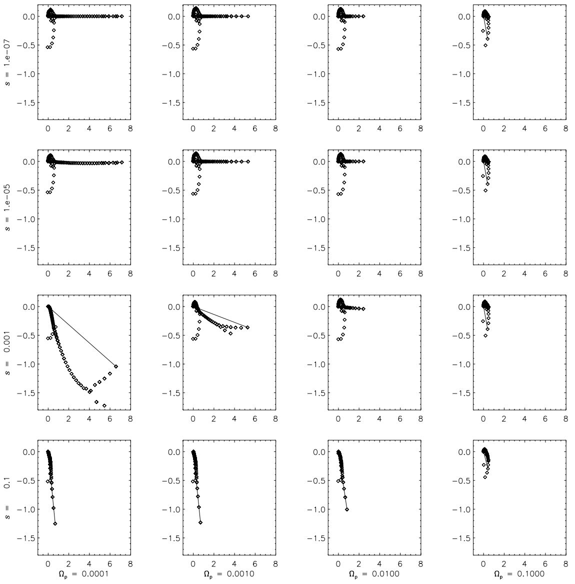


Figure 9.16: Complete eigenvalue spectrum for perturbations in a disk with $\beta = -0.25$, $N = 2$, $Q_{\rm s} = 1$. The four rows show spectra for increasing growth rate: $s = 10^{-7}$, 10^{-5}, 10^{-3}, 10^{-1}. The four columns show spectra for increasing pattern speed: $\Omega_{\rm p} = 10^{-4}$, 10^{-3}, 10^{-2}, 10^{-1}. In each case, a line is drawn from the origin to the largest eigenvalue. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2.0$.]

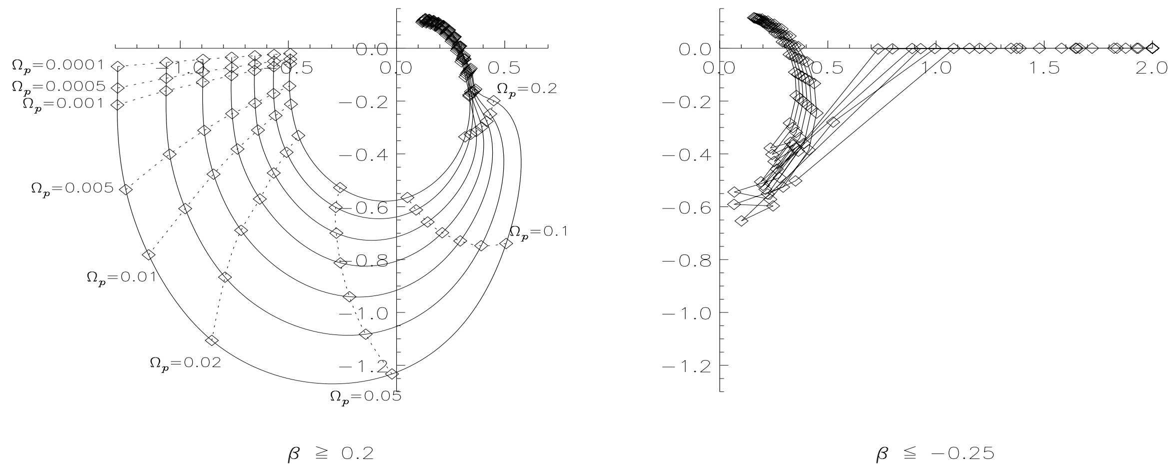


Figure 9.17: Marginal eigenvalue curves for inner cut-out disks with $N = 2$, $Q_{\rm s} = 1.0$ and various values of β. Each plot shows the largest mathematical eigenvalues for vanishing growth rate, $s = 10^{-7}$, and 20 values of the pattern speed: $\Omega_{\rm p}$ = 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, … 1.2, 1.3. The left-hand plot shows these marginal eigenvalue curves for 7 positive values of β: β = 0.2, 0.25, 0.3 … 0.45, 0.5. The right-hand plot shows the curves for 6 negative values of β: β = −0.25, -0.3, -0.35 … -0.45, -0.5. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2.$]

For intermediate values of β, $-0.25 < \beta < 0.2$, the marginal eigenvalue curves are tangled, and are not shown here. Their behaviour is closer to the curves with $\beta \geq 0.2$ than to those with $\beta \leq -0.25$. It proved numerically easier to investigate the transition at lower temperatures. Fig. 9.18 shows marginal eigenvalue curves for β between -0.50 and +0.25, for $Q_{\rm s} = 0.3$. Here, we have plotted the data on three sets of axes, due to the very different scales required. The eigenvalue curves are once again rather tangled, presumably due to swapping between eigenvalues as different eigenvalues attain dominance. However, the transition between the two different stability regimes appears to occur close to $\beta = -0.25$ – the value of β distinguishing disks where $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes from those where it does not. For $\beta \leq -0.25$, the eigenvalue ends up on the positive real axis as $s \to 0$ and $\Omega_{\rm p} \to 0$, indicating the presence of neutral modes. For this low temperature, very large positive eigenvalues are obtained. (For example, for $\beta = -0.25$, $s = 10^{-7}$ and $\Omega_{\rm p} = 10^{-4}$, $\lambda_1 \approx 180$. When $Q_{\rm s} = 1$, $\lambda_1 \approx 7$.) When β is increased above -0.25, the behaviour of the eigenvalue in the limit $s \to 0$ and $\Omega_{\rm p} \to 0$ suddenly changes. The eigenvalue arcs clockwise around the origin. For the first few values of β above -0.25, the modulus of the eigenvalue is very large. For higher values of β (> 0.1), the modulus of the eigenvalue is smaller. Curves of $\beta > 0.25$ are not shown here, since they are very tangled. However, their behaviour is qualitatively similar. In all cases with $\beta > -0.25$, the eigenvalue seems destined to end up on the negative real axis as $s \to 0$ and $\Omega_{\rm p} \to 0$.

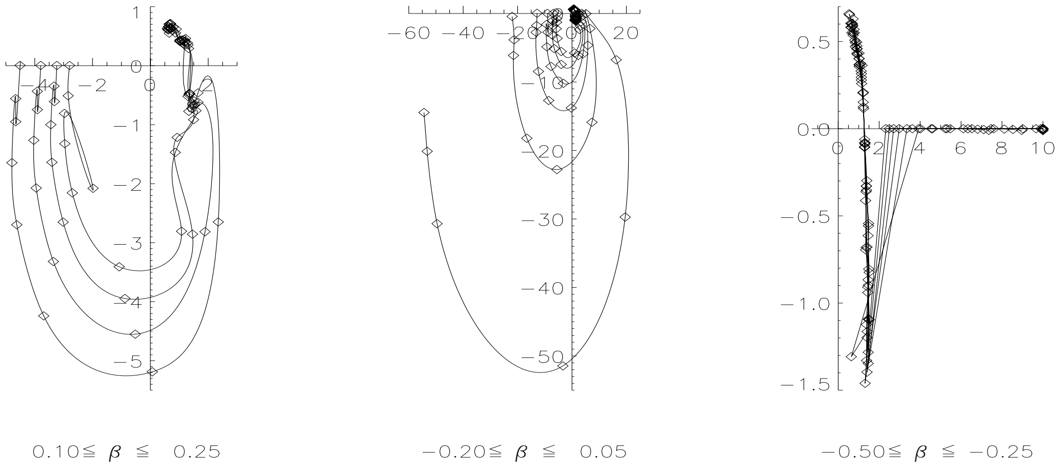


Figure 9.18: Marginal eigenvalue curves for inner cut-out disks with $N = 2$, $Q_{\rm s} = 0.3$ and various values of β. Each plot shows the largest mathematical eigenvalues for vanishing growth rate, $s = 10^{-7}$, and 20 values of the pattern speed: $\Omega_{\rm p} = 0.0001$, 0.0005, 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, ... 1.2, 1.3. The left-hand plot shows these marginal eigenvalue curves for $\beta = -0.50$, -0.45, -0.4, -0.35, -0.30 and -0.25. The middle plot shows curves for $\beta = -0.20$, -0.15, -0.10, -0.05, 0.00 and 0.05. The right-hand plot shows the curves for $\beta = 0.10$, 0.15, 0.20 and 0.25. Note the very different scales. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$.]

It appears, then, that disks with different β have very different stability to $m = 3$ perturbations. For $\beta > -0.25$, the eigenvalue curves intersect the real axis only once, at pattern speeds of about $\Omega_{\rm p} \approx 0.6$. Marginal stability occurs at the critical temperature for which the marginal eigenvalue curve intersects the real axis at $\lambda = 1$. This critical temperature is very low (fig. 9.8). These disks are extremely stable to $m = 3$ perturbations. However, the behaviour is qualitatively similar to that for $m = 2$.

For $\beta \leq -0.25$, the eigenvalue curves cross the real axis at moderate pattern speeds of about $\Omega_{\rm p} \approx 0.6$, but then return to it as $\Omega_{\rm p} \to 0$, $s \to 0$. If the dominant eigenvalue is large and real, one of the sub-dominant eigenvalues must have the value $(1, 0)$ (or could be made to do so with small adjustments in s and Ω_{p} at the same temperature). Thus for $\beta \leq -0.25$, the behaviour is qualitatively similar to that for $m = 1$. We cannot define a critical temperature above which the disk is stable. As the temperature is raised, the disk continues to admit growing modes, although their growth rate becomes vanishingly small.

For vanishing pattern speed, the inner Lindblad radius moves to infinity. Under such circumstances, we might expect the response of the cut-out disk to resemble that of the self-consistent disk. And indeed, the two stability regimes we have found for cut-out disks are reflected in the response of the self-consistent disk. In Chapter 4, we derived the response function of the self-consistent disk to neutral modes (with zero growth rate and pattern speed). Fig. 9.19 shows the response function for self-consistent disks with $\beta = \pm 0.25$. Note the different vertical scales in each case.

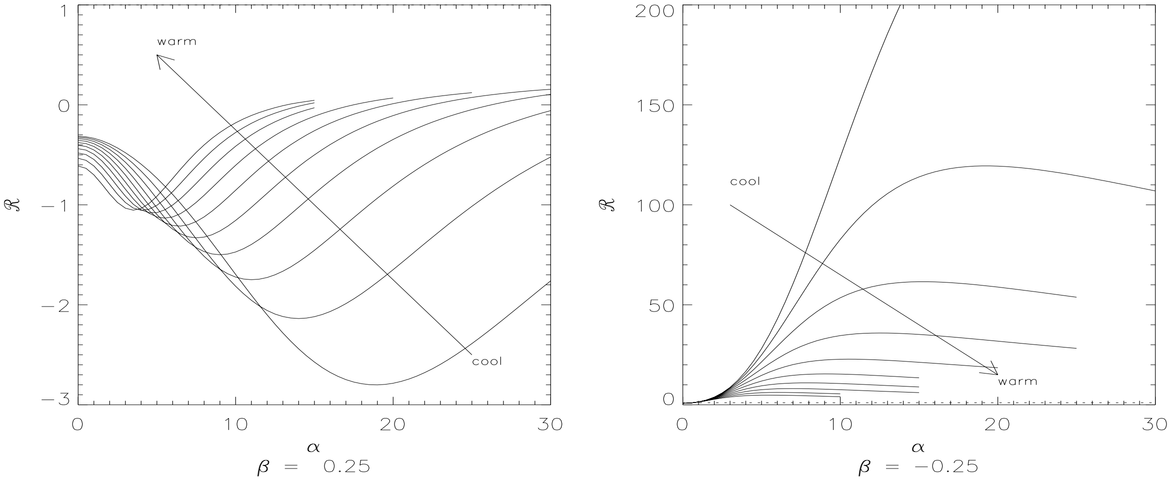


Figure 9.19: The response function for the self-consistent disks with $\beta = \pm 0.25$. For each β, curves are shown for ten values of Q_{s} in steps of 0.1 from 0.3 to 1.2. Note the different vertical scales. [Numerical accuracy parameters: $l_{\mathrm{min}} = -20$, $l_{\mathrm{max}} = +30$, $n_{GL} = 12$, $f_{\sigma} = 0.8$, $a_{\mathrm{acc}} = 50$, $b_{\mathrm{acc}} = 2.5$.]

For $\beta = +0.25$, the response function is negative for disturbances of zero wavenumber, and remains less than unity throughout the range of α and Q_{s} investigated. (For very low Q_{s}, the response function does actually exceed unity at very large wavenumbers.) For $\beta = -0.25$, the response function is large and positive throughout most of the wavenumber range shown. On the scale shown, it is not clear that the response function ever passes through unity. In fact, for all the temperatures shown, the response function is 0.75 at $\alpha = 0$, meaning that in every case there is a solution, $\mathcal{R} = 1$, for some small value of α. The $\beta = -0.25$ disk admits loose triskele neutral modes, essentially independently of temperature.[1] Fig. 9.20 shows two of these.

---

[1] At least, it is apparent from fig. 9.19 that extremely high temperatures would be required to remove these neutral modes.

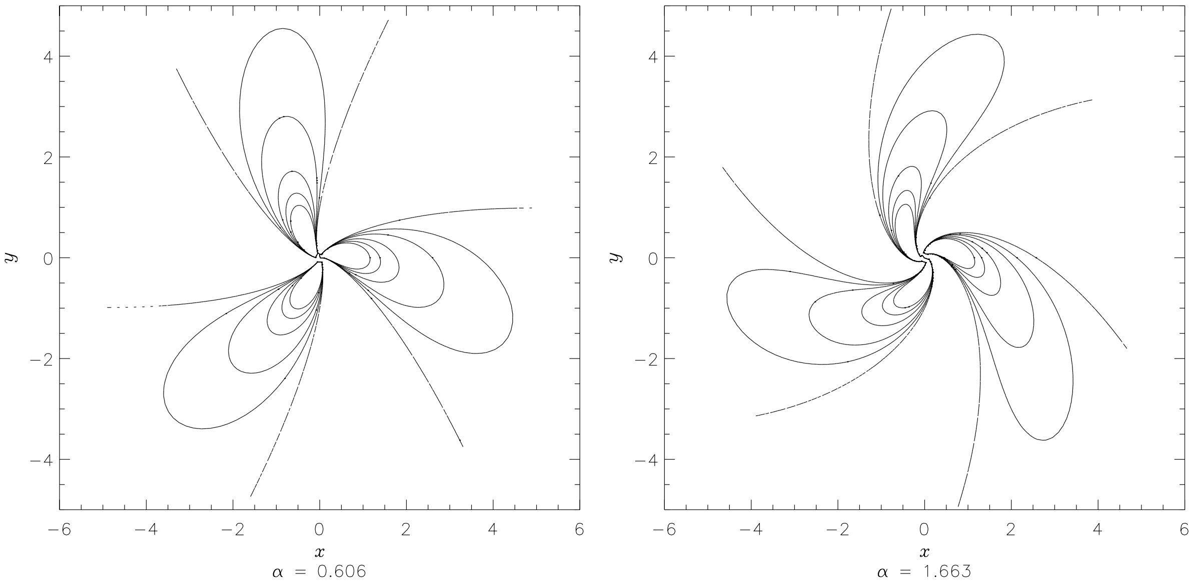


Figure 9.20: Neutral modes in $Q_\mathrm{s} = 1$ disks with $\beta = -0.25$ and $\beta = -0.50$. The left-hand plot shows a log-spiral with $\alpha = 0.606$; this is a neutral mode in the $\beta = -0.25$ disk with $Q_\mathrm{s} = 1$. The right-hand plot shows a log-spiral with $\alpha = 1.66$; this is a neutral mode in the $\beta = -0.50$ disk with $Q_\mathrm{s} = 1$. In each case, the density is calculated between $R = 0.1R_0$ and $R = 5R_0$. The solid lines mark 10%, 20%, 40%, 60% and 80% of the maximum density within this range; the dotted lines show the nodes, where the density perturbation is zero.

Thus the self-consistent disk, like the cut-out disk, has two distinct stability regimes in its response to neutral perturbations. We next investigate the value of β at which the switch between the two regimes occurs. Fig. 9.21 shows curves of $\mathcal{R}$ against wavenumber α for several different values of β between ±0.5.

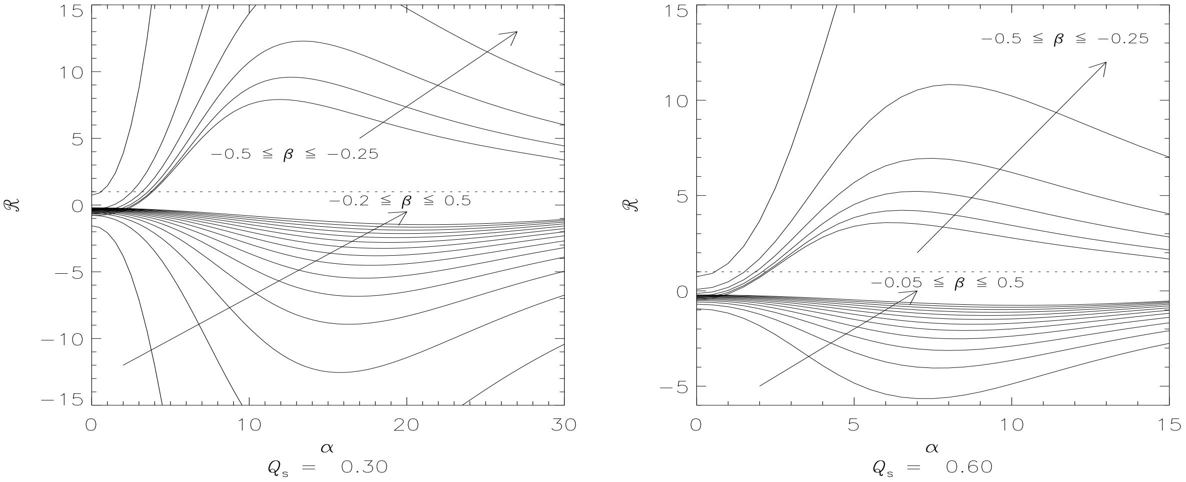


Figure 9.21: The response function for self-consistent disks with $Q_\mathrm{s} = 0.3$ (left-hand plot) and $Q_\mathrm{s} = 0.6$ (right-hand plot). In each case, curves are drawn showing the dependence of the response function on wavenumber for various values of β from -0.5 to 0.5. In each case, the curves fall into two distinct groups: $\beta \leq -0.25$ and $\beta > -0.25$. These are labelled with an arrow indicating the direction of increasing β. For $Q_\mathrm{s} = 0.6$, the convergence was very poor when β was slightly greater than -0.25. The curves with $\beta = -0.2$, -0.15 and -0.1 are therefore omitted. [Numerical accuracy parameters: $l_\mathrm{min} = -20$, $l_\mathrm{max} = +30$, $n_{GL} = 12$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 50$, $b_\mathrm{acc} = 2.5$.]

The left-hand plot shows results in disks with $Q_\mathrm{s} = 0.3$; the right-hand plot shows results when $Q_\mathrm{s} = 0.6$.

In each case, the curves fall into two groups with dramatically different behaviour. When $\beta \leq -0.25$, the curve starts with $\mathcal{R} < 1$ for $\alpha = 0$ and rises rapidly to large positive values, thus having a solution $\mathcal{R} = 1$ for small α. For more positive β, the response function is negative throughout the range of α examined. Any solutions which may exist will occur at large wavenumbers. Again, there is a degree of uncertainty in locating the exact β at which the switch occurs. The convergence was markedly worse for values of β slightly greater than -0.25. Indeed, for $Q_{\rm s} = 0.6$, reliable curves could not be obtained for $\beta = -0.20$, -0.15, -0.10; these three curves are therefore omitted. However, the evidence suggests that, as for the cut-out disks, $\beta = -0.25$ is the critical value.

The sudden change in the stability appears, then, to be related to the vanishing of $l\tilde{\kappa} + m\tilde{\Omega}$. We have found two distinct stability regimes. Disks with $\beta \leq -0.25$ contain no stars whose orbits close in an inertial frame, and admit neutral modes even at very high temperatures. Disks with $\beta > -0.25$ contain stars on closed orbits, and possess only modes with finite pattern speed, becoming stable at relatively low temperatures. We conclude that the stars on closed orbits act as absorbers and damp the waves. In the absence of such orbits, the disk is free to respond to a neutral perturbation.

9.4 Tetraskele modes

Global stability of the cut-out disks

As for $m = 3$ modes, we begin our study by examining the eigenvalue curves at moderate pattern speeds, in the range $\Omega_{\rm p} = 0.3$ to 1.3. Figs. 9.22 - 9.24 show these curves for $\beta = 0.25$, 0.00 and -0.25. In each case, the left-hand plot is for $Q_{\rm s} = 1$, and the right-hand plot for $Q_{\rm s} = 0.5$. In the latter case, the eigenvalue curves are ragged, as different eigenvalues become dominant.

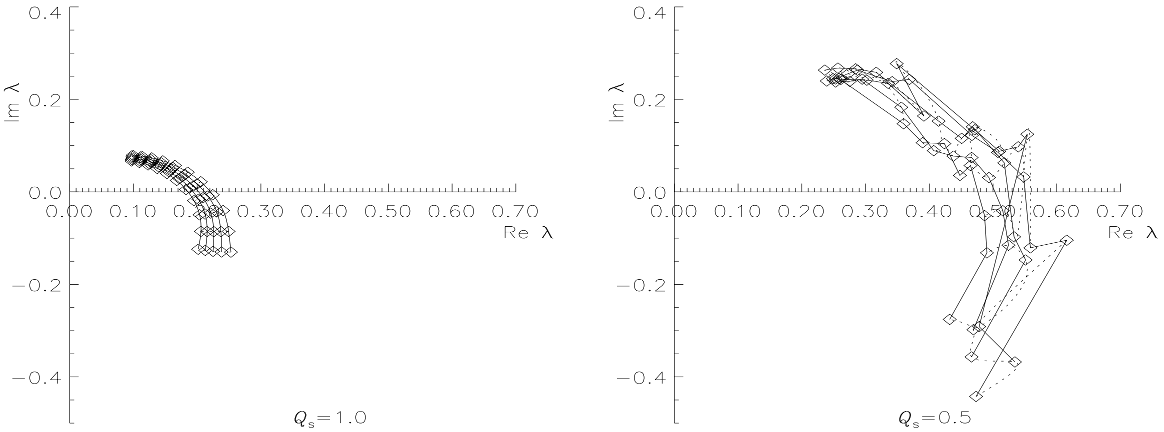


Figure 9.22: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = 0.25$. The left-hand plot is for $Q_{\rm s} = 1.0$, i.e. $\tilde{\sigma}_u = 0.283$ and $\gamma = 11.0$. The right-hand plot is for $Q_{\rm s} = 0.5$, i.e. $\tilde{\sigma}_u = 0.141$, $\gamma = 48.5$. The solid lines show curves of constant s at intervals of 0.1 from $s = 0^+$ (right-most curve) to $s = 0.4$ (left-most curve). The dotted lines show curves of constant $\Omega_{\rm p}$ at intervals of 0.1 from $\Omega_{\rm p} = 0.3$ (lowest curve) to $\Omega_{\rm p} = 1.3$ (highest curve). [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$.]

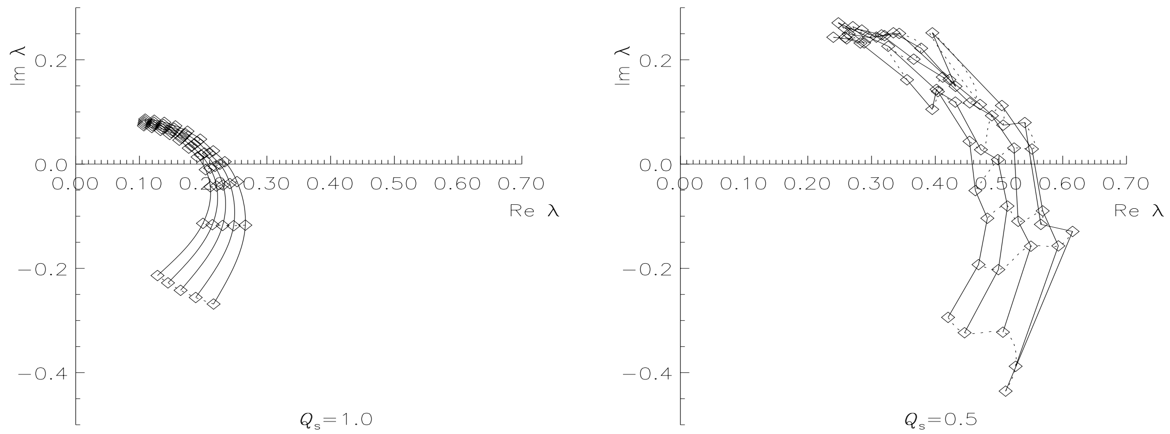


Figure 9.23: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = 0.00$. The left-hand plot is for $Q_{\rm s} = 1.0$, i.e. $\tilde{\sigma}_u = 0.378$ and $\gamma = 6.00$. The right-hand plot is for $Q_{\rm s} = 0.5$, i.e. $\tilde{\sigma}_u = 0.189$ and $\gamma = 27.0$. The solid lines show curves of constant s at intervals of 0.1 from $s = 0^+$ (right-most curve) to $s = 0.4$ (left-most curve). The dotted lines show curves of constant $\Omega_{\rm p}$ at intervals of 0.1 from $\Omega_{\rm p} = 0.5$ (lowest curve) to $\Omega_{\rm p} = 1.3$ (highest curve). [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$.]

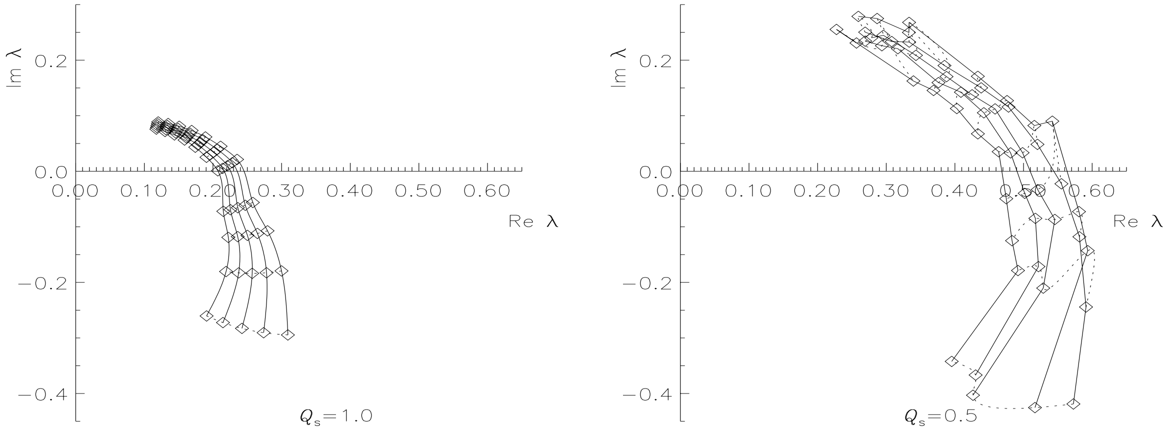


Figure 9.24: The dependence of the largest eigenvalue on growth rate and pattern speed, for $N = 2$ and $\beta = -0.25$. The left-hand plot is for $Q_{\rm s} = 1.0$, i.e. $\tilde{\sigma}_u = 0.509$ and $\gamma = 3.36$. The right-hand plot is for $Q_{\rm s} = 0.5$, i.e. $\tilde{\sigma}_u = 0.254$ and $\gamma = 14.9$. The solid lines show curves of constant s at intervals of 0.1 from $s = 0^+$ (right-most curve) to $s = 0.4$ (left-most curve). The dotted lines show curves of constant $\Omega_{\rm p}$ at intervals of 0.1 from $\Omega_{\rm p} = 0.5$ (lowest curve) to $\Omega_{\rm p} = 1.3$ (highest curve). [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\rm min} = -30$, $l_{\rm max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 1.5$.]

All three disks are very stable to $m = 4$ perturbations. We find the by now familiar result that disks with rising rotation curves are less stable than disks with falling rotation curves. Before presenting graphs of critical temperature, we investigate the convergence of the largest mathematical eigenvalue. The following tables show the convergence of the largest mathematical eigenvalue for a perturbation with $m = 4$, $\Omega_{\rm p} = 0.8$ and $s = 10^{-7}$, in a disk with $\beta = 0.25$, $N = 3$. We set $l_{\rm min} = -20$, $l_{\rm max} = +30$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$ and $b_{\rm acc} = 2.0$, and investigate the effect of changing n, $\Delta\alpha$ and $n_{\rm GL}$.

For $Q_{\mathrm{s}} = 0.8$, the convergence is good. A grid covering a range of 50 in α is sufficient to obtain 3 s.f. accuracy.

n_{GL}	Grid-spacing $\Delta\alpha$	Number of grid points n	Largest mathematical eigenvalue
7	1.000	51	$0.335239 + 0.129574i$
	0.500	101	$0.330833 + 0.114177i$
	0.403	125	$0.332074 + 0.109966i$
	0.301	167	$0.330790 + 0.114326i$
	0.200	251	$0.330825 + 0.114176i$
	0.100	501	$0.330823 + 0.114175i$
9	1.000	51	$0.343349 + 0.139573i$
	0.500	101	$0.330825 + 0.114189i$
	0.403	125	$0.330822 + 0.114188i$
	0.301	167	$0.330819 + 0.114188i$
	0.200	251	$0.330818 + 0.114187i$
	0.100	501	$0.330815 + 0.114186i$

A larger grid, covering a range of 100 in α, is sufficient for 6 s.f. accuracy. The results shown in the following table are unchanged when the grid is extended to cover 150 in α.

n_{GL}	Grid-spacing $\Delta\alpha$	Number of grid points n	Largest mathematical eigenvalue
7	1.000	101	$0.330974 + 0.114169i$
	0.500	201	$0.330970 + 0.114179i$
	0.400	251	$0.330970 + 0.114179i$
	0.299	335	$0.330969 + 0.114179i$
	0.200	501	$0.330969 + 0.114179i$
9	1.000	101	$0.330959 + 0.114173i$
	0.500	201	$0.330955 + 0.114182i$
	0.400	251	$0.330955 + 0.114182i$
	0.299	335	$0.330955 + 0.114182i$
	0.200	501	$0.330955 + 0.114182i$

However, for $Q_{\mathrm{s}} = 0.2$, the convergence is much poorer:

n_{GL}	Grid-spacing $\Delta\alpha$	Number of grid points n	Largest mathematical eigenvalue
7	1.000	51	$1.30458 + 0.230934i$
	0.500	101	$1.30095 + 0.226455i$
	0.403	125	$1.25248 + 0.164168i$
	0.301	167	$1.25350 + 0.159987i$
	0.200	251	$1.24946 + 0.163155i$
	0.100	501	$1.24774 + 0.162588i$
9	1.000	51	$1.40498 + 0.268700i$
	0.500	101	$1.25460 + 0.165147i$
	0.403	125	$1.25248 + 0.164167i$
	0.301	167	$1.25058 + 0.163531i$
	0.200	251	$1.24946 + 0.163154i$
	0.100	501	$1.24774 + 0.162588i$

n_{GL}	Grid-spacing $\Delta\alpha$	Number of grid points n	Largest mathematical eigenvalue
7	1.000	101	$1.63324 + 0.399320i$
	0.500	201	$1.63305 + 0.398125i$
	0.400	251	$1.70814 + 0.394478i$
	0.299	335	$1.64058 + 0.397632i$
9	1.000	101	$1.83738 + 0.516992i$
	0.500	201	$1.63331 + 0.398036i$
	0.400	251	$1.70814 + 0.394478i$
	0.299	335	$1.64068 + 0.397753i$

Fig. 9.25 shows how low the temperature must be, in terms of the velocity dispersion $\tilde{\sigma}_u$ and the stability parameter Q_{s}, before the marginal eigenvalue curve intersects the real axis at unity. Even lower temperatures are required than for $m = 3$ (fig. 9.8). As we have just seen, at these low temperatures the convergence is very poor. The values shown for $\tilde{\sigma}_{u,\min}$ and Q_{s} in fig. 9.25 are probably accurate to only 1 or 2 s.f.

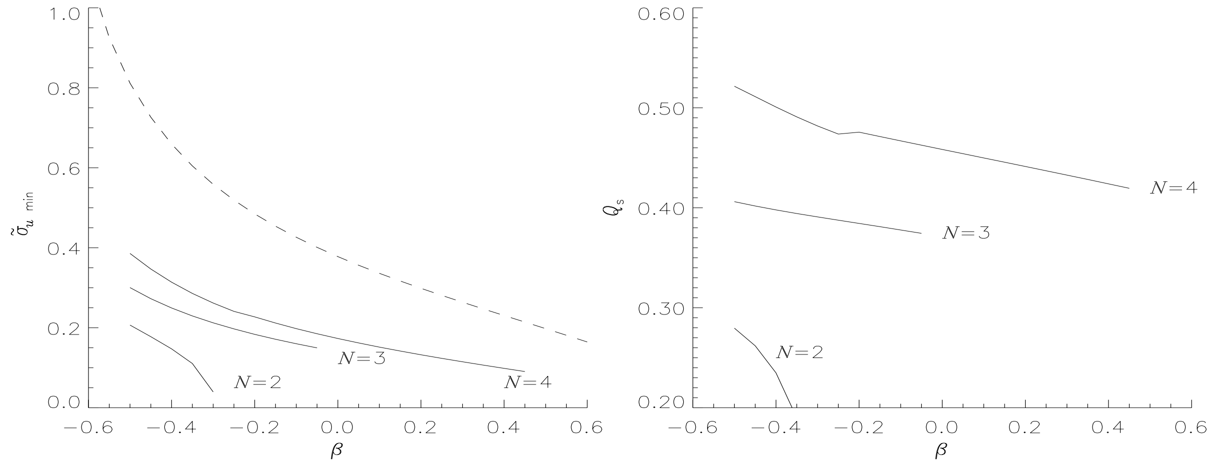
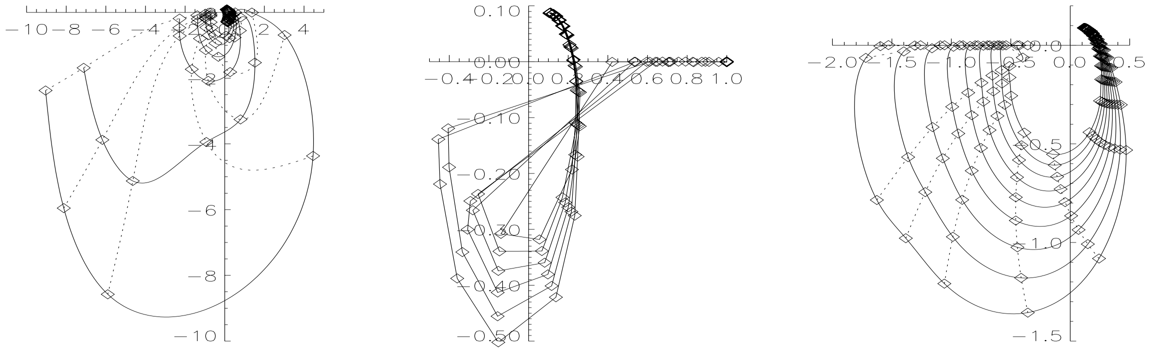


Figure 9.25: Minimum temperature plotted against β, for cut-out disks. The left-hand plot shows the minimum velocity dispersion $\tilde{\sigma}_u$; the right-hand one the same data presented in terms of the stability parameter Q_{s}. The solid lines indicate the results for inner cut-out disks with N = 1, 2, 3, 4. The broken line shows the temperature necessary for global axisymmetric stability in the self-consistent disk. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\min} = -20$, $l_{\max} = +30$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20$, $b_{\mathrm{acc}} = 2.0$, $\epsilon_\lambda = 10^{-4}$.]

Once again, disks where the centre has been cut out more sharply are more unstable than those where it has been removed relatively gently. Disks with rising rotation curves are more unstable than those with falling rotation curves, both in terms of $\tilde{\sigma}_{u,\min}$, the absolute amount of velocity dispersion required, and in terms of Q_{s}, the ratio of the velocity dispersion needed relative to that necessary for axisymmetric stability.

Neutral modes

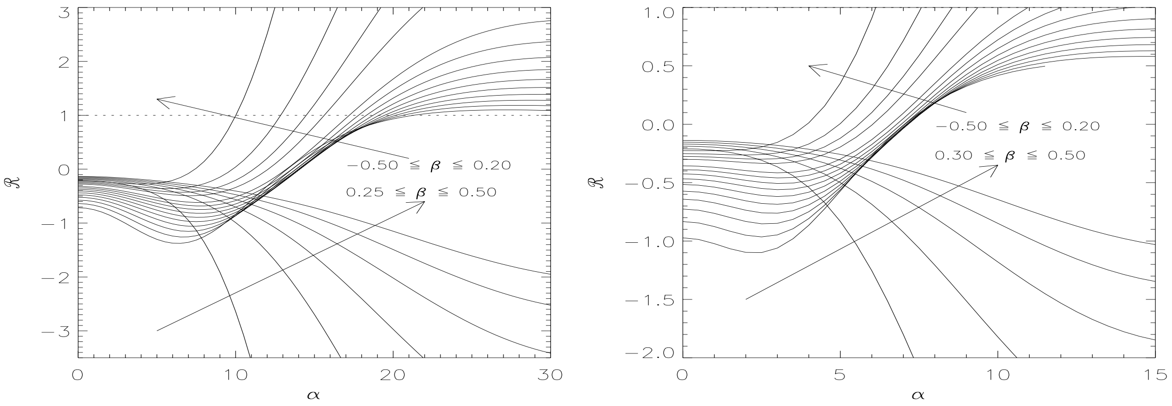
Once again, the marginal modes presented here are not the whole story. As for $m = 3$, some of the power-law disks may admit modes with vanishing growth rate and pattern speeds, even at high temperatures. Fig. 9.26 shows the marginal eigenvalue curves, down to vanishing pattern speeds, for various β.



$0.25 \leq \beta \leq 0.50$ $-0.05 \leq \beta \leq 0.20$ $-0.50 \leq \beta \leq -0.10$

Figure 9.26: Marginal eigenvalue curves for inner cut-out disks with $N = 2$, $Q_{\rm s} = 1$ and various values of β. Each plot shows the largest mathematical eigenvalues for vanishing growth rate, $s = 10^{-7}$, and 20 values of the pattern speed: $\Omega_{\rm p}$ = 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.3, ... 1.2, 1.3. The left-hand plot shows these marginal eigenvalue curves for $\beta = 0.25$, 0.30, 0.35, 0.40, 0.45, 0.50. The middle plot shows the curves for $\beta = -0.05$, -0.10, ... 0.15, 0.20. The right-hand plot shows the curves for $\beta = -0.50$, -0.45 ... -0.15, -0.10. [Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.5$, $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 12$, $f_\sigma = 0.8$, $a_{\rm acc} = 50$, $b_{\rm acc} = 2.5$.]

The left-hand plot shows values of β which are sufficiently large that $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes for some eccentric velocity. In accordance with our results for $m = 3$ (fig. 9.17), these disks do not admit neutral modes. The middle plot shows values of β just low enough that the term $l\tilde{\kappa} + m\tilde{\Omega}$ never vanishes. The eigenvalue becomes real and positive as $\Omega_{\rm p} \to 0$, again agreeing with our results for $m = 3$. However, the right-hand plot shows that for still lower values of β, the eigenvalue becomes real and *negative* as $\Omega_{\rm p} \to 0$. Even though there are no stars on closed four-lobed orbits, there are no $m = 4$ neutral modes. Once again, the response of the self-consistent disks to neutral perturbations sheds some light on the issue. Fig. 9.27 shows the neutral response function for several values of β.



$Q_{\rm s} = 0.30$ $Q_{\rm s} = 0.60$

Figure 9.27: The response function for self-consistent disks with $Q_{\rm s} = 0.3$ (left-hand plot) and $Q_{\rm s} = 0.6$ (right-hand plot). In each case, curves are drawn showing the dependence of the response function on wavenumber for 20 values of β from -0.5 to 0.5 (in steps of 0.05, omitting $\beta = 0$). In each case, the curves fall into two distinct groups: $\beta \leq 0.2$ and $\beta > 0.2$. These are labelled with an arrow indicating the direction of increasing β. Note the different vertical scales in the two plots. [Numerical accuracy parameters: $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 12$, $f_\sigma = 0.8$, $a_{\rm acc} = 50$, $b_{\rm acc} = 2.5$.]

The curves clearly fall into two groups. Once again, it is hard to obtain accurate results close to the value of β separating disks where $l\tilde{\kappa} + m\tilde{\Omega}$ vanishes from those where it does not. However, the evidence is certainly consistent with the division's occurring at $\beta = 2/9$.

Comparing the left-hand and right-hand plots in fig. 9.27, we see that raising the temperature from $Q_{\rm s} = 0.3$ to $Q_{\rm s} = 0.6$ has sufficed to banish neutral modes in many of the disks. At $Q_{\rm s} = 0.3$, all the disks with $\beta < 2/9$ admit neutral modes. At $Q_{\rm s} = 0.6$, the response curves of disks with $\beta \leq -0.2$ peak below the line $\mathcal{R} = 1$, and it appears unlikely that they will intersect it. As we approach $\beta = 0.2$, however, the response curves are rising steeply as they pass through $\mathcal{R} = 1$, and the evidence seems to be that they will require considerably higher temperatures before these neutral modes are abolished. Fig. 9.28 shows how the response function depends on temperature for three different values of β.

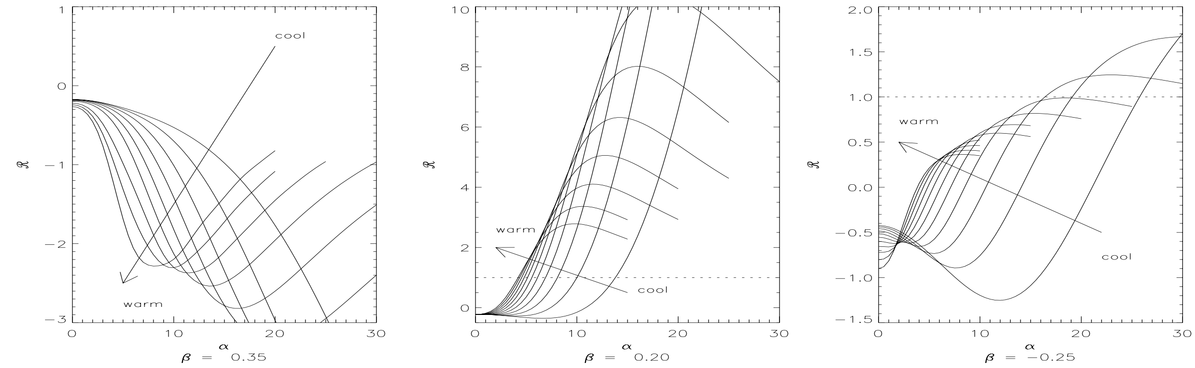


Figure 9.28: The response function for the self-consistent disks with $\beta = 0.35$, 0.20 and -0.25. For $\beta = 0.35$, curves are shown for 9 values of $Q_{\rm s}$ in steps of 0.1 from 0.2 to 1.0. For $\beta = 0.20$ and -0.25, curves are shown for 11 values of $Q_{\rm s}$ in steps of 0.1 from 0.2 to 1.2. Note the different vertical scales. [Numerical accuracy parameters: $l_{\rm min} = -20$, $l_{\rm max} = +30$, $n_{GL} = 12$, $f_\sigma = 0.8$, $a_{\rm acc} = 50$, $b_{\rm acc} = 2.5$.]

For $\beta = 0.35$, the disk admits no neutral modes at the temperatures and wavenumbers studied. For $\beta = 0.20$, the disk admits neutral modes at *all* the temperatures and wavenumbers studied. The disk with $\beta = -0.25$ is essentially of the same type as $\beta = 0.20$, but admits neutral modes only up to moderate temperatures ($Q_{\rm s} \approx 0.5$).

We can now understand the different behaviours of the cut-out disks shown in fig. 9.26. These curves were drawn at $Q_{\rm s} = 1$, which is hot enough to prevent neutral modes in disks with $\beta \leq -0.1$, while disks with $\beta \geq 0.25$ do not admit neutral modes even at low temperatures. The marginal eigenvalue curves for both these ranges of β therefore end up on the negative real axis as $s \to 0$ and $\Omega_{\rm p} \to 0$. Values of β only slightly less than the dividing value $\beta = 2/9$, however, require much higher temperatures to prevent neutral modes. They still admit neutral modes at $Q_{\rm s} = 1$, and so their marginal eigenvalue curves end up on the *positive* real axis. So for $m = 4$, as for $m = 3$, neutral modes are suppressed by the presence of closed orbits. Tetraskele neutral modes are also much more easily quelled by temperature than are triskele neutral modes.

9.5 Summary

In this Chapter, we have examined the response to perturbations with $m = 3$ and $m = 4$. The power-law disks are extremely stable to modes with these symmetries; very low temperatures are required before growing modes become possible. For values of β above a critical value, closed three- and four-lobed orbits exist in the inertial frame. The presence of these orbits somehow prevents neutral modes from occurring. For disks with β below the critical value, there are no closed orbits with the same symmetry as the perturbing potential. Neutral modes are then possible even when there are no growing modes, For $m = 3$, such neutral modes persist even at high temperatures.

Chapter 10

Summary and Conclusions

The work in this thesis describes a linear normal mode analysis of a family of self-gravitating, differentially rotating stellar dynamical disks. Our analysis includes both self-consistent disks, in which the surface density and the potential are related through Poisson's equation, and cut-out disks, in which the central core is carved out and the active matter is restricted to an annulus. The analytic and numerical procedures have exploited the self-similar potential of the models to develop a fast code that finds the pattern speeds and growth rates of the normal modes. We now summarise the main new results and bring to a resolution the question of the mode structure of the self-consistent disks left unsolved at the end of chapter 4.

The effect of the rotation curve

One of the aims of our work was to understand whether disks with falling rotation curves were more or less stable than those with rising rotation curves. Let us recall that our models have rotation curves varying like a power of radius, namely $v_{\mathrm{circ}} \propto R^{-\beta/2}$. The numerical work leaves no doubt that the disks with rising rotation curves ($\beta < 0$) are more unstable than the disks with falling rotation curves ($\beta > 0$). For all azimuthal wavenumbers, the unstable modes persist to higher temperatures and grow more vigorously if $\beta < 0$. One way of understanding this result is to recall that the disks with $\beta < 0$ are closer to the limit of solid-body rotation. Obviously, the maintenance of any spiral pattern is helped as the differential rotation or shear is minimised. Increasing the shear leads to a more rapid damping of the response to any disturbance. Individual stars continue in perturbed orbits, but the differential rotation causes the phase mixing of the disturbance distribution, so the overall response averages to zero within a few epicyclic periods.

The cut-out disks

The inner cut-out is crucial in allowing growing non-axisymmetric modes. Mathematically, as is evident from Chapter 4, this arises because the cut-out function breaks the self-similarity and introduces an additional degree of freedom, thereby permitting solutions of the integral equation. Physically, it is helpful to picture the instabilities as standing wave patterns caused by the interference of travelling waves. For $m > 1$, the inner Lindblad resonance absorbs incident waves and damps any disturbance. Growing modes therefore have pattern speeds which are large enough to push the inner Lindblad resonance within the shield of the inner cut-out. The inner cut-out is most effective as a reflector of waves when it is abrupt. Disks with lower cut-out indices allow more of the wave to seep through to the resonance and are thus more stable. The $N = 1$ disk has the mildest cut-out and is the most stable of all.

The self-consistent disks

There are no length-scales and time-scales in the scale-free disks. If any mode is admitted at some pattern speed and growth rate, then it must be present at all pattern speeds and growth rates. Is a two-dimensional continuum of non-axisymmetric modes possible? Our mathematical analysis in chapter 4 suggested that this was not likely, although it fell short of providing a rigorous proof. In the light of our analysis of the cut-out disks, there is one further piece of evidence that we now wish to adduce. Non-axisymmetric modes in disks are principally confined to the region between the inner and outer Lindblad radii. As the pattern speed becomes vanishingly small, the inner Lindblad radius moves well beyond the inner cut-out radius. Under such circumstances, the response of the inner cut-out disk must approach that of the self-consistent disk. Evidence that this limit is a reliable guide to the self-consistent disk is provided by the neutral modes – as the pattern speed and growth rate are made small, the eigenvalues of the cut-out disk tend to the positive or negative real axis according to whether or not the self-consistent disk admits neutral modes.

If a continuum of modes exists in the self-consistent disk, its presence should be sensed by the cut-out disks as the pattern speed is brought to zero. We would expect the eigenvalues to draw together and become virtually independent of growth rate as the pattern speed approaches zero. There is no evidence that this happens. Here, the reader is urged to look back at graphs such as fig. 7.15 ($m = 2$), fig. 8.1 ($m = 1$), figs. 9.12 - 9.14 ($m = 3$). Fig. 10.1 shows the eigenvalue spectra for $m = 2$ modes with four different growth rates and vanishingly small pattern speed, in a disk with $Q = 0.2$. The eigenvalues rise to the real axis as the growth rate becomes small, picking up the neutral modes. However, they show no sign of becoming independent of growth rate. Altogether, the weight of the evidence seems to us to suggest that there is no continuum and that *the self-consistent disks admit no growing non-axisymmetric modes at all.* It remains possible that the self-consistent disks admit a continuum of van Kampen modes.

The case of the axisymmetric modes is somewhat different. Our work in chapter 4 shows that for the $m = 0$ modes, the eigenvalue must be real and positive. The temperature marking the onset of the neutral modes is accurately given by local theory. On physical grounds, we expect these neutral modes to be followed by growing modes as the temperature of the disk is lowered below this critical value. But the self-similar disk cannot distinguish between modes with different growth rates. It appears, then, that the self-consistent disks must admit a one-dimensional continuum of growing axisymmetric modes.

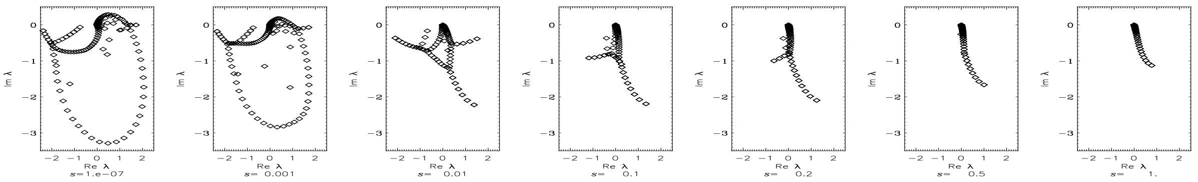


Figure 10.1: The eigenvalue spectra for $m = 2$ modes with vanishing pattern speed, in a disk with $N = 2$, $\beta = 0.25$, $Q_{\rm s} = 0.2$. Seven different growth rates are shown: $s = 10^{-7}$, 0.001, 0.01, 0.1, 0.2, 0.5, 1.0. In each case, $\Omega_{\rm p} = 10^{-7}$.

It is surprising that this continuum is not, apparently, sensed by the cut-out disks. On mathematical and physical grounds, we expect the presence of a cut-out to make little difference to the axisymmetric response. This expectation is borne out by the good agreement between the temperature at which neutral modes set into the self-consistent and cut-out disks. If the presence of the continuum were reflected in the cut-out disks, the mathematical eigenvalue would presumably become less dependent on growth rate as the temperature was lowered below $Q_{\rm s} = 1$. No such effect is seen (fig. 6.16).

The stability to different harmonics

Fig. 10.2 shows the minimum velocity dispersion needed for global stability to modes with different azimuthal wavenumbers. In almost all cases, the disks are considerably more stable to bar-like or bisymmetric modes than to axisymmetric disturbances. Less surprisingly, modes with $m = 3$ and $m = 4$ are very hard to excite. The disks must be made extremely cold before they admit growing modes with more than two arms.

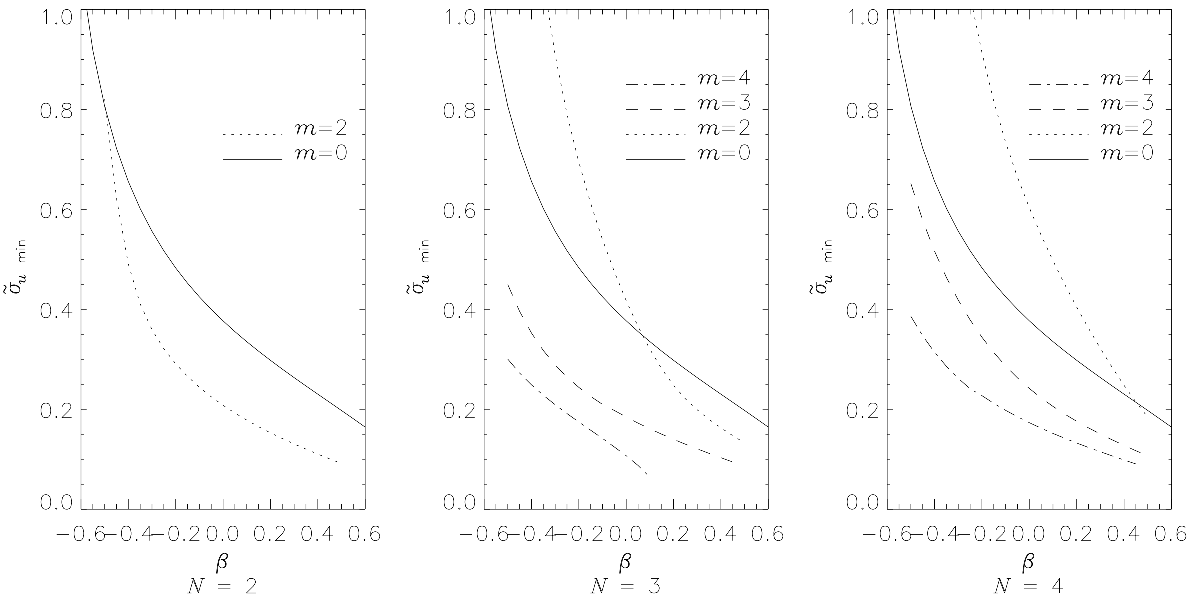


Figure 10.2: Minimum temperature for global stability to modes with different azimuthal symmetry.

The most serious unstabilities to which all the cut-out disks are susceptible are one-armed instabilities. At outset, we concede that our analysis permits the $m = 1$ modes to shift the barycentre of the galaxy. This artefact arises from the rigid core which has been carved out of the galaxy. A full treatment should really allow the core to move in response to the growing mode in such a way as to cancel the shifting of the barycentre. We do not believe this to be a serious flaw in our approach – first, because the barycentre moves only slightly within the linear régime and second, because N-body simulations by Earn (1993) appear to corroborate the conclusion that one-armed instabilities are the most serious. The prevalence of the one-armed modes can be easily understood at a simple level. First, there is no inner Lindblad resonance for $m = 1$. This removes a powerful stabilising mechanism. Second, this is the only harmonic for which the stars are driven at less than their natural frequency, thus responding in phase with the perturbation.

Appendix A

Reference Tables

A.1 Reference table of dimensional quantities

	Quantity	For the Toomre-Zang disk	For the general power-law disk
$\Sigma_{\rm eq}$	Equilibrium surface density	$\Sigma_{\rm eq} = \Sigma_0 \left(\frac{R_0}{R}\right)$	$\Sigma_{\rm eq} = \Sigma_0 \left(\frac{R_0}{R}\right)^{1+\beta}$
ψ	Self-consistent potential	$\psi(R) = v_0^2 \ln\left(\frac{R_0}{R}\right)$	$\psi(R) = \frac{v_\beta^2}{\beta}\left(\frac{R_0}{R}\right)^{\beta}$
$v_{\rm circ}$	Circular velocity	$v_{\rm circ} = v_0$	$v_{\rm circ} = v_\beta \left(\frac{R_0}{R}\right)^{\beta/2}$
v_β	Circular velocity at the reference radius R_0	$v_0^2 = 2\pi G \Sigma_0 R_0$	$v_\beta^2 = 2\pi G \Sigma_0 R_0 \frac{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}$
	Radial equation of motion	$\frac{d^2R}{dt^2} = \frac{L_z^2}{R^3} - \frac{v_0^2}{R}$	$\frac{d^2R}{dt^2} = \frac{L_z^2}{R^3} - \frac{v_\beta^2}{R}\left(\frac{R_0}{R}\right)^{\beta}$
ϕ	Potential energy $= -\psi$	$\phi(R) = -v_0^2 \ln\left(\frac{R_0}{R}\right)$	$\phi(R) = -\frac{v_\beta^2}{\beta}\left(\frac{R_0}{R}\right)^{\beta}$
E	Energy	$E = \frac{1}{2}\left(u^2+v^2\right) - v_0^2 \ln\left(\frac{R_0}{R}\right)$	$E = \frac{1}{2}\left(u^2+v^2\right) - \frac{v_\beta^2}{\beta}\left(\frac{R_0}{R}\right)^{\beta}$
L_z	Angular momentum	$L_z = R_{\rm H} v_0$	$L_z = R_{\rm H} v_\beta \left(\frac{R_0}{R_{\rm H}}\right)^{\beta/2}$
$R_{\rm H}$	Home radius	$R_{\rm H} = \frac{L_z}{v_0}$	$R_{\rm H} = R_0 \left(\frac{L_z}{v_\beta R_0}\right)^{\frac{2}{2-\beta}}$

	Quantity	For the Toomre-Zang disk	For the general power-law disk
U	Eccentric velocity	$U^2 = 2E - v_0^2 \left(2 \ln \frac{L_z}{v_0 R_0} + 1 \right)$	$U^2 = 2E + \left(\frac{2}{\beta} - 1 \right) \left(\frac{v_\beta^2 R_0^\beta}{L_z^\beta} \right)^{\frac{2}{2-\beta}}$
X	Logarithmic radius	$X = \ln R$	$X = \ln R$
ω	Complex frequency	$\omega = m\Omega_{\mathrm{p}} + is$	$\omega = m\Omega_{\mathrm{p}} + is$
κ	Radial frequency	$\kappa = \frac{v_0}{R_{\mathrm{H}}} \frac{2\pi}{\mathcal{J}_0}$	$\kappa = \frac{v_\beta}{R_{\mathrm{H}}} \left(\frac{R_0}{R_{\mathrm{H}}} \right)^{\beta/2} \frac{2\pi}{\mathcal{J}_0}$
Ω	Angular frequency	$\Omega = \frac{v_0}{R_{\mathrm{H}}} \frac{\mathcal{J}_2}{\mathcal{J}_0}$	$\Omega = \frac{v_\beta}{R_{\mathrm{H}}} \left(\frac{R_0}{R_{\mathrm{H}}} \right)^{\beta/2} \frac{\mathcal{J}_2}{\mathcal{J}_0}$
κ_0	Epicyclic frequency	$\kappa_0 = \sqrt{2} \times \frac{v_0}{R}$	$\kappa_0 = \sqrt{2-\beta} \times \frac{v_\beta}{R} \left(\frac{R_0}{R} \right)^{\beta/2}$
Ω_0	Circular frequency	$\Omega_0 = \frac{v_0}{R}$	$\Omega_0 = \frac{v_\beta}{R} \left(\frac{R_0}{R} \right)^{\beta/2}$

A.2 Reference table of dimensionless quantities

	Quantity	For the Toomre-Zang disk	For the general power-law disk
$\tilde{R}$	Radial coordinate	$\tilde{R} = \frac{R}{R_{\mathrm{H}}}$	$\tilde{R} = \frac{R}{R_{\mathrm{H}}}$
$\tilde{R}_{\mathrm{H}}$	Home radius	$\tilde{R}_H = \frac{R_{\mathrm{H}}}{R_0}$	$\tilde{R}_H = \frac{R_{\mathrm{H}}}{R_0}$
$\tilde{t}$	Time	$\tilde{t} = \frac{v_0}{R_{\mathrm{H}}} t$	$\tilde{t} = \frac{v_\beta}{R_{\mathrm{H}}} \tilde{R}_H^{-\beta/2} t$
	Radial equation of motion	$\frac{d^2\tilde{R}}{d\tilde{t}^2} = \frac{\tilde{L}_z^2}{\tilde{R}^3}\tilde{R}_H^{-2} - \frac{1}{\tilde{R}}$	$\frac{d^2\tilde{R}}{d\tilde{t}^2} = \frac{\tilde{L}_z^2}{\tilde{R}^3}\tilde{R}_H^{\beta-2} - \frac{1}{\tilde{R}^{1+\beta}}$
$\tilde{u}$	Radial velocity	$\tilde{u} = \frac{d\tilde{R}}{d\tilde{t}} = \frac{u}{v_0}$ $\tilde{u}^2 = \tilde{U}^2 + 1 - \tilde{R}^{-2} - 2\ln\tilde{R}$	$\tilde{u} = \frac{d\tilde{R}}{d\tilde{t}} = \frac{u}{v_\beta}\tilde{R}_H^{\beta/2}$ $\tilde{u}^2 = \tilde{U}^2 + 1 - \tilde{R}^{-2} + \frac{2}{\beta}\left(\tilde{R}^{-\beta} - 1\right)$
$\tilde{v}$	Tangential velocity	$\tilde{v} = \frac{\tilde{R}d\theta}{d\tilde{t}} = \frac{v}{v_0}$ $\tilde{v} = \tilde{R}^{-1}$	$\tilde{v} = \frac{\tilde{R}d\theta}{d\tilde{t}} = \frac{v}{v_\beta}\tilde{R}_H^{\beta/2}$ $\tilde{v} = \tilde{R}^{-1}$
$\tilde{E}$	Energy	$\tilde{E} = \frac{E}{v_0^2}$ $\tilde{E} = \frac{1}{2}\left(\tilde{u}^2 + \tilde{v}^2\right) - \ln\frac{R_0}{R}$	$\tilde{E} = \frac{E}{v_\beta^2}\tilde{R}_H^{\beta}$ $\tilde{E} = \frac{1}{2}\left(\tilde{u}^2 + \tilde{v}^2\right) - \frac{1}{\beta\tilde{R}^\beta}$
$\tilde{L}_z$	Angular momentum	$\tilde{L}_z = \frac{L_z}{R_0 v_0}$	$\tilde{L}_z = \frac{L_z}{R_0 v_\beta}$
$\tilde{L}_z$	Angular momentum	$\tilde{L}_z = \tilde{R}_H$	$\tilde{L}_z = \tilde{R}_H^{1-\beta/2}$
$\tilde{R}_{\mathrm{min}}$ $\tilde{R}_{\mathrm{max}}$	Extrema	The solutions of: $\tilde{U}^2 + 1 - \tilde{R}^{-2} - 2\ln\tilde{R} = 0$	The solutions of: $\tilde{U}^2 + 1 - \tilde{R}^{-2} + \frac{2}{\beta}\left(\tilde{R}^{-\beta} - 1\right) = 0$

	Quantity	For the Toomre-Zang disk	For the general power-law disk
$\tilde{U}$	Eccentric velocity	$\tilde{U} = \frac{U}{v_0}$ $\tilde{U}^2 = 2\tilde{E} - 1 - 2\ln \tilde{L}_z$ $\tilde{U}^2 = \tilde{u}^2 - 1 + \tilde{R}^{-2} + 2\ln \tilde{R}$ $\tilde{U}^2 = \tilde{u}^2 - 1 + \tilde{v}^2 - 2\ln \tilde{v}$	$\tilde{U} = \frac{U}{v_\beta}\tilde{R}_H^{\beta/2}$ $\tilde{U}^2 = 2\tilde{E} - 1 + \frac{2}{\beta}$ $\tilde{U}^2 = \tilde{u}^2 - 1 + \tilde{R}^{-2} - \frac{2}{\beta}\left(\tilde{R}^{-\beta} - 1\right)$ $\tilde{U}^2 = \tilde{u}^2 - 1 + \tilde{v}^2 - \frac{2}{\beta}\left(\tilde{v}^{\beta} - 1\right)$
$\mathcal{J}_n$	Auxiliary integral	$\mathcal{J}_n(\tilde{U}) = 2\times$ $\int_{\tilde{R}_{\min}}^{\tilde{R}_{\max}} \frac{d\tilde{R}}{\tilde{R}^n\left(\tilde{U}^2 + 1 - \tilde{R}^{-2} - 2\ln \tilde{R}\right)^{\frac{1}{2}}}$	$\mathcal{J}_n(\tilde{U}) = 2\times$ $\int_{\tilde{R}_{\min}}^{\tilde{R}_{\max}} \frac{d\tilde{R}}{\tilde{R}^n\left(\tilde{U}^2 + 1 - \tilde{R}^{-2} + \frac{2}{\beta}\left(\tilde{R}^{-\beta} - 1\right)\right)^{\frac{1}{2}}}$
$\tilde{\Omega}_{\mathrm{p}}$	Pattern speed	$\tilde{\Omega}_{\mathrm{p}} = \frac{R_0}{v_0}\Omega_{\mathrm{p}}$	$\tilde{\Omega}_{\mathrm{p}} = \frac{R_0}{v_\beta}\Omega_{\mathrm{p}}$
$\tilde{s}$	Growth rate	$\tilde{s} = \frac{R_0}{v_0}s$	$\tilde{s} = \frac{R_0}{v_\beta}s$
$\tilde{\omega}$	Complex frequency	$\tilde{\omega} = \frac{R_0}{v_0}\omega = m\tilde{\Omega}_p + i\tilde{s}$	$\tilde{\omega} = \frac{R_0}{v_\beta}\omega = m\tilde{\Omega}_p + i\tilde{s}$
$\tilde{\kappa}$	Radial frequency	$\tilde{\kappa} = \frac{R_{\mathrm{H}}}{v_0}\kappa = \frac{2\pi}{\mathcal{J}_0}$	$\tilde{\kappa} = \frac{R_{\mathrm{H}}}{v_\beta}\tilde{R}_H^{\beta/2}\kappa = \frac{2\pi}{\mathcal{J}_0}$
$\tilde{\Omega}$	Angular frequency	$\tilde{\Omega} = \frac{R_{\mathrm{H}}}{v_0}\Omega = \frac{\mathcal{J}_2}{\mathcal{J}_0}$	$\tilde{\Omega} = \frac{R_{\mathrm{H}}}{v_\beta}\tilde{R}_H^{\beta/2}\Omega = \frac{\mathcal{J}_2}{\mathcal{J}_0}$
$\tilde{\kappa}_0$	Epicyclic frequency	$\tilde{\kappa}_0 = \sqrt{2}$	$\tilde{\kappa}_0 = \sqrt{2-\beta}$
$\tilde{\Omega}_0$	Circular frequency	$\tilde{\Omega}_0 = 1$	$\tilde{\Omega}_0 = 1$

	Quantity	For the Toomre-Zang disk	For the general power-law disk
χ	Orbital phase	$\chi = \kappa t = \tilde{\kappa}\tilde{t}$	$\chi = \kappa t = \tilde{\kappa}\tilde{t}$
$\tilde{Y}$	Angular deviation	$\tilde{Y} = \theta - \Omega t = \theta - \tilde{\Omega}\tilde{t}$	$\tilde{Y} = \theta - \Omega t = \theta - \tilde{\Omega}\tilde{t}$
$\tilde{x}$	Logarithmic radius	$\tilde{x} = \ln\left(\frac{R}{R_0}\right)$	$\tilde{x} = \ln\left(\frac{R}{R_0}\right)$
$\tilde{X}$	Scaled logarithmic radius	$\tilde{X} = \ln\left(\frac{R}{R_{\mathrm{H}}}\right) = \ln \tilde{R}$	$\tilde{X} = \ln\left(\frac{R}{R_{\mathrm{H}}}\right) = \ln \tilde{R}$

A.3 Reference table of distribution functions and dependent quantities

Quantity	For the Toomre-Zang disk	For the general power-law disk
Self-consistent equilibrium distribution function	$f_s(E,L_z) = \tilde{C} L^\gamma \exp\left(-(\gamma+1)\frac{E}{v_0^2}\right)$ $\tilde{C} = \frac{C_{0\gamma}\Sigma_0}{2^{\gamma/2}\sqrt{\pi}R_0^\gamma v_0^{\gamma+2}}$ $C_{0\gamma} = \frac{(\gamma+1)^{1+\gamma/2}}{\Gamma\left[\frac{1}{2}(\gamma+1)\right]}$	$f_s(E,L_z) = \tilde{C} L^\gamma \mid E \mid^{1/\beta+\gamma/\beta-\gamma/2}$ $\tilde{C} = \frac{C_{\beta\gamma}\Sigma_0\|\beta\|^{1+\frac{1}{\beta}+\frac{\gamma}{\beta}}}{2^{\gamma/2}\sqrt{\pi}R_0^\gamma v_\beta^{2\left(1+\frac{1}{\beta}+\frac{\gamma}{\beta}\right)}}$ $\beta > 0 : C_{\beta\gamma} = \frac{\Gamma\left[2+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]}{\Gamma\left[\frac{1}{2}(\gamma+1)\right]\Gamma\left[1+\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}\right]}$ $\beta < 0 : C_{\beta\gamma} = \frac{\Gamma\left[\frac{\gamma}{2}-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}{\Gamma\left[\frac{1}{2}(\gamma+1)\right]\Gamma\left[-1-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}$
Anisotropy parameter	$\gamma = \frac{1}{\tilde{\sigma}_u^2} - 1$	$\gamma = \frac{1}{\tilde{\sigma}_u^2} - 1 - 2\beta$
Radial velocity dispersion	$\sigma_u^2 = \frac{v_0^2}{1+\gamma}$ $\tilde{\sigma}_u^2 = \frac{1}{1+\gamma}$	$\sigma_u^2 = \frac{v_\beta^2}{1+\gamma+2\beta}\left(\frac{R_0}{R}\right)^\beta$ $\tilde{\sigma}_u^2 = \frac{1}{1+\gamma+2\beta}$
Root-mean-squared tangential velocity	$\sqrt{\langle v^2\rangle} = \sqrt{1+\gamma}\cdot\sigma_u$	$\sqrt{\langle v^2\rangle} = \sqrt{1+\gamma}\cdot\sigma_u$
Mean streaming velocity	$\langle v\rangle = v_0\sqrt{\frac{2}{1+\gamma}}\frac{\Gamma\left[1+\frac{\gamma}{2}\right]}{\Gamma\left[\frac{1}{2}+\frac{\gamma}{2}\right]}$	$\beta > 0 :$ $\langle v\rangle = v_\beta\sqrt{\frac{2}{\beta}}\frac{\Gamma\left[1+\frac{\gamma}{2}\right]}{\Gamma\left[\frac{1}{2}+\frac{\gamma}{2}\right]}\frac{\Gamma\left[2+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]}{\Gamma\left[\frac{5}{2}+\frac{1}{\beta}+\frac{\gamma}{\beta}\right]}\left(\frac{R_0}{R}\right)^{\beta/2}$ $\beta < 0 :$ $\langle v\rangle = v_\beta\sqrt{\frac{2}{-\beta}}\frac{\Gamma\left[1+\frac{\gamma}{2}\right]\Gamma\left[-\frac{3}{2}-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}{\Gamma\left[\frac{1}{2}+\frac{\gamma}{2}\right]\Gamma\left[-1-\frac{1}{\beta}-\frac{\gamma}{\beta}\right]}\left(\frac{R}{R_0}\right)^{-\beta/2}$
Tangential velocity dispersion	$\sigma_v^2 = \langle v^2\rangle - \langle v\rangle^2$	$\sigma_v^2 = \langle v^2\rangle - \langle v\rangle^2$
Kinetic energy within radius R	$K(R) = \pi\Sigma_0 v_0^2 R_0^2\frac{2+\gamma}{(1+\gamma)}\frac{R}{R_0}$	$K(R) = \pi\Sigma_0 v_\beta^2 R_0^2\frac{2+\gamma}{(1+\gamma+2\beta)(1-2\beta)}\left(\frac{R}{R_0}\right)^{1-2\beta}$

Quantity	For the Toomre-Zang disk	For the general power-law disk
Potential energy within radius R	$W(R) = -2\pi\Sigma_0 v_0^2 R_0^2 \frac{R}{R_0}$	$W(R) = -\frac{2\pi\Sigma_0 v_\beta^2 R_0^2}{1-2\beta}\left(\frac{R}{R_0}\right)^{1-2\beta}$
Cut-out distribution function	$f(E, L_z) = H(L) f_s(E, L_z)$	$f(E, L_z) = H(L) f_s(E, L_z)$
Logarithmic angular momentum	$h = \ln \tilde{L}_z$	$h = \frac{2+\beta}{2-\beta} \ln \tilde{L}_z$
Inner cut-out function	$H(L) = \frac{L^N}{L^N + (v_0 R_0)^N}$ $\tilde{H}(\tilde{L}_z) = \frac{\tilde{L}_z^N}{\tilde{L}_z^N + 1}$ $\mathcal{H}(h) = \frac{e^{Nh}}{e^{Nh}+1}$	$H(L) = \frac{L_z^{N_\beta}}{L_z^{N_\beta} + (v_\beta R_0)^{N_\beta}}$ $\tilde{H}(\tilde{L}_z) = \frac{\tilde{L}_z^{N_\beta}}{\tilde{L}_z^{N_\beta} + 1}$ $\mathcal{H}(h) = \frac{e^{Nh}}{e^{Nh}+1}$
Outer cut-out function	$H(L) = \frac{L_c^M}{L_z^{M_\beta} + L_c^M}$ $\tilde{H}(\tilde{L}_z) = \frac{1}{\left(\tilde{L}_z/\tilde{L}_\mathrm{c}\right)^M + 1}$ $\mathcal{H}(h) = \frac{1}{\left[\tilde{L}_\mathrm{c}^{-M} e^{Mh} + 1\right]}$	$H(L) = \frac{L_c^{M_\beta}}{L_z^{M_\beta} + L_c^{M_\beta}}$ $\tilde{H}(\tilde{L}_z) = \frac{1}{\left(\tilde{L}_z/\tilde{L}_\mathrm{c}\right)^{M_\beta} + 1}$ $\mathcal{H}(h) = \frac{1}{\left[\tilde{L}_\mathrm{c}^{-M_\beta} e^{Mh} + 1\right]}$
Doubly cut-out function	$H(L) = \frac{L^N L_c^M}{\left[L^N + (v_0 R_0)^N\right]\left[L^M + L_c^M\right]}$ $\tilde{H}(\tilde{L}_z) = \frac{\tilde{L}_z^N}{\left[\tilde{L}_z^N + 1\right]\left[\left(\tilde{L}_z/\tilde{L}_\mathrm{c}\right)^M + 1\right]}$ $\mathcal{H}(h) = \frac{e^{Nh}}{\left[e^{Nh}+1\right]\left[\tilde{L}_\mathrm{c}^{-M} e^{Mh} + 1\right]}$	$H(L) = \frac{L_z^{N_\beta} L_c^{M_\beta}}{\left[L_z^{N_\beta} + (v_0 R_0)^{N_\beta}\right]\left[L_z^{M_\beta} + L_c^{M_\beta}\right]}$ $\tilde{H}(\tilde{L}_z) = \frac{\tilde{L}_z^{N_\beta}}{\left[\tilde{L}_z^{N_\beta} + 1\right]\left[\left(\tilde{L}_z/\tilde{L}_\mathrm{c}\right)^{M_\beta} + 1\right]}$ $\mathcal{H}(h) = \frac{e^{Nh}}{\left[e^{Nh}+1\right]\left[\tilde{L}_\mathrm{c}^{-M_\beta} e^{Mh} + 1\right]}$
Cut-out indices	N, M	N, M $N_\beta = \frac{2+\beta}{2-\beta} N, M_\beta = \frac{2+\beta}{2-\beta} M$

A.4 Reference table of quantities describing the response of the disk to a perturbation

Symbol	Meaning	Reference
α	The (dimensionless) wavenumber of a logarithmic spiral; successive peaks of the density distribution are separated by $2\pi R/\alpha$ in physical space.	Sec. 3.3, eqs. (3.21) and (3.24).
$A_{\mathrm{imp}}(\alpha)$	The transform of the imposed density perturbation; describes the amplitude of the imposed perturbation at the wavenumber α.	Sec. 3.3, eq. (3.26).
$A_{\mathrm{res}}(\alpha)$	The transform of the response density perturbation; describes the amplitude of the response at the wavenumber α.	Sec. 3.3, eq. (3.42).
λ	The mathematical eigenvalue. This is initially defined by the relationship $\lambda = A_{\mathrm{res}}(\alpha)/A_{\mathrm{imp}}(\alpha)$, i.e. λ is the ratio of the response to an imposed perturbation at a particular wavenumber. The resulting integral equation can be satisfied only for certain discrete values of λ – the eigenvalues – and the corresponding eigenvectors $A(\alpha)$. Self-consistent modes require $\lambda = 1$.	Sec. 5.5, eq. (5.6)
m	The azimuthal harmonic number.	Sec. 3.3, eq. (3.21).
l	The radial harmonic number. The time-dependence of the potential perturbation is described by a Fourier series of components with frequency l times the radial frequency κ.	Sec. 3.5, eq. (3.53).
$Q_{lm}(\alpha)$	The Fourier coefficient; describes the contribution of a log-spiral with azimuthal symmetry m and wavenumber α to the lth component of the potential perturbation experienced by a star with a particular eccentric velocity $\tilde{U}$.	Sec. 3.5, eqs. (3.53) and (3.55).
α'	The logarithmic wavenumber of a component of the imposed perturbation, used as an argument of the transfer function $\mathcal{S}_m(\alpha, \alpha')$.	Sec. 3.4.
$\mathcal{S}_m(\alpha, \alpha')$	The transfer function; describes how much the imposed component with wavenumber α' contributes to the response component with wavenumber α.	Sec. 3.4, eq. (3.45); sec. 3.5, eq. (3.79).
η	The difference between the response and imposed wavenumbers, $\eta = \alpha - \alpha'$.	Sec. 3.5.
$\hat{\eta}$	$\hat{\eta} = \eta \times \frac{2}{2+\beta}$	App. C.

Symbol	Meaning	Reference
$F_{lm}(\eta)$	The angular momentum function. Adds up orbits of similar shape at all possible scales, and describes how much the imposed component contributes to the response a distance η away in wavenumber-space.	Sec. 3.5, eq. (3.77); App. C.
α_{u}	Most unstable wavenumber, i.e. the dominant wavenumber of the first mode to set in as the disk is cooled. Neutral modes in the self-consistent disk are pure log-spirals; α_{u} is then simply the wavenumber of the first mode to appear. Where modes are superpositions of many log-spirals, the dominant wavenumber is defined to be the wavenumber at which the density transform has its maximum amplitude.	Sec. 6.2, eq. (6.10); sec. 6.6.
$\tilde{\sigma}_{u,\mathrm{min}}$	Minimum velocity dispersion needed for stability. This is the velocity dispersion at which the disk first admits marginal modes (with vanishing growth rate). For velocity dispersions less than $\tilde{\sigma}_{u,\mathrm{min}}$, the disk admits growing modes. For velocity dispersions greater than $\tilde{\sigma}_{u,\mathrm{min}}$, the disk is stable.	Sec. 6.2, eq. (6.11).
γ_{max}	Maximum anisotropy needed for stability. Equivalent to $\tilde{\sigma}_{u,\mathrm{min}}$; an alternative way of expressing the temperature of the disk.	Sec. 6.2.
Q_{s}	Minimum temperature needed for stability. Equivalent to $\tilde{\sigma}_{u,\mathrm{min}}$, but here the temperature of the disk is expressed in terms of Toomre's stability parameter Q. $Q_{\mathrm{s}} = \tilde{\sigma}_u/\tilde{\sigma}_{u,0}$, where: $\tilde{\sigma}_u$ is the velocity dispersion necessary to make the disk under examination globally stable to modes of the particular azimuthal symmetry being studied, and $\tilde{\sigma}_{u,0}$ is the velocity dispersion necessary to make the corresponding self-consistent disk locally stable to axisymmetric modes.	Sec. 6.2, eq. (6.15).
$R_{\mathrm{ILR}}, R_{\mathrm{OLR}}$	The inner and outer Lindblad resonances, at which the star encounters successive crests of the potential at the frequency of its own radial oscillations.	Sec. 7.3, eq. (7.3).
R_{CR}	The co-rotation radius, at which a star on a circular orbit co-rotates with the potential	Sec. 7.3, eq. (7.3).
$\mathcal{R}$	The response function, the ratio of the response to the imposed density. Used to describe the response of the self-consistent disk to neutral perturbations. Modes have $\mathcal{R} = 1$.	Sec. 4.3, eq. 4.12; sec. 6.2, eq. (6.7).

A.5 Reference table of numerical accuracy parameters

Quantity affected	Accuracy parameter	What it controls
Fourier components Q_{lm}	$n_\psi, a_{\rm acc}, b_{\rm acc}$	Number of steps in mid-point integration: $n_\psi = a_{\rm acc} \exp(b_{\rm acc} \tilde{U})$
Transfer function $\mathcal{S}_m$	$l_{\rm min}, l_{\rm max}$	$\sum_{l=l_{\rm min}}^{l_{\rm max}}$ approximates $\sum_{l=-\infty}^{+\infty}$
Transfer function $\mathcal{S}_m$	$n_{\rm GL}$	Number of Gauss-Laguerre abscissae
Transfer function $\mathcal{S}_m$	f_σ	Position of Gauss-Laguerre abscissae. Integration is carried out in $V = \frac{1}{2}\tilde{U}^2/(f_\sigma \tilde{\sigma}_u)^2$.
Mathematical eigenvalue λ	n	Number of points along each side of the grid
Mathematical eigenvalue λ	$\Delta\alpha$	Grid-spacing
Mode parameters $\Omega_{\rm p}, s, \gamma$	ϵ_λ	Accuracy to which unit eigenvalue is sought. Pattern speed and temperature or growth rate are adjusted until the eigenvalue is unity to within ϵ_λ.

A.6 Reference table of quantities involved in local theory.

$\alpha_{\rm crit}$	The critical wavenumber. In a cold disk, axisymmetric disturbances with wavenumbers greater than $\alpha_{\rm crit}$ are locally unstable.	$\alpha_{\rm crit} = \frac{\kappa_0^2 R}{2\pi G\Sigma}$ $\alpha_{\rm crit} = (2-\beta)\frac{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}$
$\lambda_{\rm crit}$	The critical wavelength. In a cold disk, axisymmetric disturbances with wavelength less than $\lambda_{\rm crit}$ are locally unstable.	$\lambda_{\rm crit} = \frac{2\pi R}{\alpha_{\rm crit}}$
x	Scaled wavenumber	$x = \frac{\alpha}{\alpha_{\rm crit}}$
y	Scaled temperature	$y = \frac{\alpha_{\rm crit}\sigma_u}{R\kappa_0} = \frac{\sigma_u\kappa_0}{2\pi G\Sigma}$
$\tilde{\sigma}_{u,\rm min}$	Critical velocity dispersion. Disks with velocity dispersion corresponding to $\tilde{\sigma}_u = \tilde{\sigma}_{u,\rm min}$ are locally just stable to axisymmetric perturbations.	$\sigma_{u,\rm min} = 3.3582830\frac{G\Sigma}{\kappa_0}$ $\tilde{\sigma}_{u,\rm min} = \frac{3.3582830}{2\pi\sqrt{2-\beta}}\frac{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}$
Q	Toomre's stability parameter. Disks with velocity dispersion corresponding to $Q = 1$ are locally just stable to axisymmetric perturbations.	$Q = \frac{\sigma_u}{\sigma_{u,\rm min}} = \frac{1}{3.3582830}\frac{\sigma_u\kappa_0}{G\Sigma}$ $Q = \frac{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}\frac{2\pi\sqrt{2-\beta}}{3.3582830}$
$\alpha_{\rm u}$	Most unstable wavenumber; the wavenumber of the perturbation to which the disk first becomes locally unstable as it is cooled.	$\alpha_{\rm u} = 1.8217367\frac{\kappa_0^2 R}{2\pi G\Sigma}$ $\alpha_{\rm u} = 1.8217367\,(2-\beta)\frac{\Gamma\left[\frac{1}{2}(1-\beta)\right]\Gamma\left[\frac{1}{2}(2+\beta)\right]}{\Gamma\left[\frac{1}{2}(1+\beta)\right]\Gamma\left[\frac{1}{2}(2-\beta)\right]}$
ϖ	The ratio of the forcing frequency to the natural radial frequency of the star.	$\varpi = \frac{m(\Omega_{\rm p}-\Omega_0)}{\kappa_0}$ $\varpi = \frac{m}{\sqrt{2-\beta}}\left\{\tilde{\Omega}_{\rm p}\left(\frac{R}{R_0}\right)^{\frac{2+\beta}{2}} - 1\right\}$
$\mathcal{F}$	The reduction factor describing how the effects of the spiral perturbation are lessened for stars with random motion in the radial direction.	$\mathcal{F} = \frac{\langle u\rangle_{\rm imp}}{\langle u\rangle_{\rm imp,\ cold}} = -\frac{\kappa_0(1-\varpi^2)}{\varpi k\psi_{\rm imp}}\langle u\rangle_{\rm imp}$

Appendix B

Single-Eccentricity Distribution Functions

In this thesis, we have used distribution functions built from powers of energy and angular momentum. We could equally construct distribution functions using alternative integrals of motion, such as the eccentric velocity and home radius. Disks with such distribution functions would in general contain orbits of many shapes (described by the eccentric velocity U) and sizes (described by the home radius R_{H}). In this Appendix, we consider disks in which all the stars have the same-shaped orbit, characterised by $U = U_r$. Mathematically, we look for distribution functions of the form

$$f_{\mathrm{s}} = F(R_{\mathrm{H}})\delta(U - U_r). \tag{B.1}$$

The surface density is

$$\Sigma_{\mathrm{eq}} = \Sigma_0 \left(\frac{R_0}{R}\right)^{1+\beta} = \iint f_{\mathrm{s}}(u,v)du\,dv. \tag{B.2}$$

We transform this to an integral over U and R_H, using the Jacobian given in (4.19), and substitute our assumed distribution function:

$$\Sigma_0 \left(\frac{R_0}{R}\right)^{1+\beta} = \left(1 - \frac{\beta}{2}\right)\frac{v_\beta}{R}\iint \left(\frac{R_0}{R_H}\right)^{\beta/2} \frac{\tilde{U}\,F(R_{\mathrm{H}})\,\delta(U-U_r)\,dU\,dR_H}{\sqrt{\tilde{U}^2 + 1 - \tilde{R}^{-2} + \frac{2}{\beta}(\tilde{R}^\beta - 1)}}. \tag{B.3}$$

We integrate over U, and then transform the integral over R_{H} to one over $\tilde{R} = R/R_H$:

$$\Sigma_0 \left(\frac{R_0}{R}\right)^{1+\beta/2} = \left(1 - \frac{\beta}{2}\right) 2v_\beta \tilde{U}_r \int_{\tilde{R}_{\mathrm{min}}}^{\tilde{R}_{\mathrm{max}}} \frac{\tilde{R}^{\beta/2} F\left(R/\tilde{R}\right)\,d\tilde{R}}{\tilde{R}^2\sqrt{\tilde{U}_r^2 + 1 - \tilde{R}^{-2} + \frac{2}{\beta}(\tilde{R}^\beta - 1)}}, \tag{B.4}$$

where $\tilde{U}_r$ is the value of U_r in dimensionless units.

Looking at the powers of R, we can guess at a solution of this integral equation, namely

$$F\left(\frac{R}{\tilde{R}}\right) = k\left(\frac{\tilde{R}}{R}\right)^{1+\beta/2}. \tag{B.5}$$

Substituting this into (B.4), and remembering the definition of the auxiliary integral $\mathcal{J}_n(\tilde{U})$ (2.27), we can solve for k and obtain

$$k = \frac{\Sigma_0 R_0^{1+\beta/2}}{(1-\frac{\beta}{2}) v_\beta \tilde{U}_r \mathcal{J}_{1-\beta}(\tilde{U}_r)} = \frac{\Sigma_0 R_0^{1+\beta}}{\left(1-\frac{\beta}{2}\right) R_{\mathrm{H}}^{\beta/2} U_r \mathcal{J}_{1-\beta}(\tilde{U}_r)}. \tag{B.6}$$

The distribution function is then

$$f_{\mathrm{s}}(U, R_{\mathrm{H}}) = \frac{\Sigma_0 R_0^{1+\beta}}{\left(1-\frac{\beta}{2}\right) U_r \mathcal{J}_{1-\beta}(\tilde{U}_r)} \frac{\delta(U-U_r)}{R_{\mathrm{H}}^{1+\beta}}. \tag{B.7}$$

This expression is valid for $\beta = 0$, but in that case the appropriate form for $\mathcal{J}_1$ (2.28) must be used.

All the stars in this disk have orbits of the same shape, but different home radii. Each star sweeps out an annulus as it orbits. The relative *width* of this annulus depends on U_r, but the overall *size* of the annulus depends on R_{H}. It is easy to imagine that if we add up annuli with a variety of R_{H} in suitable proportions, we could recover the surface density of the equilibrium disk; and this is what the distribution function (B.7) does. It seems intuitively right that the number of annuli needed falls off with R_{H} at the same rate $(1+\beta)$ as the surface density falls off with R. Similarly, we can see why the total number of annuli depends inversely on U_r: for high U_r, the annuli are very wide, and few are needed; for low U_r, the converse is true. Although the single-eccentricity distribution function is never used in this thesis, it seemed sufficiently interesting to be worth presenting here.

Appendix C

The Angular Momentum Function

In this Appendix, we perform the integration over angular momentum $\tilde{L}_z$ to arrive at the angular momentum function F_{lm} defined in Chapter 3 (3.77) as

$$F_{lm} = \frac{1}{2\pi}\int_0^\infty \frac{e^{-i\eta\frac{2}{2-\beta}\ln\tilde{L}_z}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}\tilde{L}_z^{\frac{2+\beta}{2-\beta}}}\left[\left\{(l\tilde{\kappa}+m\tilde{\Omega})\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\gamma m\right\}\tilde{H}-m\tilde{L}_z\frac{d\tilde{H}}{d\tilde{L}_z}\right]\frac{d\tilde{L}_z}{\tilde{L}_z}.$$

The integration is carried out over the variable h, where

$$h = \frac{2+\beta}{2-\beta}\ln\tilde{L}_z, \qquad \tilde{H}(\tilde{L}_z) = \mathcal{H}(h). \tag{C.1}$$

It is performed for three different cut-out factors, namely

(1) the *self-consistent* disk:

$$\mathcal{H}(h) = 1, \qquad \frac{d\mathcal{H}}{dh} = 0; \tag{C.2}$$

(2) the *inner cut-out* disk:

$$\mathcal{H}(h) = \frac{e^{Nh}}{e^{Nh}+1}, \qquad \frac{d\mathcal{H}}{dh} = \frac{Ne^{Nh}}{\left(e^{Nh}+1\right)^2}; \tag{C.3}$$

(3) and the *doubly cut-out* disk:

$$\mathcal{H}(h) = \frac{e^{Nh}}{\left[e^{Nh}+1\right]\left[\tilde{L}_c^{-M_\beta}e^{Mh}+1\right]}, \tag{C.4}$$

$$\frac{d\mathcal{H}}{dh} = \frac{Ne^{Nh}\left(\tilde{L}_c^{-M_\beta}e^{Mh}+1\right)-M\tilde{L}_c^{-M_\beta}e^{Nh}e^{Mh}}{\left[e^{Nh}+1\right]^2\left[\tilde{L}_c^{-M_\beta}e^{Mh}+1\right]^2}. \tag{C.5}$$

We proceed by splitting up the integral into two parts:

$$F_{lm}(\eta) = \left\{(l\tilde{\kappa}+m\tilde{\Omega})\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\gamma m\right\}J_1 - mJ_2, \tag{C.6}$$

where

$$J_1(\beta,\eta,\tilde{L}_c) = \frac{1}{2\pi}\frac{2-\beta}{2+\beta}\int_{-\infty}^{\infty}\frac{e^{-i\eta\frac{2}{2+\beta}h}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^h}\mathcal{H}(h)dh, \tag{C.7}$$

and

$$J_2(\beta,\eta,\tilde{L}_c) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-i\eta\frac{2}{2+\beta}h}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^h}\frac{d\mathcal{H}}{dh}dh. \tag{C.8}$$

We need only evaluate these for the case $\beta = 0$, since the general integrals are related to the $\beta = 0$ integrals as

$$J_1(\beta, \eta, \tilde{L}_{\rm c}) = \frac{2-\beta}{2+\beta} J_1\left(0, \hat{\eta}, \tilde{L}_{\rm c}^{\frac{2+\beta}{2-\beta}}\right), \qquad J_2(\beta, \eta, \tilde{L}_{\rm c}) = J_2\left(0, \hat{\eta}, \tilde{L}_{\rm c}^{\frac{2+\beta}{2-\beta}}\right), \tag{C.9}$$

where

$$\hat{\eta} \equiv \eta \times \frac{2}{2+\beta}. \tag{C.10}$$

$\tilde{L}_{\rm c}$ has been included as an argument, although clearly this is only relevant to the doubly cut-out disk. We suppress the third argument unless needed.

(1) The self-consistent disk

Here

$$J_1(0, \eta) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{e^{-i\eta h}}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^h} dh, \qquad J_2(0, \eta) = 0. \tag{C.11}$$

We first consider the case $\omega \neq 0$. J_1 may be evaluated by contour integration. The integrand has poles when

$$h_n = \ln \frac{l\tilde{\kappa} + m\tilde{\Omega}}{\tilde{\omega}} + 2ni\pi, \tag{C.12}$$

where n is an integer. The poles occur at intervals along a line parallel to the imaginary axis; note that $\tilde{\omega}$ is not in general real ($\omega = m\Omega_{\rm p} + is$), so that the poles are displaced from the lines $h = 2ni\pi$.

We evaluate the integral

$$\frac{1}{2\pi} \oint \frac{e^{\lambda h} e^{-i\eta h}}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^h} dh, \tag{C.13}$$

around a rectangular contour as shown in fig. C.1, with long sides at $h = 0$, $h = 2i\pi$, and short edges at $\pm L$, enclosing the pole at $h = \ln[(l\tilde{\kappa} + m\tilde{\Omega})/\tilde{\omega}]$. Here the logarithm denotes the principal value, i.e. the imaginary part lies between 0 and 2π.

Then

$$\begin{aligned} &\frac{1}{2\pi} \int_{-L}^{L} \frac{e^{\lambda h} e^{-i\eta h} dh}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^h} - \frac{e^{2i\pi\lambda} e^{2\pi\eta}}{2\pi} \int_{-L}^{L} \frac{e^{\lambda h} e^{-i\eta h} dh}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^h} + \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{\lambda L} e^{y\eta} e^{i\lambda y} e^{-i\eta L} dh}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^{iy} e^{L}} \\ &- \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{-\lambda L} e^{y\eta} e^{i\lambda y} e^{i\eta L} dh}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^{iy} e^{-L}} = -\frac{i}{l\tilde{\kappa} + m\tilde{\Omega}} e^{\lambda \ln \frac{l\tilde{\kappa} + m\tilde{\Omega}}{\tilde{\omega}}} e^{-\eta \ln \frac{l\tilde{\kappa} + m\tilde{\Omega}}{\tilde{\omega}}} \end{aligned} \tag{C.14}$$

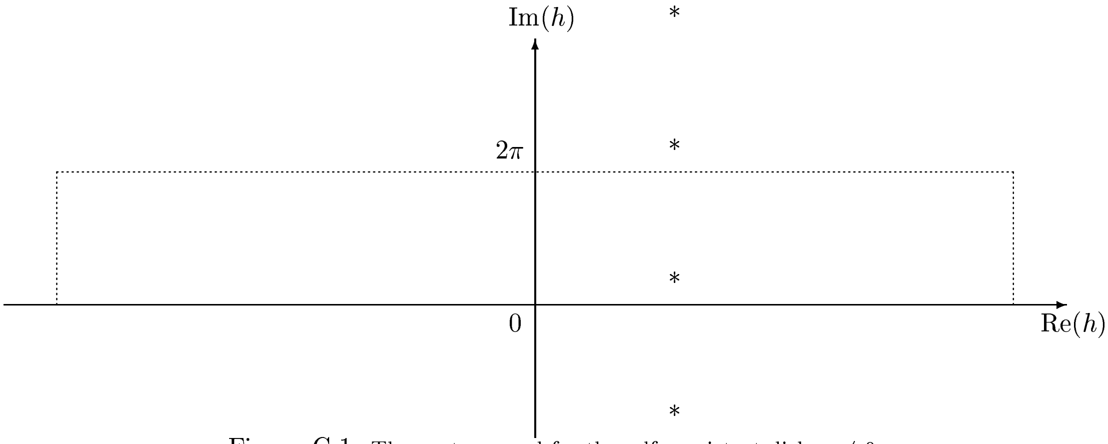


Figure C.1: The contour used for the self-consistent disk, $\eta \neq 0$.

Taking $L \to \infty$, and then $\lambda \to 0$, we obtain the result

$$J_1(0,\eta) = -\frac{i e^{-i\eta \ln \frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}}}{(l\tilde{\kappa}+m\tilde{\Omega})\left(1-e^{2\pi\eta}\right)}. \tag{C.15}$$

Then, generalising this result to all β as described above, we obtain

$$F_{lm}(\eta) = -\frac{2-\beta}{2+\beta}\left(\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right)\frac{i e^{-i\hat{\eta} \ln \frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}}}{\left(1-e^{2\pi\hat{\eta}}\right)}, \qquad \eta \neq 0. \tag{C.16}$$

This expression is singular when $\eta = 0$. As we now demonstrate, $F_{lm}(\eta)$ contains a delta-function at $\eta = 0$. For the case $\Omega_{\mathrm{p}} = 0$, $s = 0$, we can integrate the expression for $F_{lm}(\eta)$ directly:

$$J_1(0,\eta) = \frac{1}{l\tilde{\kappa}+m\tilde{\Omega}}\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-i\eta h} dh = \frac{1}{l\tilde{\kappa}+m\tilde{\Omega}}\delta(\eta). \tag{C.17}$$

and so for the general power-law disk

$$F_{lm}(\eta) = \frac{2-\beta}{2+\beta}\left\{\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right\}\delta(\hat{\eta}). \tag{C.18}$$

To obtain the expression for general Ω_{p} and s, we write

$$J_1(0,\eta) = J_1^{(1)}(0,\eta) + J_1^{(2)}\delta(\eta), \tag{C.19}$$

where $J_1^{(1)}(0,\eta)$ is equal to our previous expression (C.15) for $J_1(0,\eta)$ when $\eta \neq 0$, and equal to zero a$\eta = 0$. We integrate (C.19) over η, thus smoothing out the delta-function.

$$\int_{\eta=-\epsilon}^{+\epsilon} d\eta\, J_1(0,\eta) = \int_{\eta=-\epsilon}^{+\epsilon} d\eta\, J_1^{(1)}(0,\eta) + J_1^{(2)}\int_{\eta=-\epsilon}^{+\epsilon} d\eta\, \delta(\eta), \tag{C.20}$$

We now let $\epsilon \to 0$. We see from the expression for $J_1^{(1)}(\eta)$ already obtained (C.15) that $J_1^{(1)}(\eta)$ is odd in the limit of vanishingly small η. It therefore contributes nothing in (C.20). Using the expression (C.7) for $J_1(0,\eta)$, we obtain

$$J_1^{(2)} = \lim_{\epsilon \to 0} \frac{1}{2\pi} \int_{\eta=-\epsilon}^{+\epsilon} d\eta \int_{h=-\infty}^{\infty} dh \frac{e^{-i\eta h}}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^h} \tag{C.21}$$

Swapping the order of integration and performing the integration over η, we obtain

$$J_1^{(2)} = \lim_{\epsilon \to 0} \frac{1}{2\pi i} \int_{h=-\infty}^{\infty} \frac{dh}{h} \frac{e^{i\epsilon h} - e^{-i\epsilon h}}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^h} \tag{C.22}$$

We evaluate this by carrying out two separate contour integrals. For the integral with $e^{-i\epsilon h}$, we close the contour in the lower half-plane, as shown in fig. C.2.

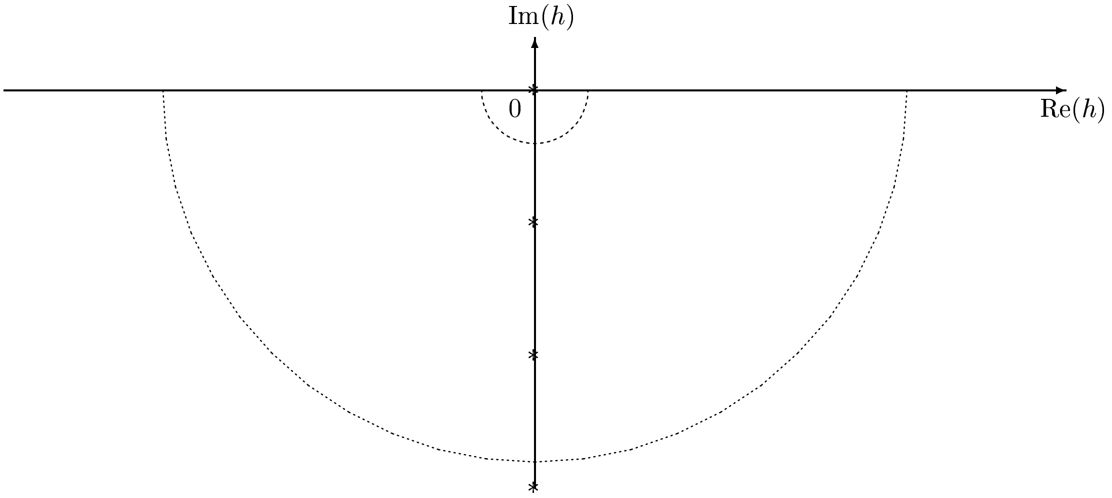


Figure C.2: The contour used for the self-consistent disk, $\eta = 0$.

As usual, we take the radius of the inner semi-circle to zero, and the that of the outer semi-circle to infinity. The integrand then vanishes on the outer semi-circle, and we obtain

$$\frac{1}{2\pi i} \int_{h=-\infty}^{\infty} \frac{dh}{h} \frac{e^{-i\epsilon h}}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^h} = \frac{1}{l\tilde{\kappa} + m\tilde{\Omega}} \sum_{n=-1}^{-\infty} \frac{e^{-i\epsilon \ln \frac{l\tilde{\kappa} + m\tilde{\Omega}}{\tilde{\omega}}} e^{2n\pi\epsilon}}{\ln \frac{l\tilde{\kappa} + m\tilde{\Omega}}{\tilde{\omega}} + 2in\pi} - \frac{1}{2} \frac{1}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega}} \tag{C.23}$$

The integral with $e^{+i\epsilon h}$ is evaluated over an analogous contour in the upper half-plane. We obtain

$$\frac{1}{2\pi i} \int_{h=-\infty}^{\infty} \frac{dh}{h} \frac{e^{i\epsilon h}}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^h} = -\frac{1}{l\tilde{\kappa} + m\tilde{\Omega}} \sum_{n=0}^{\infty} \frac{e^{i\epsilon \ln \frac{l\tilde{\kappa} + m\tilde{\Omega}}{\tilde{\omega}}} e^{-2n\pi\epsilon}}{\ln \frac{l\tilde{\kappa} + m\tilde{\Omega}}{\tilde{\omega}} + 2in\pi} + \frac{1}{2} \frac{1}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega}} \tag{C.24}$$

Subtracting these two results and taking the limit $\epsilon \to 0$, we obtain

$$\frac{1}{2\pi i} \int_{h=-\infty}^{\infty} \frac{dh}{h} \frac{e^{i\epsilon h} - e^{-i\epsilon h}}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^h} = \frac{1}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega}} + \frac{i}{l\tilde{\kappa} + m\tilde{\Omega}} \sum_{n=-\infty}^{\infty} \frac{1}{2n\pi - i \ln \frac{l\tilde{\kappa} + m\tilde{\Omega}}{\tilde{\omega}}} \tag{C.25}$$

The sum can be explicitly evaluated.

$$\begin{aligned}\sum_{n=-\infty}^{\infty}\frac{1}{2n\pi-iz} &= -\frac{1}{iz}+\sum_{n=-1}^{-\infty}\frac{1}{2n\pi-iz}+\sum_{n=1}^{\infty}\frac{1}{2n\pi-iz} = \frac{i}{z}+2iz\sum_{n=1}^{\infty}\frac{1}{(2n\pi)^2+z^2} \\ &= \frac{iz}{4\pi^2}\sum_{n=-\infty}^{\infty}\frac{1}{(n+iz/2\pi)(n-iz/2\pi)} = -\frac{1}{2}\cot\frac{iz}{2} = -\frac{i}{2}\frac{e^{-z}+1}{e^{-z}-1}\end{aligned} \tag{C.26}$$

where we have used the standard result (Prudnikov *et al.* 1986, eq. (5.1.6.4))

$$\sum_{n=-\infty}^{\infty}\frac{1}{(n+a)(n+b)} = \frac{\pi}{b-a}(\cot\pi a-\cot\pi b) \tag{C.27}$$

Thus we obtain

$$J_1^{(2)} = \frac{1}{2(l\tilde{\kappa}+m\tilde{\Omega})} \tag{C.28}$$

The angular momentum function for the self-consistent disk is then

$$F_{lm}(\eta) = \frac{2-\beta}{2+\beta}\left(\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right| - \frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right)\left(\frac{1}{2}\delta(\hat{\eta}) - \frac{ie^{-i\hat{\eta}\ln\frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}}}{1-e^{2\pi\hat{\eta}}}\right). \tag{C.29}$$

The corresponding result for the Toomre-Zang disk is (correcting a typographical error in Zang (1976), eq. (Z3.43))

$$F_{lm}(\eta) = \left(\gamma+1-\frac{\gamma m}{l\tilde{\kappa}+m\tilde{\Omega}}\right)\left(\frac{1}{2}\delta(\eta) - \frac{ie^{-i\eta\ln\frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}}}{1-e^{2\pi\eta}}\right). \tag{C.30}$$

(2) The inner cut-out disk

Here, we must evaluate

$$J_1(0,\eta) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-i\eta h}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^h}\frac{e^{Nh}}{e^{Nh}+1}dh, \tag{C.31}$$

$$J_2(0,\eta) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-i\eta h}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^h}\frac{Ne^{Nh}}{\left(e^{Nh}+1\right)^2}dh. \tag{C.32}$$

Again the integrands have poles when $h=\ln[(l\tilde{\kappa}+m\tilde{\Omega})/\tilde{\omega}]+2ni\pi$. However, they now also have poles along the imaginary axis, at

$$h_j = \frac{2j-1}{N}i\pi, \tag{C.33}$$

where j is an integer. We integrate around the same rectangular contour as for the self-consistent disk (fig. C.1). As well as the pole at $\ln[(l\tilde{\kappa}+m\tilde{\Omega})/\tilde{\omega}]$, we now also enclose N poles lying on the imaginary axis. Adding up the residues from all these poles, we obtain the result:

$$J_1(0,\eta) = \frac{-i}{1-e^{2\pi\eta}}\left\{\frac{(l\tilde{\kappa}+m\tilde{\Omega})^{N-1}e^{-i\eta\ln\frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}}}{(l\tilde{\kappa}+m\tilde{\Omega})^N+\tilde{\omega}^N} - \frac{1}{N}\sum_{j=1}^{N}\frac{e^{\frac{2j-1}{N}\pi\eta}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^{\frac{2j-1}{N}i\pi}}\right\}, \tag{C.34}$$

$$\begin{aligned} J_2(0,\eta) = -\frac{i}{1-e^{2\pi\eta}}\times&\left\{\frac{(l\tilde{\kappa}+m\tilde{\Omega})^{N-1}N\tilde{\omega}^N e^{-i\eta\ln\frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}}}{\left[(l\tilde{\kappa}+m\tilde{\Omega})^N+\tilde{\omega}^N\right]^2}\right. \\ &\left.+\frac{1}{N}\sum_{j=1}^{N}\frac{e^{\frac{2j-1}{N}\pi\eta}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^{\frac{2j-1}{N}i\pi}}\left[i\eta-\frac{\tilde{\omega}e^{\frac{2j-1}{N}i\pi}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^{\frac{2j-1}{N}i\pi}}\right]\right\}. \end{aligned} \tag{C.35}$$

For $\beta \neq 0$, we obtain the angular momentum function

$$\begin{aligned} F_{lm}(\eta\neq 0) = -\frac{i}{1-e^{2\pi\hat{\eta}}}\times&\left\{\frac{(l\tilde{\kappa}+m\tilde{\Omega})^{N-1}e^{-i\hat{\eta}\ln\frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}}}{(l\tilde{\kappa}+m\tilde{\Omega})^N+\tilde{\omega}^N}\right. \\ &\times\left[\frac{2-\beta}{2+\beta}\left((l\tilde{\kappa}+m\tilde{\Omega})\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\gamma m\right)-\frac{mN\tilde{\omega}^N}{(l\tilde{\kappa}+m\tilde{\Omega})^N+\tilde{\omega}^N}\right] \\ &-\frac{1}{N}\sum_{j=1}^{N}\frac{e^{\frac{2j-1}{N}\pi\hat{\eta}}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^{\frac{2j-1}{N}i\pi}} \\ &\left.\times\left[\frac{m\tilde{\omega}e^{\frac{2j-1}{N}i\pi}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^{\frac{2j-1}{N}i\pi}}-im\hat{\eta}+\frac{2-\beta}{2+\beta}\left((l\tilde{\kappa}+m\tilde{\Omega})\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\gamma m\right)\right]\right\}. \end{aligned} \tag{C.36}$$

We can use l'Hôpital's rule to obtain the result for $\eta = 0$. This requires that both the numerator and denominator of (C.36) are zero in the limit $\eta \to 0$. The denominator $(1-e^{2\pi\eta})$ is obviously zero in this limit, and it can be shown that the numerator is also (e.g. it is readily apparent for $N = 1$).

Application of l'Hôpital's rule then yields

$$\begin{aligned} F_{lm}(\eta=0) = -\frac{i}{2\pi}\times&\left\{\frac{i(l\tilde{\kappa}+m\tilde{\Omega})^{N-1}\ln\frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}}{(l\tilde{\kappa}+m\tilde{\Omega})^N+\tilde{\omega}^N}\right. \\ &\left[\frac{2-\beta}{2+\beta}\left((l\tilde{\kappa}+m\tilde{\Omega})\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\gamma m\right)-\frac{mN\tilde{\omega}^N}{(l\tilde{\kappa}+m\tilde{\Omega})^N+\tilde{\omega}^N}\right] \\ &+\frac{1}{N}\sum_{j=1}^{N}\frac{1}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^{\frac{2j-1}{N}i\pi}}\times \\ &\left.\left((2j-1)\frac{\pi}{N}\left[\frac{m\tilde{\omega}e^{\frac{2j-1}{N}i\pi}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^{\frac{2j-1}{N}i\pi}}+\frac{2-\beta}{2+\beta}\left((l\tilde{\kappa}+m\tilde{\Omega})\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\gamma m\right)\right]-im\right)\right\}. \end{aligned} \tag{C.37}$$

N.B. For $m = 0$, the term $l\tilde{\kappa}+m\tilde{\Omega}$ vanishes at the $l = 0$ radial harmonic. In this case, the only poles are those at $h_j = (2j-1)i\pi/N$, for which the residues are zero. Thus $F_{00} = 0$.

(3) The doubly cut-out disk

Here

$$\begin{aligned} J_1(0,\eta,\tilde{L}_{\mathrm{c}}) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-i\eta h}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^{h}}\frac{e^{Nh}}{\left[e^{Nh}+1\right]\left[\tilde{L}_{\mathrm{c}}^{-M_\beta}e^{Mh}+1\right]}dh, && \text{(C.38)}\\ J_2(0,\eta,\tilde{L}_{\mathrm{c}}) &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{e^{-i\eta h}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^{h}}\frac{Ne^{Nh}\left(\tilde{L}_{\mathrm{c}}^{-M_\beta}e^{Mh}+1\right)-M\tilde{L}_{\mathrm{c}}^{-M_\beta}e^{Nh}e^{Mh}}{\left[e^{Nh}+1\right]^2\left[\tilde{L}_{\mathrm{c}}^{-M_\beta}e^{Mh}+1\right]^2}dh. && \text{(C.39)}\end{aligned}$$

The integrand now has additional poles at

$$h_k=\frac{2k-1}{M}i\pi+\frac{2+\beta}{2-\beta}\ln\tilde{L}_{\mathrm{c}}, \tag{C.40}$$

where k is an integer. We thus obtain a second sum, and F_{lm} becomes

$$\begin{aligned} F_{lm}(\eta\neq 0) = \frac{-i}{1-e^{2\pi\hat{\eta}}}\times\Bigg\{ & \frac{(l\tilde{\kappa}+m\tilde{\Omega})^{N-1}\tilde{L}_c^{M_\beta}\tilde{\omega}^{M}e^{-i\hat{\eta}\ln\frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}}}{\left[(l\tilde{\kappa}+m\tilde{\Omega})^{N}+\tilde{\omega}^{N}\right]\left[(l\tilde{\kappa}+m\tilde{\Omega})^{M}+\tilde{L}_c^{M_\beta}\tilde{\omega}^{M}\right]}\\ & \times\left[\frac{2-\beta}{2+\beta}\left((l\tilde{\kappa}+m\tilde{\Omega})\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\gamma m\right)-\frac{mN\tilde{\omega}^{N}}{(l\tilde{\kappa}+m\tilde{\Omega})^{N}+\tilde{\omega}^{N}}+\frac{mM(l\tilde{\kappa}+m\tilde{\Omega})^{M}}{(l\tilde{\kappa}+m\tilde{\Omega})^{M}+\tilde{L}_c^{M_\beta}\tilde{\omega}^{M}}\right]\\ & -\frac{\tilde{L}_c^{M_\beta}}{N}\sum_{j=1}^{N}\frac{e^{\frac{2j-1}{N}\pi\hat{\eta}}}{\left[e^{(2j-1)\frac{M}{N}i\pi}+\tilde{L}_c^{M_\beta}\right]\left[l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^{\frac{2j-1}{N}i\pi}\right]}\\ & \times\left[\frac{m\tilde{\omega}e^{\frac{2j-1}{N}i\pi}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}e^{\frac{2j-1}{N}i\pi}}-im\hat{\eta}+\frac{2-\beta}{2+\beta}\left((l\tilde{\kappa}+m\tilde{\Omega})\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\gamma m\right)\right]\\ & +\frac{\tilde{L}_{\mathrm{c}}^{N_\beta}}{M}e^{-i\hat{\eta}\frac{2+\beta}{2-\beta}\ln\tilde{L}_{\mathrm{c}}}\sum_{k=1}^{M}\frac{e^{\frac{2k-1}{M}\pi\hat{\eta}}}{\left[e^{-(2k-1)\frac{N}{M}i\pi}+\tilde{L}_{\mathrm{c}}^{N_\beta}\right]\left[l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}\tilde{L}_{\mathrm{c}}^{\frac{2+\beta}{2-\beta}}e^{\frac{2k-1}{M}i\pi}\right]}\\ & \times\left[\frac{m\tilde{\omega}\tilde{L}_{\mathrm{c}}^{\frac{2+\beta}{2-\beta}}e^{\frac{2k-1}{M}i\pi}}{l\tilde{\kappa}+m\tilde{\Omega}-\tilde{\omega}\tilde{L}_{\mathrm{c}}^{\frac{2+\beta}{2-\beta}}e^{\frac{2k-1}{M}i\pi}}-im\hat{\eta}+\frac{2-\beta}{2+\beta}\left((l\tilde{\kappa}+m\tilde{\Omega})\left|\frac{2+2\gamma-\beta\gamma}{\beta\tilde{U}^2+\beta-2}\right|-\gamma m\right)\right]\Bigg\}. \end{aligned}\tag{C.41}$$

In the limit $\eta \to 0$:

$$
\begin{aligned}
F_{lm}(\eta = 0) = -\frac{i}{2\pi} \times \Bigg\{ & \frac{i(l\tilde{\kappa} + m\tilde{\Omega})^{N-1} \tilde{L}_c^{M_\beta} \tilde{\omega}^M \ln \frac{l\tilde{\kappa}+m\tilde{\Omega}}{\tilde{\omega}}}{\left[(l\tilde{\kappa} + m\tilde{\Omega})^N + \tilde{\omega}^N\right] \left[(l\tilde{\kappa} + m\tilde{\Omega})^M + \tilde{L}_c^{M_\beta} \tilde{\omega}^M\right]} \\
& \times \left[\frac{2-\beta}{2+\beta} \left((l\tilde{\kappa} + m\tilde{\Omega}) \left| \frac{2 + 2\gamma - \beta\gamma}{\beta\tilde{U}^2 + \beta - 2} \right| - \gamma m \right) - \frac{mN\tilde{\omega}^N}{(l\tilde{\kappa} + m\tilde{\Omega})^N + \tilde{\omega}^N} + \frac{mM(l\tilde{\kappa} + m\tilde{\Omega})^M}{(l\tilde{\kappa} + m\tilde{\Omega})^M + \tilde{L}_c^{M_\beta} \tilde{\omega}^M} \right] \\
& + \frac{\tilde{L}_c^{M_\beta}}{N} \sum_{j=1}^{N} \frac{1}{\left[e^{(2j-1)\frac{M}{N} i\pi} + \tilde{L}_c^{M_\beta}\right] \left[l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^{\frac{2j-1}{N} i\pi}\right]} \\
& \times \left[(2j-1)\frac{\pi}{N} \left\{ \frac{m\tilde{\omega} e^{\frac{2j-1}{N} i\pi}}{\left[l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} e^{\frac{2j-1}{N} i\pi}\right]} \right. \right. \\
& \qquad \left. \left. + \frac{2-\beta}{2+\beta} \left((l\tilde{\kappa} + m\tilde{\Omega}) \left| \frac{2 + 2\gamma - \beta\gamma}{\beta\tilde{U}^2 + \beta - 2} \right| - \gamma m \right) \right\} - im \right] \\
& - \frac{\tilde{L}_\mathrm{c}^{N_\beta}}{M} \sum_{k=1}^{M} \frac{1}{\left[e^{-(2k-1)\frac{N}{M} i\pi} + \tilde{L}_\mathrm{c}^{N_\beta}\right] \left[l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} \tilde{L}_\mathrm{c}^{\frac{2+\beta}{2-\beta}} e^{\frac{2k-1}{M} i\pi}\right]} \\
& \times \left[\left\{ \frac{(2k-1)\pi}{M} - i \ln \tilde{L}_\mathrm{c}^{\frac{2+\beta}{2-\beta}} \right\} \times \left\{ \frac{m\tilde{\omega} \tilde{L}_\mathrm{c}^{\frac{2+\beta}{2-\beta}} e^{\frac{2k-1}{M} i\pi}}{l\tilde{\kappa} + m\tilde{\Omega} - \tilde{\omega} \tilde{L}_\mathrm{c}^{\frac{2+\beta}{2-\beta}} e^{\frac{2k-1}{M} i\pi}} \right. \right. \\
& \qquad \left. \left. + \frac{2-\beta}{2+\beta} \left((l\tilde{\kappa} + m\tilde{\Omega}) \left| \frac{2 + 2\gamma - \beta\gamma}{\beta\tilde{U}^2 + \beta - 2} \right| - \gamma m \right) \right\} - im \right] \Bigg\}.
\end{aligned}
\tag{C.42}
$$

Numerical calculation of F_{lm}

Before we can begin to implement the equations given above, we need to consider the quantity $\ln[(l\tilde{\kappa} + m\tilde{\Omega})/\tilde{\omega}]$. As noted above, our treatment requires the principal value of this logarithm, i.e. that with imaginary part between 0 and 2π. The version of FORTRAN we used can cope with complex logarithms, but returns an answer with imaginary part between $-\pi$ and π. We must therefore add on a multiple of 2π "by hand". The imaginary part of ω is s, which is positive by assumption, so the argument of ω always lies between 0 and π. $l\tilde{\kappa} + m\tilde{\Omega}$ can be positive or negative according to the value of l, so its argument may be 0 or π. The correction necessary thus differs by π depending on the sign of $l\tilde{\kappa} + m\tilde{\Omega}$.

$$\ln \frac{l\tilde{\kappa} + m\tilde{\Omega}}{\tilde{\omega}} = \ln |l\tilde{\kappa} + m\tilde{\Omega}| - \ln \tilde{\omega} + 2i\pi, \qquad l\tilde{\kappa} + m\tilde{\Omega} > 0, \tag{C.43}$$

$$\ln \frac{l\tilde{\kappa} + m\tilde{\Omega}}{\tilde{\omega}} = \ln |l\tilde{\kappa} + m\tilde{\Omega}| - \ln \tilde{\omega} + i\pi, \qquad l\tilde{\kappa} + m\tilde{\Omega} < 0. \tag{C.44}$$

These corrections hold for both positive and negative pattern speed Ω_p, but assume that the disturbance is growing ($\mathrm{Im}(\tilde{\omega}) > 0$).

Appendix D

The Reduction Factor

In this Appendix, we derive the reduction factor $\mathcal{F}$ describing the local response of the power-law disks to a tightly-wound spiral perturbation. Increased velocity dispersion tends to stabilise a disk. A hot disk thus responds less vigorously to a given potential perturbation $\psi_{\rm imp}$ than a cold disk does. The reduction factor $\mathcal{F}$ describes the amount by which the response is reduced. $\mathcal{F}$ therefore lies between 1 (for a totally cold disk) and 0 (for an infinitely hot disk, which is unaffected by the perturbation). If we define the response of the disk in terms of the average change in radial velocity caused by the imposed perturbation, $\langle u \rangle_{\rm imp}$, then

$$\mathcal{F} = \frac{\langle u \rangle_{\rm imp}}{\langle u \rangle_{\rm imp,\ cold}}. \tag{D.1}$$

We now proceed to derive the reduction factor for a tightly-wound spiral perturbation, following the procedure of app. 6.A of Binney & Tremaine (1987). First, we write the perturbing potential as

$$\psi_{\rm imp}(R,\theta) = F(R) e^{ig(R)} e^{i(m\theta - \omega t)}. \tag{D.2}$$

This involves no loss of generality, since a general perturbation can be described by summing many such components. $F(R)$ represents the slow change in the amplitude of the potential as we move outwards through the disk; $g(R)$ represents the rapid oscillations in amplitude as we pass through successive spiral arms. The arms are separated by ΔR, where

$$g(R + \Delta R) = g(R) + 2\pi. \tag{D.3}$$

We can define a dimensional wavenumber k characterising the perturbation:

$$k(R) = \frac{2\pi}{\Delta R} = \frac{g(R + \Delta R) - g(R)}{\Delta R}. \tag{D.4}$$

We make the assumption that the perturbation is tightly-wound, i.e. ΔR is small in comparison to R.

In this case, $k = dg/dR$.[1] Therefore we rewrite eq. (D.2) as

$$\psi_{\rm imp}(R,\theta) = F(R) e^{i\int^R k(R'')dR''} e^{i(m\theta-\omega t)}. \tag{D.5}$$

If the disk is perfectly cold – that is, all the stars are initially on circular orbits – the resultant average radial velocity acquired by the stars is (Binney & Tremaine 1987, p. 360)

$$\langle u\rangle_{\rm imp,\ cold} = \frac{\varpi k \psi_{\rm imp}}{\kappa_0(1-\varpi^2)}. \tag{D.6}$$

where $\varpi = m(\Omega_{\rm p} - \Omega_0)/\kappa_0$; $\Omega_{\rm p}$ is the pattern speed and growth rate of the disturbance, and Ω_0, κ_0 are the circular and epicyclic frequencies (2.42). In a hot disk, stars already have a finite radial velocity, and so they sample different parts of the spiral potential, both positive and negative. The effect of the perturbing potential is partially cancelled out, leading to a smaller change in radial velocity. According to the definition of the reduction factor given in eq. (D.1) (substituting for $\langle u\rangle_{\rm imp,\ cold}$ from eq. (D.6) and for $\psi_{\rm imp}$ from eq. (D.5)), we have

$$\mathcal{F} = \frac{\kappa_0(1-\varpi^2)}{\varpi k(R)F(R)} e^{-i\int^R k(R'')dR''} e^{-i(m\theta-\omega t)} \langle u\rangle_{\rm imp}. \tag{D.7}$$

To evaluate the reduction factor, we need an expression for the perturbed mean radial velocity $\langle u\rangle_{\rm imp}$. By definition,

$$\langle u\rangle_{\rm imp} = \frac{\iint du\,dv\,u(f+f_{\rm imp})}{\iint du\,dv\,(f+f_{\rm imp})}. \tag{D.8}$$

In the unperturbed state, the stars are as likely to be travelling outwards as inwards, so we have $\iint fududv = 0$. Then to first order in the perturbation quantities

$$\langle u\rangle_{\rm imp} = \frac{\iint du\,dv\,u f_{\rm imp}}{\Sigma_{\rm eq}}. \tag{D.9}$$

Substituting this into (D.7), we obtain

$$\mathcal{F} = \frac{\kappa_0(1-\varpi^2)}{\varpi\Sigma_{\rm eq}} \frac{e^{-i\int^R k(R'')dR''} e^{-i(m\theta-\omega t)}}{k(R)F(R)} \iint du\,dv\,u f_{\rm imp}. \tag{D.10}$$

We now need to evaluate the perturbed distribution function $f_{\rm imp}$ for this choice of $\psi_{\rm imp}$, just as we did in Chapters 3 and 4 when the perturbing potential was a log-spiral.

---

[1] For comparison, we saw in eq. (3.24) that successive maxima of a log-spiral perturbation were separated by $\Delta R = 2\pi R/\alpha$. For a log-spiral, the function corresponding to $g(R)$ in eq. (D.2) is $\alpha \ln R$. Clearly $k = dg/dR = \alpha/R$.

The perturbed distribution function $f_{\rm imp}$ is given by eqs. (3.17) and (3.18). In order to calculate $f_{\rm imp}$, we need the derivative of the perturbation potential $\psi_{\rm imp}$ (D.5) with respect to radius and azimuth. For a tightly-wound spiral perturbation, by far the most significant change in potential occurs as we move between spiral arms. We can neglect the slow changes that occur as we move along an arm, or the gradual fall-off in amplitude with radius. We can thus neglect $\partial\psi_{\rm imp}/\partial\theta$ entirely, and approximate the radial derivative by

$$\frac{\partial\psi_{\rm imp}}{\partial R} \approx ik(R)F(R)e^{i\int^R k(R'')dR''}e^{i(m\theta-\omega t)} \tag{D.11}$$

Then (writing $k(R'')=k''$) eq. (3.17) becomes

$$f_{\rm imp}(t) = -\frac{\partial f}{\partial E}\int_{-\infty}^{t} u' ik(R')F(R')e^{i(m\theta'-\omega t')}e^{i\int^{R'}k''dR'}dt' \tag{D.12}$$

We also make the approximation

$$e^{i\int^{R'}k''dR''} \approx e^{i\left[\int^R k''dR''+k(R)(R'-R)\right]}. \tag{D.13}$$

Similarly, we approximate $F(R')\approx F(R)$ and $k(R')\approx k(R)$, and take them outside the integral. Furthermore, we pretend that the star is moving on a circular orbit so that $\theta'=\theta+\Omega_0(t'-t)$. Note that we cannot make a similar approximation and set $R'=R$ in eq. (D.13), since there the small difference $(R'-R)$ is multiplied by the wavenumber k, which by assumption is large. Then the perturbed distribution function becomes

$$\begin{aligned} f_{\rm imp}(t) = &-\frac{\partial f}{\partial E}ik(R)F(R)e^{i(m\theta-\omega t)}e^{i\int^R k''dR''} \\ &\times\int_{-\infty}^{t} u' e^{i(m\Omega_0(t'-t)-\omega(t'-t))}e^{ik(R)(R'-R)}dt'. \end{aligned} \tag{D.14}$$

This is analogous to the perturbed distribution functions derived for log-spiral perturbations in Chapters 3 and 4. There, we obtained the unperturbed orbit of the star by numerically solving the equations of motion (2.10). Here, we use epicyclic approximations to the radial position R and the radial velocity u. The epicyclic orbit equations are (Binney & Tremaine 1987, p. 691):

$$R' = R_g + X\cos\kappa_0 t' - Y\sin\kappa_0 t', \tag{D.15}$$

$$u' = -\kappa_0 X\sin\kappa_0 t' - \kappa_0 Y\cos\kappa_0 t', \tag{D.16}$$

$$v' - v_{\rm circ}(R') = 2BX\cos\kappa_0 t' - 2BY\sin\kappa_0 t', \tag{D.17}$$

where the guiding centre R_g and amplitudes X and Y are determined by the boundary condition that at time $t' = t$, we have $R' = R$ and $u' = u$. B is Oort's parameter (Binney & Tremaine 1987, eq. (6-27)):

$$B = -\left(\frac{R}{2}\frac{d\Omega_0}{dR} + \Omega_0\right). \tag{D.18}$$

Defining $\chi = \kappa_0(t' - t)$, the epicyclic approximations to R' and u' are

$$R' - R = \frac{v - v_{\rm circ}}{2B}(\cos\chi - 1) + \frac{u}{\kappa_0}\sin\chi, \tag{D.19}$$

$$u' = u\cos\chi - \frac{\kappa_0(v - v_{\rm circ})}{2B}\sin\chi. \tag{D.20}$$

Substituting these approximations into the perturbed distribution function, eq. (D.14), and then into the expression for the reduction factor, eq. (D.10), we obtain

$$\begin{aligned}\mathcal{F} = &-\frac{i(1-\varpi^2)}{\varpi\Sigma_{\rm eq}}\int_{-\infty}^{\infty} du\, u \int_{-\infty}^{\infty} dv\, \frac{\partial f}{\partial E}\int_{-\infty}^{0} e^{-i\varpi\chi} \\ &\times \left\{u\cos\chi - \frac{\kappa_0(v - v_{\rm circ})}{2B}\sin\chi\right\}\exp ik\left\{\frac{(v - v_{\rm circ})}{2B}(\cos\chi - 1) + \frac{u}{\kappa_0}\sin\chi\right\} d\chi.\end{aligned} \tag{D.21}$$

(We delay substituting in the equilibrium density $\Sigma_{\rm eq}$ and the derivative of the equilibrium distribution function $\partial f/\partial E$, so as to keep the derivation fully general for as long as possible.) Changing variables from χ to $-\chi$, we rewrite eq. (D.21) as

$$\begin{aligned}\mathcal{F} = &-\frac{i(1-\varpi^2)}{\varpi\Sigma_{\rm eq}}\int_{-\infty}^{\infty} du\, u \int_{-\infty}^{\infty} dv\, \frac{\partial f}{\partial E}\int_{0}^{\infty} d\chi e^{i\varpi\chi} \\ &\times \left\{u\cos\chi + \frac{\kappa_0(v - v_{\rm circ})}{2B}\sin\chi\right\}\exp ik\left\{\frac{(v - v_{\rm circ})}{2B}(\cos\chi - 1) - \frac{u}{\kappa_0}\sin\chi\right\}.\end{aligned} \tag{D.22}$$

The infinite integral over χ is periodic, and can therefore be split up into a geometric sum of integrals over an interval of 2π. We then make the further adjustment $\chi \to \chi - \pi$:

$$\begin{aligned}\mathcal{F} = &-\frac{(1-\varpi^2)}{2\varpi\sin\pi\varpi}\frac{1}{\Sigma_{\rm eq}}\int_{-\infty}^{\infty} du\, u \int_{-\infty}^{\infty} dv\, \frac{\partial f}{\partial E}\int_{-\pi}^{\pi} d\chi e^{i\varpi\chi} \\ &\times \left\{u\cos\chi + \frac{\kappa_0(v - v_{\rm circ})}{2B}\sin\chi\right\}\exp\left[\frac{ik}{\kappa_0}\left\{u\sin\chi - \frac{\kappa_0(v - v_{\rm circ})}{2B}(\cos\chi + 1)\right\}\right].\end{aligned} \tag{D.23}$$

We notice that the first term in curly brackets is the derivative with respect to χ of the second such term. We can therefore integrate by parts to obtain

$$\begin{aligned}\mathcal{F} = &-\frac{(1-\varpi^2)}{2\varpi\sin\pi\varpi}\frac{1}{\Sigma_{\rm eq}}\int_{-\infty}^{\infty} du\, u \int_{-\infty}^{\infty} dv\, \frac{\partial f}{\partial E} \\ &\times \left\{\frac{2\kappa_0}{k}\sin\pi\varpi - \frac{\kappa_0\varpi}{k}\int_{-\pi}^{\pi} d\chi e^{i\varpi\chi}\exp\left[\frac{ik}{\kappa_0}\left\{u\sin\chi - \frac{\kappa_0(v - v_{\rm circ})}{2B}(\cos\chi + 1)\right\}\right]\right\}.\end{aligned} \tag{D.24}$$

We now consider the integration over u. Clearly only terms that are even in u will survive the integration. The derivative $\partial f/\partial E$ is even in u (since f is built out of isolating integrals), so on the second line of eq. (D.24) we need retain only terms which are odd in u.

$$\begin{aligned}\mathcal{F} = {} & i\frac{(1-\varpi^2)}{\sin\pi\varpi}\frac{1}{\Sigma_{\rm eq}}\frac{\kappa_0}{k}\int_{-\pi}^{\pi} d\chi e^{i\varpi\chi} \\ & \times\int_0^\infty du\, u\sin\left[\frac{ku}{\kappa_0}\sin\chi\right]\int_{-\infty}^{\infty} dv\,\frac{\partial f}{\partial E}\exp\left[-\frac{ik(v-v_{\rm circ})}{2B}(\cos\chi+1)\right].\end{aligned} \tag{D.25}$$

We now carry out a similar procedure, abandoning terms which are odd in χ.

$$\begin{aligned}\mathcal{F} = {} & -\frac{(1-\varpi^2)}{\sin\pi\varpi}\frac{2\kappa_0}{k\Sigma_{\rm eq}}\int_0^{\pi} d\chi\sin\varpi\chi \\ & \times\int_0^\infty du\, u\sin\left[\frac{ku}{\kappa_0}\sin\chi\right]\int_{-\infty}^{\infty} dv\,\frac{\partial f}{\partial E}\exp\left[-\frac{ik(v-v_{\rm circ})}{2B}(\cos\chi+1)\right].\end{aligned} \tag{D.26}$$

Our working so far has been completely general in that we have not assumed an equilibrium surface density or distribution function for our disk. Kalnajs (1965) and Lin & Shu (1966) derived the reduction factor for the Schwarzschild distribution function

$$f_{\blacktriangledown} = \frac{\Sigma_{\rm eq}}{2\pi\sigma_u\sigma_v}\exp\left[-\frac{u^2}{2\sigma_u^2}-\frac{(v-v_{\rm circ})^2}{2\sigma_v^2}\right]. \tag{D.27}$$

In this case, the integration over u and v can be carried out analytically. There is no dependence on surface density, so the results are valid for any density profile, provided appropriate values are substituted for κ_0, $v_{\rm circ}$ and B. In general, the Schwarzschild reduction factor is (Lin & Shu 1966; Binney & Tremaine 1987, app. 6.A)

$$\mathcal{F}_{\blacktriangledown} = \frac{(1-\varpi^2)}{\sin\pi\varpi}\int_0^{\pi} d\chi e^{-\zeta(1+\cos\chi)}\sin\varpi\chi\sin\chi, \tag{D.28}$$

where $\zeta = (k\sigma_u/\kappa_0)^2$. For the power-law disks, $\zeta = \alpha^2\tilde{\sigma}_u^2/(2-\beta)$.

To find the form of the reduction factor for our power-law distribution function (eqs. (2.54), (2.81)), we return to our general expression (D.26) and specialise to the case of the power-law disks. We rewrite the integrals in terms of the dimensionless velocities $\tilde{u}$ and $\tilde{v}$. We are using the epicyclic approximation, in which the star is roughly at its home radius. Under these circumstances, $\tilde{u} = u/v_{\rm circ}$ and $\tilde{v} = v/v_{\rm circ}$ (see eq. (2.20)). We also have $kv_{\rm circ}/\kappa_0 = \alpha/\sqrt{2-\beta}$ and $-kv_{\rm circ}/2B = 2\alpha/(2-\beta)$. Then eq. (D.26) becomes

$$\begin{aligned}\mathcal{F} = {} & -\frac{(1-\varpi^2)}{\sin\pi\varpi}\frac{2\sqrt{2-\beta}}{\alpha}\frac{v_\beta^4}{\Sigma_0}\left(\frac{R}{R_0}\right)^{1-\beta}\int_0^{\pi} d\chi\sin\varpi\chi \\ & \times\int_0^\infty d\tilde{u}\,\tilde{u}\sin\left[\frac{\alpha\tilde{u}\sin\chi}{\sqrt{2-\beta}}\right]\int_{-\infty}^{\infty} d\tilde{v}\,\exp\left[\frac{2i\alpha(\tilde{v}-1)}{2-\beta}(\cos\chi+1)\right]\frac{\partial f}{\partial E}\end{aligned} \tag{D.29}$$

We substitute for $\partial f/\partial E$ from eq. (3.61), and for the constant $\tilde{C}$ from eq. (2.55). We then obtain

$$\boxed{\begin{aligned}\mathcal{F} = {} & \frac{(1-\varpi^2)}{\sin\pi\varpi}\left|\frac{1}{\beta}+\frac{\gamma}{\beta}-\frac{\gamma}{2}\right|\frac{C_{\beta\gamma}\sqrt{2-\beta}|\beta|^{1+\frac{1}{\beta}+\frac{\gamma}{\beta}}}{2^{\gamma/2-1}\alpha\sqrt{\pi}}\int_0^{\pi} d\chi\sin\varpi\chi \\ & \times\int_0^\infty d\tilde{u}\,\tilde{u}\sin\left[\frac{\alpha}{\sqrt{2-\beta}}\tilde{u}\sin\chi\right]\int_{-\infty}^{\infty} d\tilde{v}\,\exp\left[\frac{2i\alpha(\tilde{v}-1)}{2-\beta}(\cos\chi+1)\right] \\ & \times\tilde{v}^{\gamma}\tilde{H}\left[\tilde{v}\left(\frac{R}{R_0}\right)^{\frac{2-\beta}{2}}\right]\left|\frac{\tilde{u}^2}{2}+\frac{\tilde{v}^2}{2}-\frac{1}{\beta}\right|^{1/\beta+\gamma/\beta-\gamma/2-1}\end{aligned}} \tag{D.30}$$

Numerical methods

To evaluate the reduction factor (D.30) for the power-law disks, the integration over velocities must be carried out numerically. We use Gauss-Hermite integration over $\tilde{u}$. This is suitable for integrands containing a factor $\exp(-x^2)$:

$$\int_{-\infty}^{\infty} e^{-x^2} f(x)dx = \sum_{i=1}^{n_{\mathrm{H}}} w_i f(x_i). \tag{D.31}$$

where the order of Gauss-Hermite integration is n_{H}. If we substitute $\tilde{u} = x\tilde{\sigma}_u\sqrt{2}$ in eq. (D.31), and write $g(\tilde{u}) = e^{-\tilde{u}^2/2\tilde{\sigma}_u^2} f\left(\frac{\tilde{u}}{\tilde{\sigma}_u\sqrt{2}}\right)$, then we obtain

$$\int_{-\infty}^{\infty} g(\tilde{u})d\tilde{u} = \tilde{\sigma}_u\sqrt{2}\sum_{i=1}^{n_{\mathrm{H}}} w_i e^{x_i^2} g(x_i\tilde{\sigma}_u\sqrt{2}). \tag{D.32}$$

Abscissae on either side of the origin have the same weights, and our integrand $g(\tilde{u})$ is even. We therefore need to use abscissae on one side of the origin only.

We used the midpoint method to integrate over $\tilde{v}$ and χ. When evaluating the Schwarzschild reduction factor, we again used the midpoint method to integrate over χ.

The following table shows the values obtained for the reduction factor using different numbers of Gauss-Hermite abscissae to integrate over $\tilde{u}$. In each case, convergence to 5 s.f. was demanded from the midpoint integration over v, and to 4 s.f. from the final midpoint integration over χ. The last row shows the Schwarzschild reduction factor.

		$\beta = 0.25$, $\alpha = 20$					
		$\varpi = 0.25$		$\varpi = 0.50$		$\varpi = 0.75$	
	n_{H}	$Q_{\mathrm{s}} = 1$	$Q_{\mathrm{s}} = 0.5$	$Q_{\mathrm{s}} = 1$	$Q_{\mathrm{s}} = 0.5$	$Q_{\mathrm{s}} = 1$	$Q_{\mathrm{s}} = 0.5$
Power-law d.f. eq. (2.54)	6	0.05389	0.1711	0.04673	0.1543	0.03407	0.1238
	7	0.04305	0.1703	0.03642	0.1536	0.02483	0.1231
	8	0.04780	0.1704	0.04098	0.1537	0.02899	0.1231
	9	0.04598	0.1704	0.03921	0.1537	0.02735	0.1231
	10	0.04658	0.1704	0.03980	0.1537	0.02790	0.1231
	12	0.04645	0.1704	0.03967	0.1537	0.02778	0.1231
	16	0.04645	0.1704	0.03967	0.1537	0.02778	0.1231
Schwarzschild d.f. eq. (D.27)		0.04728	0.1711	0.04041	0.1544	0.02837	0.1238

The Schwarzschild and power-law distribution functions give the same value for the reduction factor, to within 2 s.f. (see also fig. 7.1). Gauss-Hermite quadrature works well in carrying out the integration over $\tilde{u}$. For $Q_{\mathrm{s}} = 1$, as many as twelve Gauss-Hermite points are required for convergence to 4 s.f. The convergence is faster for colder disks.

Appendix E

Tables of Marginal Modes

In this Appendix, tables of data are presented summarising the marginal modes of the power-law disks. These are the modes with vanishing growth rate which occur at the minimum temperature necessary for stability.

E.1 Axisymmetric modes

The minimum temperature necessary for stability is given in terms of the anisotropy parameter, γ_{max}, the velocity dispersion $\tilde{\sigma}_{u,\mathrm{min}}$ and Toomre's stability criterion Q_{s} (see eq. (6.15)). For the singular disk, the most unstable wavenumber α_{crit} is also given.

The singular disk

This table contains the data used in figs. 6.12, and 6.13. Q_{s} gives the ratio of the critical velocity dispersion obtained by global theory to that obtained by local theory.

	γ_{max}		$\tilde{\sigma}_{u,\mathrm{min}}$		Q_{s}	α_{crit}	
β	Local	Global	Local	Global		Local	Global
-0.70	1.13000	0.673755	1.17041	1.91126	1.63298	1.367	1.30046
-0.65	1.26362	0.845274	1.01870	1.35423	1.32937	1.556	1.44040
-0.60	1.42444	1.04331	0.903715	1.08895	1.20497	1.737	1.60151
-0.55	1.61221	1.26754	0.813194	0.925476	1.13807	1.912	1.76973
-0.50	1.82724	1.51765	0.739779	0.811735	1.09727	2.081	1.93529
-0.45	2.07039	1.79432	0.678783	0.726563	1.07039	2.245	2.09530
-0.40	2.34306	2.09907	0.627078	0.659514	1.05173	2.406	2.25040
-0.35	2.64726	2.43415	0.582493	0.604768	1.03824	2.563	2.40195
-0.30	2.98566	2.80255	0.543474	0.558795	1.02819	2.717	2.55121
-0.25	3.36167	3.20803	0.508877	0.519312	1.02051	2.870	2.69932
-0.20	3.77962	3.65523	0.477839	0.484773	1.01451	3.022	2.84737
-0.15	4.24490	4.14980	0.449698	0.454086	1.00976	3.175	3.00000
-0.10	4.76425	4.69871	0.423932	0.426452	1.00594	3.328	3.14749
-0.05	5.34608	5.31057	0.400125	0.401268	1.00285	3.484	3.30179
0.00	6.00092	5.99611	0.377940	0.378069	1.00034	3.643	3.46058
0.05	6.74208	6.76881	0.357096	0.356488	0.998300	3.808	3.62528
0.10	7.58651	7.64579	0.337359	0.336226	0.996643	3.978	3.79760
0.15	8.55605	8.64905	0.318529	0.317036	0.995315	4.158	3.97956
0.20	9.67922	9.80727	0.300431	0.298710	0.994271	4.348	4.17363

	$\gamma_{\max}$		$\tilde{\sigma}_{u,\min}$		Q_{s}	α_{crit}	
β	Local	Global	Local	Global		Local	Global
0.25	10.9938	11.1584	0.282912	0.281067	0.993478	4.553	4.38289
0.30	12.5509	12.7536	0.265832	0.263949	0.992914	4.776	4.61125
0.35	14.4207	14.6634	0.249063	0.247208	0.992555	5.022	4.86380
0.40	16.7022	16.9870	0.232481	0.230713	0.992393	5.298	5.14730
0.45	19.5393	19.8680	0.215971	0.214334	0.992421	5.613	5.47080
0.50	23.1470	23.5226	0.199415	0.197942	0.992614	5.980	5.84720
0.55	27.8603	28.2853	0.182695	0.181413	0.992983	6.418	6.29490
0.60	34.2263	34.7035	0.165689	0.164614	0.993513	6.953	6.84143
0.65	43.1908	43.7235	0.148265	0.147404	0.994196	7.630	7.53027
0.70	56.5158	57.1080	0.130282	0.129632	0.995012	8.521	8.43370
0.75	77.8175	78.4730	0.111582	0.111130	0.995944	9.756	9.70000

The cut-out disks

The following tables contain the data used in fig. 6.17. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.1$, $l_{\min} = 40$, $l_{\max} = 50$, $n_{\mathrm{GL}} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20.0$, $b_{\mathrm{acc}} = 1.5$, $\epsilon_\lambda = 10^{-6}$.]

Inner cut-out index $N = 1$

	Inner cut-out disk			Doubly cut-out disk with $N = 1$, $\tilde{R}_{\mathrm{c}} = 100$								
	$N = 1$			$M = 2$			$M = 4$			$M = 6$		
β	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Q_{s}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Q_{s}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Q_{s}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Q_{s}
-0.50	1.554	0.802	1.084	2.812	0.596	0.806	2.404	0.645	0.872	2.290	0.661	0.893
-0.45	1.834	0.719	1.059	3.210	0.550	0.810	2.766	0.591	0.870	2.643	0.604	0.889
-0.40	2.142	0.653	1.042	3.630	0.511	0.815	3.150	0.546	0.871	3.020	0.557	0.889
-0.35	2.480	0.600	1.030	4.074	0.478	0.821	3.560	0.509	0.874	3.422	0.518	0.890
-0.30	2.852	0.555	1.020	4.548	0.450	0.827	4.000	0.477	0.877	3.855	0.485	0.892
-0.25	3.260	0.516	1.013	5.055	0.424	0.834	4.473	0.448	0.881	4.322	0.455	0.895
-0.20	3.710	0.482	1.008	5.601	0.402	0.840	4.986	0.423	0.885	4.828	0.429	0.898
-0.15	4.208	0.451	1.004	6.192	0.381	0.847	5.545	0.400	0.890	5.380	0.406	0.902
-0.10	4.759	0.424	1.000	6.838	0.362	0.854	6.157	0.379	0.894	5.985	0.384	0.906
-0.05	5.374	0.399	0.998	7.547	0.344	0.860	6.832	0.360	0.899	6.654	0.364	0.909
0.00	6.063	0.376	0.996	8.331	0.327	0.866	7.581	0.341	0.903	7.396	0.345	0.913
0.05	6.840	0.355	0.994	9.206	0.311	0.872	8.419	0.324	0.908	8.227	0.327	0.917
0.10	7.721	0.335	0.992	10.190	0.296	0.878	9.364	0.308	0.912	9.164	0.311	0.921
0.15	8.729	0.316	0.991	11.308	0.282	0.884	10.439	0.292	0.916	10.231	0.294	0.925
0.20	9.891	0.298	0.991	12.591	0.267	0.890	11.675	0.277	0.921	11.457	0.279	0.928
0.25	11.248	0.280	0.990	14.080	0.253	0.895	13.110	0.262	0.925	12.882	0.264	0.932
0.30	12.851	0.263	0.990	15.831	0.240	0.901	14.799	0.247	0.929	14.559	0.249	0.936
0.35	14.769	0.246	0.989	17.918	0.226	0.906	16.815	0.232	0.933	16.561	0.234	0.940
0.40	17.102	0.230	0.989	20.449	0.212	0.912	19.261	0.218	0.937	18.990	0.219	0.943
0.45	19.994	0.214	0.990	23.578	0.198	0.917	22.286	0.203	0.941	21.994	0.205	0.947
0.50	23.661	0.197	0.990	27.534	0.184	0.923	26.113	0.189	0.946	25.795	0.190	0.951

Inner cut-out index $N = 2$

	Inner cut-out disk			Doubly cut-out disk with $N = 2$, $\tilde{R}_{\mathrm{c}} = 100$								
	$N = 2$			$M = 2$			$M = 4$			$M = 6$		
β	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Q_{s}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Q_{s}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Q_{s}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Q_{s}
-0.50	1.540	0.806	1.089	2.158	0.681	0.920	1.910	0.724	0.978	1.848	0.736	0.994
-0.45	1.819	0.722	1.064	2.489	0.622	0.916	2.222	0.656	0.967	2.158	0.666	0.980
-0.40	2.126	0.656	1.046	2.841	0.573	0.914	2.559	0.602	0.960	2.492	0.609	0.972
-0.35	2.463	0.602	1.033	3.221	0.533	0.915	2.924	0.557	0.956	2.854	0.563	0.967
-0.30	2.833	0.556	1.023	3.632	0.498	0.916	3.320	0.518	0.954	3.248	0.524	0.963
-0.25	3.241	0.517	1.016	4.077	0.467	0.919	3.751	0.485	0.953	3.677	0.489	0.962
-0.20	3.690	0.483	1.010	4.561	0.440	0.921	4.222	0.455	0.953	4.147	0.459	0.961
-0.15	4.187	0.452	1.006	5.092	0.416	0.924	4.739	0.429	0.953	4.662	0.432	0.960
-0.10	4.738	0.425	1.002	5.675	0.393	0.927	5.312	0.405	0.954	5.231	0.407	0.960
-0.05	5.352	0.400	1.000	6.321	0.372	0.930	5.946	0.382	0.955	5.864	0.384	0.961
0.00	6.040	0.377	0.997	7.041	0.353	0.933	6.653	0.361	0.956	6.570	0.363	0.962
0.05	6.815	0.355	0.995	7.849	0.334	0.936	7.448	0.342	0.958	7.363	0.344	0.963
0.10	7.695	0.335	0.994	8.762	0.317	0.939	8.348	0.324	0.959	8.261	0.325	0.964
0.15	8.701	0.316	0.993	9.803	0.300	0.942	9.375	0.306	0.961	9.286	0.307	0.965
0.20	9.862	0.298	0.992	11.002	0.284	0.945	10.558	0.289	0.963	10.467	0.290	0.966
0.25	11.215	0.280	0.991	12.397	0.268	0.948	11.936	0.273	0.964	11.842	0.274	0.968
0.30	12.815	0.263	0.991	14.041	0.253	0.951	13.560	0.257	0.966	13.463	0.258	0.969
0.35	14.730	0.247	0.991	16.006	0.238	0.954	15.501	0.241	0.968	15.401	0.242	0.971
0.40	17.059	0.230	0.990	18.393	0.223	0.957	17.860	0.226	0.970	17.756	0.226	0.973
0.45	19.947	0.214	0.991	21.348	0.207	0.960	20.782	0.210	0.972	20.672	0.210	0.975
0.50	23.609	0.198	0.991	25.090	0.192	0.963	24.482	0.194	0.974	24.366	0.195	0.977

Inner cut-out index $N = 3$

	Inner cut-out disk			Doubly cut-out disk with $N = 3$, $\tilde{R}_{\mathrm{c}} = 100$								
	$N = 3$			$M = 2$			$M = 4$			$M = 6$		
β	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Q_{s}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Q_{s}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Q_{s}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Q_{s}
-0.50	1.537	0.807	1.090	1.989	0.709	0.959	1.798	0.746	1.008	1.754	0.755	1.021
-0.45	1.816	0.722	1.064	2.305	0.645	0.950	2.102	0.674	0.993	2.057	0.681	1.003
-0.40	2.123	0.656	1.046	2.645	0.593	0.945	2.433	0.616	0.983	2.384	0.622	0.992
-0.35	2.460	0.602	1.033	3.014	0.549	0.943	2.789	0.569	0.977	2.739	0.574	0.985
-0.30	2.830	0.556	1.024	3.413	0.512	0.942	3.177	0.529	0.973	3.126	0.533	0.980
-0.25	3.237	0.517	1.017	3.847	0.480	0.943	3.602	0.494	0.970	3.550	0.497	0.976
-0.20	3.686	0.483	1.011	4.321	0.451	0.943	4.069	0.463	0.969	4.015	0.465	0.974
-0.15	4.183	0.453	1.006	4.842	0.425	0.945	4.582	0.435	0.968	4.526	0.437	0.973
-0.10	4.734	0.425	1.003	5.416	0.401	0.946	5.147	0.410	0.967	5.091	0.412	0.972

	Inner cut-out disk			Doubly cut-out disk with $N = 3$, $\tilde{R}_{\mathrm{c}} = 100$								
	$N = 3$			$M = 2$			$M = 4$			$M = 6$		
β	γ_{max}	$\tilde{\sigma}_{u,\mathrm{min}}$	Q_{s}	γ_{max}	$\tilde{\sigma}_{u,\mathrm{min}}$	Q_{s}	γ_{max}	$\tilde{\sigma}_{u,\mathrm{min}}$	Q_{s}	γ_{max}	$\tilde{\sigma}_{u,\mathrm{min}}$	Q_{s}
-0.05	5.347	0.400	1.000	6.053	0.379	0.948	5.774	0.387	0.967	5.718	0.389	0.972
0.00	6.035	0.377	0.998	6.763	0.359	0.950	6.478	0.366	0.968	6.420	0.367	0.971
0.05	6.810	0.356	0.996	7.560	0.340	0.952	7.267	0.346	0.968	7.208	0.347	0.972
0.10	7.690	0.335	0.994	8.463	0.322	0.954	8.163	0.327	0.969	8.102	0.328	0.972
0.15	8.696	0.316	0.993	9.493	0.304	0.956	9.182	0.309	0.970	9.122	0.310	0.972
0.20	9.857	0.298	0.992	10.679	0.288	0.958	10.359	0.292	0.971	10.298	0.292	0.973
0.25	11.209	0.281	0.991	12.061	0.272	0.960	11.730	0.275	0.972	11.666	0.276	0.974
0.30	12.809	0.263	0.991	13.690	0.256	0.962	13.346	0.259	0.973	13.280	0.259	0.975
0.35	14.723	0.247	0.991	15.637	0.240	0.964	15.278	0.243	0.974	15.210	0.243	0.976
0.40	17.052	0.230	0.991	18.002	0.225	0.967	17.627	0.227	0.976	17.556	0.227	0.978
0.45	19.939	0.214	0.991	20.931	0.209	0.969	20.535	0.211	0.978	20.461	0.211	0.979
0.50	23.600	0.198	0.991	24.641	0.194	0.972	24.221	0.195	0.979	24.142	0.196	0.981

Inner cut-out index $N = 4$

	Inner cut-out disk			Doubly cut-out disk with $N = 4$, $\tilde{R}_{\mathrm{c}} = 100$								
	$N = 4$			$M = 2$			$M = 4$			$M = 6$		
β	γ_{max}	$\tilde{\sigma}_{u,\mathrm{min}}$	Q_{s}	γ_{max}	$\tilde{\sigma}_{u,\mathrm{min}}$	Q_{s}	γ_{max}	$\tilde{\sigma}_{u,\mathrm{min}}$	Q_{s}	γ_{max}	$\tilde{\sigma}_{u,\mathrm{min}}$	Q_{s}
-0.50	1.536	0.807	1.091	1.921	0.722	0.975	1.757	0.754	1.020	1.723	0.762	1.030
-0.45	1.815	0.723	1.065	2.232	0.655	0.965	2.061	0.680	1.002	2.023	0.686	1.011
-0.40	2.122	0.656	1.047	2.569	0.601	0.958	2.387	0.622	0.991	2.345	0.627	1.000
-0.35	2.459	0.602	1.034	2.934	0.556	0.955	2.740	0.574	0.985	2.698	0.578	0.992
-0.30	2.829	0.557	1.024	3.330	0.518	0.953	3.126	0.533	0.980	3.084	0.536	0.986
-0.25	3.236	0.517	1.017	3.760	0.485	0.952	3.550	0.497	0.976	3.507	0.500	0.982
-0.20	3.685	0.483	1.011	4.230	0.455	0.952	4.016	0.465	0.974	3.970	0.468	0.979
-0.15	4.181	0.453	1.006	4.748	0.428	0.953	4.526	0.437	0.973	4.478	0.439	0.977
-0.10	4.732	0.425	1.003	5.319	0.404	0.954	5.089	0.412	0.972	5.042	0.414	0.976
-0.05	5.346	0.400	1.000	5.953	0.382	0.955	5.715	0.389	0.972	5.668	0.390	0.975
0.00	6.033	0.377	0.998	6.660	0.361	0.956	6.417	0.367	0.972	6.369	0.368	0.975
0.05	6.808	0.356	0.996	7.455	0.342	0.957	7.205	0.347	0.972	7.155	0.348	0.975
0.10	7.688	0.335	0.994	8.354	0.324	0.959	8.098	0.328	0.972	8.047	0.329	0.975
0.15	8.694	0.316	0.993	9.381	0.306	0.961	9.118	0.310	0.973	9.065	0.311	0.975
0.20	9.854	0.298	0.992	10.564	0.289	0.962	10.293	0.292	0.973	10.240	0.293	0.976
0.25	11.207	0.281	0.992	11.941	0.273	0.964	11.660	0.276	0.974	11.607	0.276	0.976
0.30	12.806	0.263	0.991	13.565	0.257	0.966	13.273	0.259	0.975	13.220	0.260	0.977
0.35	14.720	0.247	0.991	15.506	0.241	0.968	15.203	0.243	0.977	15.146	0.244	0.978
0.40	17.049	0.230	0.991	17.865	0.226	0.970	17.548	0.227	0.978	17.488	0.228	0.979
0.45	19.936	0.214	0.991	20.786	0.210	0.972	20.452	0.212	0.979	20.390	0.212	0.981
0.50	23.596	0.198	0.991	24.486	0.194	0.974	24.133	0.196	0.981	24.067	0.196	0.982

E.2 Bisymmetric modes

The following tables contain the data used in the graphs of minimum temperature presented in Chapter 7, figs. 7.26 and 7.27. The temperature necessary for stability is given in terms of the anisotropy parameter $\gamma_{\max}$ and the velocity dispersion $\tilde{\sigma}_{u,\min}$. For these bisymmetric modes, we also need to specify the pattern speed Ω_{p}. [Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\min} = 30$, $l_{\max} = 40$, $n_{\mathrm{GL}} = 9$, $f_\sigma = 0.8$, $a_{\mathrm{acc}} = 20.0$, $b_{\mathrm{acc}} = 2.0$, $\epsilon_\lambda = 10^{-6}$.]

Inner cut-out index $N = 2$

	Inner cut-out disk			Doubly cut-out disk with $N = 2$, $\tilde{R}_{\mathrm{c}} = 10$								
	$N = 2$			$M = 2$			$M = 4$			$M = 6$		
β	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Ω_{p}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Ω_{p}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Ω_{p}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Ω_{p}
-0.50	1.48	0.821	0.420	2.07	0.694	0.453	1.54	0.806	0.425	1.44	0.833	0.418
-0.45	2.48	0.622	0.451	3.40	0.534	0.487	2.50	0.620	0.453	2.47	0.624	0.451
-0.40	3.95	0.491	0.482	4.96	0.440	0.513	3.96	0.491	0.483	3.94	0.491	0.482
-0.35	5.62	0.411	0.505	6.53	0.383	0.531	5.61	0.411	0.505	5.61	0.411	0.505
-0.30	7.35	0.359	0.521	8.16	0.342	0.543	7.34	0.360	0.521	7.35	0.359	0.521
-0.25	9.19	0.321	0.533	9.92	0.310	0.552	9.17	0.322	0.533	9.19	0.321	0.533
-0.20	11.2	0.291	0.543	11.9	0.283	0.559	11.2	0.291	0.543	11.2	0.291	0.543
-0.15	13.4	0.266	0.551	14.1	0.260	0.566	13.4	0.266	0.551	13.4	0.266	0.551
-0.10	16.0	0.244	0.559	16.5	0.240	0.571	16.0	0.244	0.559	16.0	0.244	0.559
-0.05	18.8	0.225	0.565	19.4	0.222	0.577	18.8	0.225	0.565	18.8	0.225	0.565
0.00	22.2	0.208	0.572	22.7	0.205	0.582	22.2	0.208	0.571	22.2	0.208	0.571
0.05	25.9	0.193	0.577	26.4	0.191	0.586	25.9	0.193	0.577	25.9	0.193	0.577
0.10	30.2	0.178	0.582	30.7	0.177	0.591	30.2	0.178	0.582	30.2	0.178	0.582
0.15	35.3	0.165	0.587	35.8	0.164	0.595	35.3	0.165	0.587	35.3	0.165	0.587
0.20	41.3	0.153	0.592	41.8	0.152	0.599	41.2	0.153	0.591	41.3	0.153	0.592
0.25	48.4	0.142	0.596	48.8	0.141	0.602	48.4	0.141	0.597	48.4	0.142	0.596
0.30	57.0	0.131	0.601	57.4	0.130	0.606	57.0	0.131	0.601	56.9	0.131	0.600
0.35	67.4	0.120	0.605	67.9	0.120	0.610	67.4	0.120	0.605	67.4	0.120	0.606
0.40	80.4	0.110	0.609	80.7	0.110	0.614	80.3	0.110	0.609	80.0	0.111	0.609
0.45	96.4	0.101	0.614	96.8	0.101	0.618	96.4	0.101	0.614	96.0	0.101	0.612
0.50	117.	0.092	0.621	118.	0.091	0.624	117.	0.092	0.621	117.	0.092	0.622

Inner cut-out index $N = 3$

	Inner cut-out disk			Doubly cut-out disk with $N = 3$, $\tilde{R}_{\mathrm{c}} = 10$								
	$N = 3$			$M = 2$			$M = 4$			$M = 6$		
β	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Ω_{p}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Ω_{p}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Ω_{p}	$\gamma_{\max}$	$\tilde{\sigma}_{u,\min}$	Ω_{p}
-0.40	0.428	1.262	0.313	0.521	1.178	0.320	0.450	1.241	0.314	0.450	1.240	0.315
-0.35	0.589	1.060	0.323	0.700	1.000	0.330	0.617	1.044	0.325	0.610	1.048	0.325

	Inner cut-out disk			Doubly cut-out disk with $N = 3$, $\tilde{R}_\mathrm{c} = 10$								
	$N = 3$			$M = 2$			$M = 4$			$M = 6$		
β	γ_max	$\tilde{\sigma}_{u,\mathrm{min}}$	Ω_p	γ_max	$\tilde{\sigma}_{u,\mathrm{min}}$	Ω_p	γ_max	$\tilde{\sigma}_{u,\mathrm{min}}$	Ω_p	γ_max	$\tilde{\sigma}_{u,\mathrm{min}}$	Ω_p
-0.30	0.811	0.909	0.334	0.940	0.864	0.341	0.839	0.898	0.336	0.824	0.904	0.336
-0.25	1.10	0.791	0.346	1.25	0.756	0.353	1.12	0.785	0.348	1.10	0.790	0.347
-0.20	1.47	0.694	0.358	1.66	0.666	0.365	1.50	0.691	0.360	1.47	0.695	0.359
-0.15	1.97	0.612	0.372	2.19	0.588	0.378	1.99	0.610	0.373	1.97	0.612	0.372
-0.10	2.64	0.539	0.386	2.90	0.520	0.392	2.65	0.538	0.386	2.63	0.540	0.386
-0.05	3.53	0.475	0.401	3.86	0.458	0.407	3.54	0.474	0.401	3.53	0.475	0.401
0.00	4.77	0.416	0.416	5.17	0.402	0.422	4.78	0.416	0.417	4.77	0.416	0.416
0.05	6.38	0.366	0.432	6.86	0.354	0.438	6.39	0.366	0.432	6.38	0.366	0.432
0.10	8.56	0.320	0.449	9.15	0.311	0.455	8.57	0.320	0.449	8.56	0.320	0.449
0.15	11.4	0.281	0.467	12.1	0.274	0.472	11.4	0.281	0.467	11.4	0.281	0.467
0.20	14.9	0.248	0.484	15.6	0.243	0.490	14.9	0.248	0.484	14.9	0.248	0.484
0.25	18.9	0.221	0.501	19.7	0.217	0.506	18.9	0.221	0.501	18.9	0.221	0.501
0.30	23.7	0.199	0.516	24.4	0.196	0.521	23.7	0.199	0.516	23.7	0.199	0.516
0.35	29.2	0.180	0.530	29.9	0.178	0.534	29.2	0.180	0.530	29.2	0.180	0.530
0.40	35.9	0.163	0.542	36.5	0.162	0.545	35.9	0.163	0.542	35.9	0.163	0.542
0.45	44.2	0.147	0.553	44.8	0.146	0.556	44.2	0.147	0.553	44.2	0.147	0.553
0.50	54.6	0.133	0.563	55.2	0.132	0.566	54.6	0.133	0.563	54.6	0.133	0.563

Inner cut-out index $N = 4$

	Inner cut-out disk			Doubly cut-out disk with $N = 4$, $\tilde{R}_\mathrm{c} = 10$								
	$N = 4$			$M = 2$			$M = 4$			$M = 6$		
β	γ_max	$\tilde{\sigma}_{u,\mathrm{min}}$	Ω_p	γ_max	$\tilde{\sigma}_{u,\mathrm{min}}$	Ω_p	γ_max	$\tilde{\sigma}_{u,\mathrm{min}}$	Ω_p	γ_max	$\tilde{\sigma}_{u,\mathrm{min}}$	Ω_p
-0.25	0.446	1.028	0.312	0.512	0.994	0.315	0.460	1.020	0.312	0.458	1.021	0.313
-0.20	0.597	0.914	0.321	0.674	0.886	0.325	0.612	0.908	0.322	0.607	0.910	0.322
-0.15	0.800	0.816	0.332	0.886	0.794	0.335	0.815	0.812	0.333	0.807	0.815	0.332
-0.10	1.06	0.734	0.342	1.15	0.717	0.345	1.07	0.732	0.343	1.06	0.734	0.343
-0.05	1.37	0.664	0.352	1.46	0.650	0.355	1.38	0.663	0.353	1.37	0.664	0.353
0.00	1.76	0.602	0.363	1.86	0.591	0.366	1.76	0.601	0.363	1.76	0.602	0.363
0.05	2.23	0.548	0.374	2.35	0.539	0.376	2.23	0.548	0.374	2.23	0.548	0.374
0.10	2.84	0.497	0.385	2.98	0.489	0.388	2.85	0.497	0.385	2.84	0.497	0.385
0.15	3.65	0.450	0.397	3.81	0.443	0.400	3.65	0.449	0.397	3.65	0.449	0.397
0.20	4.71	0.405	0.410	4.89	0.399	0.412	4.72	0.404	0.410	4.71	0.404	0.410
0.25	6.13	0.362	0.423	6.34	0.357	0.425	6.14	0.362	0.423	6.13	0.362	0.423
0.30	8.05	0.322	0.436	8.31	0.318	0.438	8.06	0.322	0.436	8.06	0.322	0.436
0.35	10.7	0.284	0.450	11.0	0.280	0.452	10.7	0.284	0.451	10.7	0.284	0.451
0.40	14.5	0.248	0.466	14.9	0.245	0.468	14.5	0.248	0.466	14.5	0.248	0.466
0.45	19.8	0.215	0.483	20.3	0.212	0.485	19.9	0.214	0.483	19.9	0.214	0.483
0.50	27.4	0.184	0.502	28.1	0.182	0.504	27.5	0.184	0.502	27.5	0.184	0.502

Appendix F

Tables of Fastest-Growing Modes

F.1 Introduction

In this Appendix we present some of the modes which exist in a disk of a particular temperature, concentrating on the fastest-growing modes. Not all disks admit modes at a particular temperature. By way of illustration, consider the following sketches, which show the dependence of the largest eigenvalue on s and Ω_{p} for the first six inner cut-out indices N. The position of the unit eigenvalue is marked by the dotted line. This set of sketches is for $\beta = 0.25$ disks exposed to an $m = 2$ perturbation.

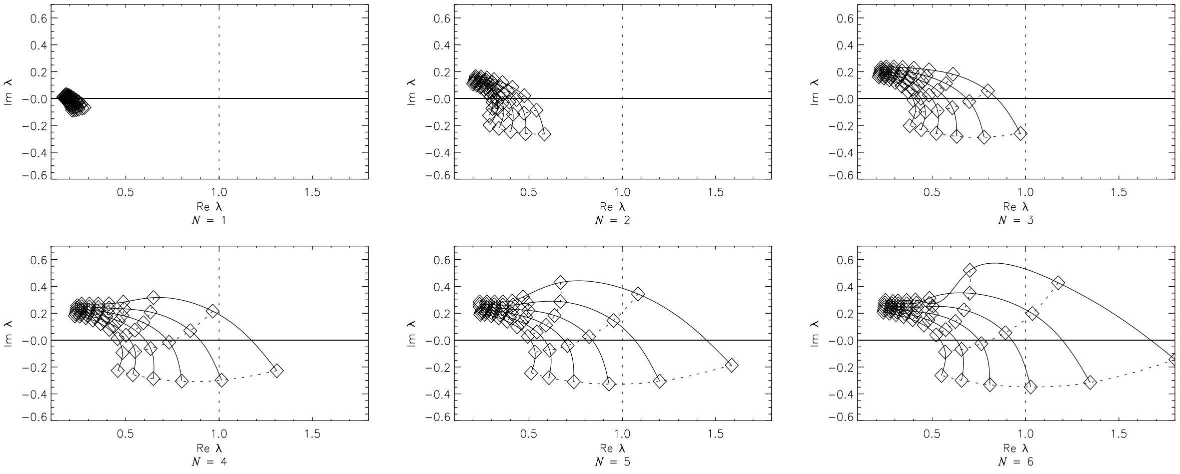


Figure F.1: $m = 2$, $\beta = 0.25$: dependence of the largest eigenvalue on growth rate and pattern speed. This is for $Q_{\mathrm{s}} = 1$, i.e. $\tilde{\sigma}_u = 0.283$ and $\gamma = 11.0$. The solid lines show curves of constant s at intervals of 0.1 from $s = 0^+$ to $s = 0.5$; the dotted lines show curves of constant Ω_{p} at intervals of 0.1 from $\Omega_{\mathrm{p}} = 0.4$ to 1.2.

We see immediately that disks with $N = 1$, 2 or 3 admit no modes for $Q_{\mathrm{s}} = 1$. The disk with $N = 4$ admits a mode at roughly $s \approx 0.05$, $\Omega_{\mathrm{p}} \approx 0.45$. Plots like this were used to obtain estimates of the position of the unit eigenvalue (if one existed). A Newton-Raphson routine was then used to locate it exactly, as described in section 5.7. Note that the more unstable disks may admit secondary modes. The above plots show the largest eigenvalue; if we were to plot the second largest eigenvalue for $N \geq 4$, we might well find that it passed through (1,0) for some lower growth rate.

In the Appendix, we consider disks with $Q_{\mathrm{s}} = 1$. This is the temperature at which a self-consistent disk with this potential would be locally just stable to axisymmetric modes.

F.2 One-armed modes

The power-law disks are highly susceptible to one-armed modes. The following table shows patterns speed and growth rates of the four fastest-growing modes present at $Q_\mathrm{s} = 1$ in inner cut-out disks with cut-out indices from $N = 1$ to $N = 4$. These data are presented graphically in figs. 8.13 and 8.14.

[Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_\mathrm{min} = 20$, $l_\mathrm{max} = 30$, $n_\mathrm{GL} = 9$, $f_\sigma = 0.8$, $a_\mathrm{acc} = 0.0$, $b_\mathrm{acc} = 0.0$, $\epsilon_\lambda = 10^{-6}$.]

	$N = 1$		$N = 2$		$N = 3$		$N = 4$	
β	Ω_p	s	Ω_p	s	Ω_p	s	Ω_p	s
0.50	0.116031	0.004346	0.182697	0.028965	0.224783	0.053632	0.252009	0.072953
	0.115773	0.003758	0.178884	0.029333	0.216303	0.050112	0.239136	0.064532
	0.077871	0.007007	0.113135	0.026346	0.132634	0.039889	0.145034	0.048827
	0.076778	0.008366	0.106751	0.023694	0.122052	0.032038	0.130663	0.036962
0.45	0.111193	0.006420	0.177255	0.033803	0.217336	0.059636	0.243004	0.079495
	0.111252	0.005531	0.172575	0.033328	0.208016	0.054308	0.229364	0.068592
	0.072140	0.009062	0.105614	0.029071	0.124025	0.042631	0.135747	0.051392
	0.070245	0.010098	0.098733	0.025051	0.112997	0.033088	0.120948	0.037815
0.40	0.107273	0.007552	0.172017	0.038339	0.210333	0.065096	0.234681	0.085365
	0.106798	0.008602	0.166631	0.036969	0.200365	0.058068	0.220462	0.072219
	0.066883	0.010877	0.098848	0.031213	0.116403	0.044699	0.127572	0.053261
	0.064426	0.011448	0.091638	0.026062	0.105049	0.033851	0.112468	0.038424
0.35	0.103679	0.009700	0.167036	0.042563	0.203808	0.070078	0.227035	0.090670
	0.102768	0.010783	0.161073	0.040306	0.193323	0.061486	0.212351	0.075520
	0.062099	0.012379	0.092793	0.032864	0.109643	0.046226	0.120341	0.054592
	0.059276	0.012482	0.085359	0.026815	0.098051	0.034411	0.105034	0.038864
0.30	0.100387	0.011881	0.162347	0.046492	0.197769	0.074651	0.220035	0.095509
	0.099067	0.012907	0.155906	0.043389	0.186850	0.064641	0.204957	0.078581
	0.057780	0.013578	0.087381	0.034110	0.103631	0.047323	0.113913	0.055509
	0.054729	0.013267	0.079795	0.027380	0.091870	0.034827	0.098486	0.039189
0.25	0.097362	0.014041	0.157969	0.050158	0.192207	0.078881	0.213645	0.099969
	0.095674	0.014952	0.151122	0.046268	0.180907	0.067599	0.198208	0.081469
	0.053903	0.014513	0.082547	0.035029	0.098266	0.048084	0.108175	0.056112
	0.050720	0.013862	0.074854	0.027811	0.086392	0.035146	0.092695	0.039438
0.20	0.094584	0.016149	0.153910	0.053598	0.187108	0.082831	0.207827	0.104126
	0.092573	0.016914	0.146706	0.048989	0.175453	0.070413	0.192041	0.084239
	0.050435	0.015230	0.078225	0.035689	0.093468	0.048588	0.103037	0.056480
	0.047183	0.014316	0.070459	0.028146	0.081523	0.035401	0.087555	0.039643
0.15	0.092045	0.018197	0.150173	0.056853	0.182453	0.086561	0.202544	0.108049
	0.089752	0.018804	0.142643	0.051595	0.170451	0.073131	0.186404	0.086938
	0.047344	0.015774	0.074361	0.036148	0.089169	0.048899	0.098427	0.056679
	0.044061	0.014666	0.066544	0.028418	0.077185	0.035618	0.082980	0.039825
0.10	0.089742	0.020188	0.146757	0.059965	0.178224	0.090126	0.197765	0.111803
	0.087201	0.020636	0.138918	0.054123	0.165869	0.075796	0.181250	0.089607
	0.044593	0.016185	0.070908	0.036457	0.085318	0.049073	0.094291	0.056761
	0.041306	0.014945	0.063052	0.028653	0.073314	0.035820	0.078899	0.040004

	$N=1$		$N=2$		$N=3$		$N=4$	
β	Ω_{p}	s	Ω_{p}	s	Ω_{p}	s	Ω_{p}	s
0.05	0.087673	0.022135	0.143660	0.062976	0.174407	0.093580	0.193463	0.115448
	0.084913	0.022429	0.135518	0.056611	0.161682	0.078445	0.176542	0.092284
	0.042155	0.016498	0.067829	0.036661	0.081874	0.049155	0.090586	0.056771
	0.038876	0.015176	0.059938	0.028872	0.069857	0.036024	0.075254	0.040199
0.00	0.085847	0.024053	0.140888	0.065923	0.170997	0.096969	0.189628	0.119035
	0.082886	0.024203	0.132440	0.059090	0.157878	0.081111	0.172262	0.095000
	0.040006	0.016743	0.065101	0.036798	0.078812	0.049188	0.087288	0.056750
	0.036740	0.015381	0.057169	0.029092	0.066779	0.036249	0.072008	0.040422
-0.05	0.084258	0.025976	0.138431	0.068868	0.167973	0.100368	0.186227	0.122648
	0.081108	0.025994	0.129662	0.061611	0.154422	0.083849	0.168368	0.097811
	0.038114	0.016949	0.062686	0.036906	0.076096	0.049210	0.084358	0.056735
	0.034859	0.015579	0.054703	0.029335	0.064030	0.036511	0.069107	0.040694
-0.10	0.082929	0.027921	0.136316	0.071844	0.165359	0.103817	0.183285	0.126332
	0.079596	0.027817	0.127205	0.064199	0.151336	0.086684	0.164876	0.100742
	0.036478	0.017140	0.060588	0.037019	0.073731	0.049259	0.081804	0.056766
	0.033226	0.015786	0.052534	0.029617	0.061606	0.036829	0.066548	0.041031
-0.15	0.081875	0.029926	0.134558	0.074911	0.163167	0.107388	0.180810	0.130167
	0.078360	0.029711	0.125078	0.066902	0.148621	0.089670	0.161788	0.103849
	0.035085	0.017342	0.058800	0.037173	0.071714	0.049377	0.079622	0.056884
	0.031825	0.016022	0.050646	0.029959	0.059490	0.037226	0.064312	0.041457
-0.20	0.081128	0.032031	0.133192	0.078134	0.161432	0.111163	0.178835	0.134244
	0.077422	0.031714	0.123307	0.069775	0.146306	0.092867	0.159126	0.107195
	0.033936	0.017579	0.057331	0.037408	0.070056	0.049610	0.077827	0.057140
	0.030650	0.016305	0.049039	0.030389	0.057682	0.037730	0.062398	0.042002
-0.25	0.080732	0.034289	0.132275	0.081594	0.160210	0.115238	0.177418	0.138674
	0.076822	0.033875	0.121939	0.072886	0.144436	0.096350	0.156939	0.110862
	0.033043	0.017882	0.056205	0.037773	0.068789	0.050017	0.076453	0.057598
	0.029707	0.016661	0.047726	0.030939	0.056195	0.038379	0.060822	0.042707
-0.30	0.080758	0.036769	0.131889	0.085397	0.159592	0.119745	0.176651	0.143604
	0.076619	0.036262	0.121046	0.076326	0.143090	0.100225	0.155306	0.114960
	0.032432	0.018287	0.055466	0.038330	0.067964	0.050673	0.075554	0.058342
	0.029015	0.017121	0.046735	0.031656	0.055064	0.039225	0.059618	0.043627
-0.35	0.081307	0.039564	0.132160	0.089688	0.159710	0.124860	0.176673	0.149233
	0.076909	0.038965	0.120742	0.080221	0.142389	0.104636	0.154351	0.119648
	0.032148	0.018844	0.055186	0.039163	0.067664	0.051684	0.075223	0.059491
	0.028609	0.017731	0.046119	0.032605	0.054345	0.040347	0.058849	0.044847
-0.40	0.082537	0.042812	0.133275	0.094669	0.160770	0.130831	0.177695	0.155843
	0.077836	0.042117	0.121200	0.084756	0.142518	0.109794	0.154268	0.125153
	0.032269	0.019623	0.055481	0.040396	0.068023	0.053205	0.075598	0.061216
	0.028553	0.018556	0.045964	0.033885	0.054138	0.041856	0.058617	0.046488
-0.45	0.084693	0.046720	0.135523	0.100644	0.163077	0.138025	0.180034	0.163847
	0.079627	0.045922	0.122684	0.090205	0.143763	0.116017	0.155351	0.131821
	0.032923	0.020732	0.056531	0.042214	0.069243	0.055464	0.076897	0.063777
	0.028951	0.019700	0.046414	0.035647	0.054598	0.043929	0.059087	0.048740

F.3 Bisymmetric modes

The power-law disks are remarkably stable to bisymmetric modes. For such modes to be possible at $Q_{\mathrm{s}} = 1$, the disk must be sharply cut-out at the centre. Disks with rising rotation curves are less stable than disks with falling rotation curves. However, even the disk with $N = 4$ and $\beta = -0.5$ admits only three modes at $Q_{\mathrm{s}} =$. Disks with lower cut-out indices or higher β admit even fewer modes. Thus the following table has many empty cells!

[Numerical accuracy parameters: $n = 251$, $\Delta\alpha = 0.2$, $l_{\mathrm{min}} = 20$, $l_{\mathrm{max}} = 30$, $n_{\mathrm{GL}} = 9$, $f_{\sigma} = 0.8$, $a_{\mathrm{acc}} = 20.0$, $b_{\mathrm{acc}} = 2.0$, $\epsilon_{\lambda} = 10^{-6}$.]

	$N = 2$		$N = 3$		$N = 4$	
β	Ω_{p}	s	Ω_{p}	s	Ω_{p}	s
-0.50	0.440576	0.017128	0.430631	0.195548	0.434722	0.274554
	-	-	0.366152	0.064841	0.356211	0.131223
	-	-	-	-	0.300804	0.036166
-0.45	-	-	0.426456	0.173846	0.430563	0.254195
	-	-	0.365842	0.044340	0.356711	0.114589
	-	-	-	-	0.302216	0.018754
-0.40	-	-	0.424017	0.154224	0.428110	0.236340
	-	-	0.366615	0.025186	0.358439	0.099353
	-	-	-	-	0.304429	0.002656
-0.35	-	-	0.422879	0.136016	0.426966	0.220258
	-	-	0.368121	0.006956	0.361028	0.085007
-0.30	-	-	0.422735	0.118765	0.426848	0.205435
	-	-	-	-	0.364235	0.071232
-0.25	-	-	0.423360	0.102143	0.427558	0.191497
	-	-	-	-	0.367896	0.057828
-0.20	-	-	0.424591	0.085912	0.428947	0.178167
	-	-	-	-	0.371900	0.044667
-0.15	-	-	0.426303	0.069894	0.430904	0.165227
	-	-	-	-	0.376176	0.031674
-0.10	-	-	0.428403	0.053949	0.433347	0.152509
	-	-	-	-	0.380683	0.018805
-0.05	-	-	0.430823	0.037970	0.436212	0.139873
	-	-	-	-	0.385398	0.006036
0.00	-	-	0.433514	0.021833	0.439455	0.127182
0.05	-	-	0.436418	0.005560	0.443019	0.114415
0.10	-	-	-	-	0.446896	0.101413
0.15	-	-	-	-	0.451058	0.088125
0.20	-	-	-	-	0.455493	0.074482
0.25	-	-	-	-	0.460195	0.060416
0.30	-	-	-	-	0.465166	0.045857
0.35	-	-	-	-	0.470415	0.030731
0.40	-	-	-	-	0.475963	0.014956

F.4 Stability to different azimuthal harmonics

The following sets of tables compare the stability of inner cut-out disks to perturbations of different azimuthal symmetry m. For each symmetry m and inner cut-out index N, the pattern speed and growth rate of the fastest-growing mode are recorded. The results for $\beta = 0.25$, $\beta = 0.00$ and $\beta = -0.25$ are shown in different tables.

[Numerical accuracy parameters: $n = 301$, $\Delta\alpha = 0.1$, $l_{\min} = -30$, $l_{\max} = +40$, $n_{GL} = 9$, $f_\sigma = 0.8$, $a_{\rm acc} = 20$, $b_{\rm acc} = 2$, $\epsilon_\lambda = 10^{-6}$.]

$Q_{\rm s} = 1.0$, $\beta = 0.25$: $\tilde{\sigma}_u = 0.283$, $\gamma = 11.0$						
	$m = 1$		$m = 2$		$m = 3$	
N	$\Omega_{\rm p}$	s	$\Omega_{\rm p}$	s	$\Omega_{\rm p}$	s
1	0.097362	0.014041	-	-	-	-
2	0.157969	0.050158	-	-	-	-
3	0.192207	0.078881	-	-	-	-
4	0.213645	0.099969	0.460196	0.060416	-	-
5	0.228252	0.115400	0.466727	0.128279	-	-
6	0.238798	0.126749	0.472435	0.170714	-	-
7	0.246718	0.135160	0.477593	0.199573	0.619188	0.016944
8	0.252824	0.141454	0.482190	0.220295	0.623895	0.076268

$Q_{\rm s} = 1.0$, $\beta = 0.00$: $\tilde{\sigma}_u = 0.378$, $\gamma = 6.00$						
	$m = 1$		$m = 2$		$m = 3$	
N	$\Omega_{\rm p}$	s	$\Omega_{\rm p}$	s	$\Omega_{\rm p}$	s
1	0.085847	0.024053	-	-	-	-
2	0.140888	0.065923	-	-	-	-
3	0.170997	0.096969	0.433518	0.021843	-	-
4	0.189628	0.119035	0.439460	0.127192	-	-
5	0.202258	0.134777	0.445317	0.185191	-	-
6	0.211327	0.146108	0.451135	0.222089	0.600215	0.006528
7	0.218103	0.154365	0.456487	0.247575	0.605870	0.081236
8	0.223301	0.160461	0.461228	0.266062	0.610127	0.134753

$Q_{\rm s} = 1.0$, $\beta = -0.25$: $\tilde{\sigma}_u = 0.509$, $\gamma = 3.36$						
	$m = 1$		$m = 2$		$m = 3$	
N	$\Omega_{\rm p}$	s	$\Omega_{\rm p}$	s	$\Omega_{\rm p}$	s
1	0.080733	0.034289	-	-	-	-
2	0.132275	0.081594	-	-	-	-
3	0.160210	0.115238	0.423391	0.102223	-	-
4	0.177418	0.138674	0.427591	0.191581	-	-
5	0.189107	0.155207	0.433935	0.242558	-	-
6	0.197443	0.166961	0.440325	0.276073	0.599256	0.141945
7	0.203663	0.175475	0.446097	0.299705	0.599256	0.141945
8	0.208430	0.181740	0.451131	0.317072	0.603496	0.190117

The above data are presented in graphical form in the following plots:

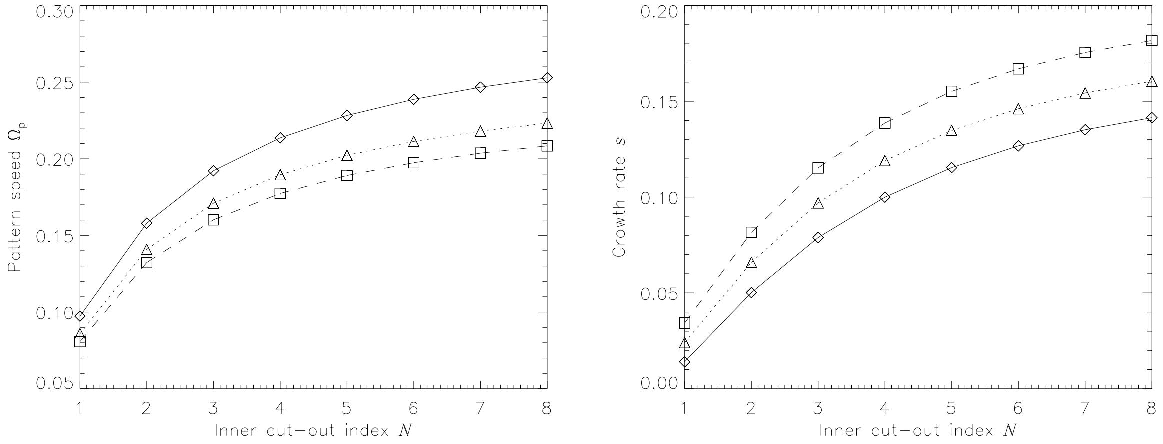


Figure F.2: The fastest-growing modes with $m = 1$, $Q_s = 1.0$. The solid line marked with diamonds is $\beta = 0.25$; the dotted line with triangles is $\beta = 0.00$; the dashed line with squares is $\beta = -0.25$.

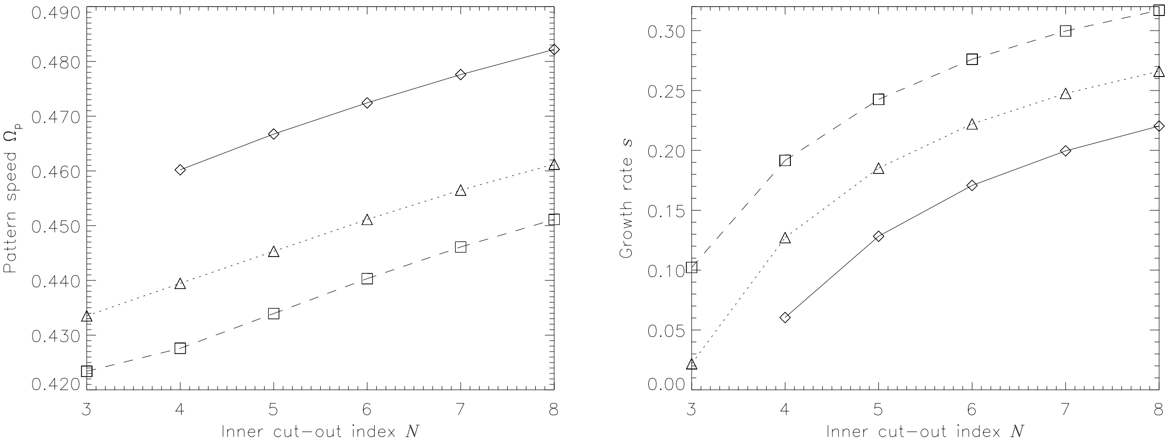


Figure F.3: The fastest-growing modes with $m = 2$, $Q_s = 1.0$. The solid line marked with diamonds is $\beta = 0.25$; the dotted line with triangles is $\beta = 0.00$; the dashed line with squares is $\beta = -0.25$.

The pattern speed of the fastest-growing mode depends almost linearly on N, although there is not a great deal of variation with either N or β. The fastest growth rate increases quite steeply with N, reflecting the increased instability of the more sharply cut-out disks.

Bibliography

Arfken G., 1966, *Mathematical Methods for Physicists*, Academic Press, London

Athanassoula E. & Sellwood J. A., 1986, *MNRAS*, **221**, 213

Berezin I. S. & Zhidkov N. P., 1965, *Computing Methods*, Pergamon Press, Oxford

Bertin G. & Lin C. C., 1996, *Spiral Structure in Galaxies*, MIT, Massachussetts

Binney J. J. & Tremaine S., 1987, *Galactic Dynamics*, Princeton, Princeton

Bryan G. H., 1889, *Phil. Trans. Roy. Soc. London A*, **180**, 187

Casertano S. & van Gorkom J. H., 1991, *A.J.*, **101**, 1231

Chamberlin T. C., 1901, *ApJ*, **14**, 17

Clemmow P. & Dougherty J. P., 1969, *Electrodynamics of Particles and Plasmas*, Addison-Wesley, Reading, Massachusetts

Delves L. M. & Mohamed J. L., 1985, *Computational Methods for Integral Equations*, Cambridge University Press, Cambridge

Dyson J. & Schutz B. F., 1979, *Proc. Roy. Soc. A*, **368**, 389

Earn D., 1993, Ph.D. thesis, Cambridge University, Cambridge

Elmegreen D. M. & Elmegreen B. G., 1982, *MNRAS*, **201**, 1021

Elmegreen D. M. & Elmegreen B. G., 1985, *ApJ*, **288**, 438

Evans N. W., 1994, *MNRAS*, **267**, 333

Evans N. W. & Collett J. L., 1993, *MNRAS*, **264**, 353

Fricke W., 1951, *Astron. Nachr.*, **280**, 193

Goldreich P. & Lynden-Bell D., 1965a, *MNRAS*, **130**, 125

Goldreich P. & Lynden-Bell D., 1965b, *MNRAS*, **130**, 97

Goldreich P. & Tremaine S., 1978, *ApJ*, **222**, 850

Gradshteyn I. & Ryzhik I., 1978, *Table of Integrals, Series and Products*, Academic Press, San Diego

Hénon M., 1973, *A&A*, **24**, 229

Hohl F., 1971, *Ap.J.*, **168**, 343

Hunter C., 1963, *MNRAS*, **126**, 299

Jeans J. H., 1915, *MNRAS*, **76**, 70

Jeans J. H., 1919, *Problems of Cosmogony & Stellar Dynamics*, Cambridge University Press, Cambridge

Julian W. H. & Toomre A., 1966, *Ap.J.*, **146**, 810

Kalnajs A. J., 1965, Ph.D. thesis, Harvard University, Cambridge, Massachusetts

Kalnajs A. J., 1971, *Ap.J.*, **166**, 275

Kalnajs A. J., 1972, *Ap.J.*, **175**, 63

Landau L. D. & Lifshitz E. M., 1960, *Mechanics*, Pergamon Press, Oxford

Lemos J. P. S., Kalnajs A. J., & Lynden-Bell D., 1991, *Ap.J.*, **375**, 484

Lin C. C. & Shu F., 1964, *Ap.J.*, **140**, 646

Lin C. C. & Shu F., 1966, *Proc. Nat. Acad. Sci. U.S.A.*, **55**, 229

Lynden-Bell D., 1962, *MNRAS*, **124**, 1

Lynden-Bell D. & Lemos J. P. S., 1996, *Equiangular spiral modes of power law disks*, unpublished

Lynden-Bell D. & Ostriker J. P., 1967, *MNRAS*, **136**, 101

Mark J. W.-K., 1974, *Ap.J.*, **193**, 539

Maxwell J. C., 1859, in W. D. Niven (ed.), *The Scientific Papers of James Clerk Maxwell*, Dover

Mestel L., 1963, *MNRAS*, **126**, 553

Mihalas D. & Binney J. J., 1987, *Galactic Astronomy*, W. H. Freeman, San Francisco

Miyamoto M., 1971, *PASJ*, **23**, 21

Miyamoto M., 1974, *A&A*, **30**, 441

Ostriker J. P. & Peebles P. J. E., 1973, *Ap.J.*, **186**, 467

Palmer P. L. & Papaloizou J., 1990, *MNRAS*, **243**, 263

Parsons W., 1850, *Philos. Trans. R. Soc.*, 503

Poincaré H., 1911, *Lectures on cosmogonical hypotheses*, Herman, Paris

Press W., Flannery B., Teukolsky S., & Vetterling W., 1989, *Numerical Recipes: the art of scientific computing*, Cambridge University Press, Cambridge

Prudnikov A. P., Brychkov Y. A., & Mavichev O., 1986, *Integrals and Series, vol. 1*, Gordon and Breach, New York

Rubin V. C., W. Kent Ford J., & Thonnard N., 1978, *Ap.J.*, **225**, L107

Rudin W., 1976, *Principles of Mathematical Analysis*, McGraw-Hill, New York

Sandage A., 1961, *The Hubble Atlas of Galaxies*, Carnegie Institute of Washington, Washington

Schmitz F. & Ebert R., 1987, *A&A*, **181**, 41

Sellwood J. A., 1983, *J. Comp. Phys.*, **50**, 337

Sellwood J. A. & Athanassoula E., 1986, *MNRAS*, **221**, 195

Sellwood J. A. & Wilkinson A., 1993, *Reports on Progress in Physics*, **56**, 173

Smith B. T., Boyle J. M., Dongarra J. J., Garbow B. S., Ikebe Y., Klema V. C., & Moler C. B., 1972, *Lecture Notes in Computer Science*, Vol. 6, Springer-Verlag, Berlin-Heidelberg-New York, 2nd edition

Syer D. & Tremaine S., 1996, *MNRAS*, **281**, 925

Toomre A., 1964, *Ap.J.*, **139**, 1217

Toomre A., 1966, in *Proceedings of the Woods Hole Oceanographic Institute*

Toomre A., 1969, *Ap.J.*, **158**, 899

Toomre A., 1972, *QJRAS*, **13**, 266

Toomre A., 1975, unpublished

Toomre A., 1977, *ARAA*, **15**, 437

Toomre A., 1981, in S. M. Fall & D. Lynden-Bell (eds.), *The structure and evolution of normal galaxies*,

p. 111, Cambridge University Press, Cambridge

Toomre A., 1995, Talk given at the 36th Herstmonceux Conference, Cambridge

Toomre A. & Toomre J., 1972, *Ap.J.*, **178**, 623

Tricomi F., 1985, *Integral Equations*, Dover, New York

van Kampen N. G., 1955, *Physica*, **21**, 949

Vauterin P. & Dejonghe H., 1996, *A&A*, **313**, 465

Visser H. C. D., 1980, *A&A*, **88**, 159

Zang T. A., 1976, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts