Dynamics of
self-gravitating disks

Christophe Pichon
Clare College
and
Institute of Astronomy

A dissertation submitted to the University of Cambridge
in accordance with the regulations for admission
to the degree of Doctor of Philosophy
Summary

A fair fraction of observed galactic disks present a bar corresponding to elongated isophotes beyond the central core. Numerical simulations also suggest that bar instabilities are quite common. Why should some galaxies have bars and others not? This question is addressed here by investigating the stability of a self gravitating disk with respect to instabilities induced by the adiabatic re-alignment of resonant orbits. It is shown that the dynamical equations can be recast in terms of the interaction of resonant flows. Concentrating on the stars belonging to the inner Lindblad resonance, it is argued that this interaction yields an instability of Jeans' type. The Jeans instability traps stars moving in phase with respect to a growing potential; here the azimuthal instability corresponds to a rotating growing potential that captures the lobe of orbital streams of resonant stars. The precession rate of this growing potential is identified with the pattern speed of the bar. The instability criterion puts constraints on the relative dispersion in angular frequency of the underlying distribution function. The outcome of orbital instability is investigated while constructing distribution functions induced by the adiabatic re-alignment of inner Lindblad orbits. These distribution functions maximise entropy while preserving the underlying axisymmetric component of the galaxy and the constraints of angular momentum, total energy, and detailed circulations conservation. Below a given temperature, the disk prefers a barred configuration with a given pattern speed. A analysis of the thermodynamics of these systems is presented when the interaction between the orbits is independent of their shapes.

Axisymmetric distribution functions characterise the equilibrium state of a galaxy. Their formal expression in terms of the invariants of the motion is desirable both from a theoretical and observational point of view: they constrain observed radial and azimuthal velocity distributions measured from the absorption lines in order to assess the gravitational nature of the equilibrium; they correspond to the core of all numerical and linear analyses studying non-axisymmetric features to describe the unperturbed axisymmetric initial state. A general inversion method is presented here leading to the construction of distribution functions chosen to match either a given potential-density pair and a given temperature profile, or alternatively to account for detailed observed kinematics.

Complete sequences of new analytic solutions of Einstein’s equations which describe thin supermassive disks are constructed. These solutions are derived geometrically. The identification of points across two symmetrical cuts through a vacuum solution of Einstein’s equations defines the gradient discontinuity from which the properties of the disk can be deduced. The subset of possible cuts which lead to physical solutions is presented. At large distances, all these disks become Newtonian, but in their central regions they exhibit relativistic features such as velocities close that of light, and large redshifts. For static metrics, the vacuum solution is derived via line superposition of fictitious sources placed symmetrically on each side of the disk by analogy with the classical method of images. The corresponding solutions describe two counter-rotating stellar streams. Sections with zero extrinsic curvature yield cold disks. Curved sections may induce disks which are stable against radial instability. The general counter rotating flat disk with planar pressure tensor is found. Owing to gravomagnetic forces, there is no systematic method of constructing vacuum stationary fields for which the non-diagonal component of the metric is a free parameter. However, all static vacuum solutions may be extended to fully stationary fields via simple algebraic transformations. Such disks can generate a great variety of different metrics including Kerr’s metric with any ratio of $a$ to $m$. A simple inversion formula is given which yields all distribution functions compatible with the characteristics of the flow, providing formally a complete description of the stellar dynamics of flattened relativistic disks.

Cambridge
February 1994
This dissertation is the outcome of my own work, except where explicit reference has been made to the work of others.

Section 2 of chapter 2 has already been published as Pichon, C. “Orbital instabilities in galaxies”, contribution to the “Ecole de Physique des Houches” on: Transport phenomena in Astrophysics, Plasma Physics, and Nuclear Physics. In press. The idea of applying thermodynamics to study the evolution of bars within galaxies was suggested to me by Jim Collett.

Most of the material presented in chapter 4 was published as Bicák, J., Lynden-Bell, D & Pichon C. “Relativistic discs and flat galaxy models”, Monthly Notices of the Royal Astronomical Society (1993) 265 126-144, where the basic method was suggested by Lynden-Bell, and carried out independently Bicák, Lynden-Bell and myself.

The material of chapter 5 has been submitted to Physical Review D for publication as: Pichon, C. & Lynden-Bell, D. “New sources of Kerr and other metrics: Rapidly rotating relativistic disks with pressure support”, and the main ideas developed here are presented in Pichon, C. & Lynden-Bell, D. “Quasar engines and relativistic disk models”, contribution to the conference: N-body problems and gravitational dynamics, CNRS Editions.

I hereby declare that this dissertation is not substantially the same as any that I have submitted for a degree or diploma or other qualification at any other University, and no part of it has been or is submitted for any such degree, diploma or any other qualification.

This dissertation does not exceed 60,000 words in length.

Christophe Pichon
Cambridge
February 1994
Dedicated to
the memory of John Treherne
Un de mes amis definissait l’observateur comme celui qui ne comprend rien à la théorie, et le théoricien comme un exalté qui ne comprend rien à rien.

L. Boltzmann (1890)
1

Prologue

Various aspects of the dynamics of self-gravitating disks are presented here. Chapter 2 addresses the stability of thin axisymmetric galactic stellar disks with respect to non-axisymmetric perturbations. Models of such disks accounting for observed or parameterised kinematics are constructed in chapter 3. Chapters 4 and 5 present a derivation of the relativistic stellar dynamics of flattened systems applied to models of rotating and counter-rotating super-massive disks.

The dynamics of disks has been the subject of many excellent reviews in recent years, of which those of Sellwood (1993) [101] for bars, and Tremaine (1989) [109] are particularly worthy of mention for their scope and detail. In view of this, given that each chapter has its own introduction, a different perspective is taken in this prologue: the main common features and hypotheses, such as that of an axisymmetric infinitely thin disk, are sketched. These assumptions provide the mathematical simplicity which allows more detailed investigation without resorting to sophisticated numerical analysis.

The infinitely thin disk approximation is used throughout as it reduces the number of dimensions involved, and is essential in constructing solutions of Einstein’s equation geometrically. As in electromagnetism, the properties of the distribution of sources is deduced in this context from the jump conditions at the boundary of the discontinuity. This approximation allows tremendous simplification in comparison with the fully three dimensional set of Einstein’s equations. This assumption together with that of axisymmetry makes the distribution function inversion problem a degenerate case (two degrees of functional freedom in the distribution function varying with energy and angular momentum, only one functional constraint while requiring that the distribution adds up to the right surface density). In effect, the thin disk approximation corresponds to an extra explicit integral of the motion which allows the inversion while also requiring that the solution satisfies all observed or prescribed kinematical properties. The infinitely thin disk approximation turns out to be useful also when studying the stability of the disk as it reduces the number of phase space dimensions required to describe the dynamics to twice the number of available invariants for an axisymmetric disk. This in turn implies that the motion of stars is regular (not chaotic), and suggests labelling the trajectories by combinations of these invariants called action variables. These variables have two attractive properties: the conjugate variable varies uniformly between zero and $2\pi$ as the star describes one oscillation of the corresponding degree of freedom\(^1\); actions may behave adiabatically, \textit{i.e.} remain approximately constant, as the slow perturbation is affecting the symmetry on which the conservation of these actions relies.

All disks are assumed to have initially reached a stationary equilibrium. The energy of each

\(^1\) The formalism of angle and action variables is presented in appendix A.
star, or fluid element, is therefore an invariant of the motion corresponding to the fact that the equilibrium is left unchanged by translations in time.

All disks are also assumed to be (at least initially) axisymmetric. This symmetry hypothesis yields a supplementary invariant corresponding to the conservation of angular momentum. As mentioned above, this invariant is crucial both for the construction of distribution functions presented in chapter 3 and to impose regular motions within the disk. In the context of relativity, axial symmetry reduces the number of unknown functions to be found from ten to three, and the number of variables correspondingly.

Once all symmetry hypothesis have been made, it is instructive to analyse the behaviour of these disks while assuming that some combinations of the actions behave like quasi-invariants as done in chapter 2, sections 3 and 4. At this level of description, it is then possible to analyse the instability in some detail, pointing out a clear analogy with the Jeans instability.

The above assumptions have a limited scope of validity which obviously depends on the type of physical disks to which they are applied. Relativistic gaseous disks are, for instance, crudely described as infinitely thin systems or even as stationary objects given that they radiate their binding energy at a luminosity close to the Eddington luminosity. Some galactic disks, however, are more likely to be well described as flat and axisymmetric, though it remains to be shown that all bars arise from initially axisymmetric disks. Adiabatic invariance assumes that the instability occurs slowly compared with the typical period of a star describing its orbit which is not strictly in agreement compared to the results of some numerical simulations. The assumption of adiabatic invariance yields nevertheless the correct criterion for the marginally unstable modes, though it is unclear that non-axisymmetric features observed in real galaxies correspond to such modes.

There are, however, good reasons for believing that most of the above assumptions are satisfactory within the scope of this work. For instance, chaotic orbits are present in real galaxies, though the diffusion of chaotic orbits away from the periodic ones takes somewhat longer than the age of the universe. Similarly, it is a simple matter, using the linearity of Poisson's equation, to thicken infinitely thin classical disks, and it is likely that a similar procedure is available for relativistic disks.
2

Orbital instabilities and bar formation

2.1 Introduction

Galactic bars were first mentioned in Astronomy by Curtis (1918) [22] who described “a fairly common type of spiral nebula whose main characteristic is a band of matter extending diametrically across the nucleus and inner part of the spiral. [...] The general appearance is that of the Greek letter Φ”. Later Hubble (1925 onwards) [48] introduced the term “barred galaxy” and distinguished two parallel series of spirals: normal or ordinary (S), and barred (SB). Roughly one third of all spiral galaxies show developed bars (de Vaucouleur 1963) [24] while another third present weak bars. Barred galaxies do not differ from non-barred ones as far as integral properties, like luminosities, fractional HI content, colours etc, are concerned. Why some spiral galaxies should form bars, and others not is therefore a puzzling challenge for astronomers.

Bar formation or growth has been variously attributed to

(1) the gradual outward transfer of stellar orbital angular momentum by spiral density waves and eccentricity pumping (Lynden-Bell & Kalnajs 1972 [76]),

(2) the density response of a stellar disk to a weak perturbation in the underlying potential which is amplified into a standing wave (Toomre 1981 [105], Zang 1976 [114], Kalnajs [54,56,58,59]), and

(3) the trapping of elongated stellar orbits by a bar-like perturbation in the inner part of galaxies. (Lynden-Bell 1973 [77], 1979 [82]). Sellwood 1993 [101] gives a extensive review of the pro and cons of these theories. This chapter addresses the question of bar formation from the viewpoint of orbital instability.

2.1.1 Context

Consider a rotating dissipative protogalactic cloud collapsing into what will be a galaxy. The cloud rapidly loses its energy of motion along the axis of rotation, but conserves its total angular momentum. It therefore settles into a flat system. While the process of baryonic matter infall from the halo continues, it can be argued that the disk becomes thin and axisymmetric. While cooling, the gas produces stars. The disk surface density in stars increases and the velocity dispersion in stars decreases as more and more gas is injected. This process goes on until
eventually an azimuthal instability occurs which redistributes angular momentum. This scenario is compatible with the observations of galactic disks, and provides a symmetry necessary for the analytical treatment. It does not rely on unknown initial conditions. The influence from external fields arising from encounters with other galaxies undoubtedly amplifies the growth rate of unstable eigen-modes. This analysis is nevertheless restricted to the normal modes of an isolated galaxy. The red colour of bar suggests neglecting the influence of the gas, though it is acknowledged that the inherent instability of the scarce gaseous component might trigger the stellar instability. The self consistent gravitational field may conceptually be split into two components: a mean field, satisfying Poisson equation for density, calculated by averaging locally over a Dirac distribution corresponding to point-like stars and the fluctuation of this mean field due to neighbouring stars, which can be evaluated statistically in order to quantify its relevance. In fact, the long range interaction of gravity and the large numbers of stars to be found in galaxies suggest that neglecting this latter component is a good approximation. The kinetic theory then provides a natural framework in which to follow statistically the behaviour of all the stars as they describe their orbits. While the fluid description is attractive for describing the global effect of the other stars on a given star, it is obviously insufficient to account for self-gravity enhancement induced by the motion of two stars tracing orbits which have a constant relative phase delay (i.e. orbits which resonate). Indeed, in stellar dynamics, as opposed to fluid mechanics, the mean free path of a given particle is quite large compared to the size of the system; therefore resonance will generally last until the weak but cumulative effect of the torque has modified the nature of the two orbits. This reorganisation of resonant orbits may restructure the galaxy.\footnote{From an energy point of view, it can be argued that the disk was maintained far from its minimum energy state by the conservation of angular momentum. Hence the dominant factor controlling its dynamical evolution is the way in which specific angular momentum may be redistributed within the disc. The mean field theory claims on the other hand that no angular momentum is exchanged between stars except on resonance. It follows that the interaction of resonant orbits provides a way to restructure the galaxy.}

The core of this analysis therefore relies on the two fundamental equations of galactic dynamics, namely

- Boltzmann’s equation, which describes the evolution of the distribution function in a given potential and characterises the dynamics,
- Poisson’s equation, which provides a prescription for the mean field induced by a given distribution of stars.

These equations may be re-expressed exactly in order to identify the contribution from the resonances of orbits, independently of the contribution corresponding to the detailed phase of the stars on each orbit. The gravitational interactions of individual stars may be re-expressed in terms of the gravitational interactions of whole orbital streams, i.e. steady streams of stars that follow the same orbit. The net effect of this interaction is to distort and re-orientate the orbits. Such instabilities restructure the galaxy only if the orbits co-operate by moving in the direction of the torques on them. Orbits corresponding to the symmetry of bars present such a cooperative behaviour. A collective re-alignment of quasi-resonant orbits is therefore expected to produce the azimuthal analogue to Jeans gravitational instability. Jeans instability works by trapping particles in a growing potential well. Azimuthal gravitational instability works here not by trapping the stars but by trapping the lobes of orbital streams of stars azimuthally in a growing potential well. As with the classical instability the potential well may be rotating relative to inertial axes. The rotation velocity of this frame is to be identified with the pattern speed of the bar.

The actual instability criterion is derived by very classical means. The starting point is a distribution function specifying a given galaxy. For small perturbations away from equilibrium,
the equations may be linearised to find a dispersion relation for the normal modes of the system. The condition for the existence of linear marginally-stable modes is derived and related to the constraints the existence of these modes would impose on the underlying galaxy.

This analysis yield the following results:

- an azimuthal instability criterion for a given galaxy, and
- the angular frequency at which the instability corresponding to this criterion propagates.

2.1.2 Outline

The fully self consistent model of a real galaxy sketched in this introduction is postponed to section 4. Section 2 introduces a toy model, the purpose of which is to give a general idea of the most relevant physical concepts; the behaviour of an assembly of tumbling ellipses as they go through bi-symmetric azimuthal instability is described in some detail. The idea of an effective interaction potential between the ellipses is introduced. This leads to a Jeans-like instability criterion. Section 3 focusses on a fictitious galaxy in which the only relevant resonance is the inner Lindblad one. Introducing an effective interaction potential similar to that of the first model, together with some assumptions about its analytical expression leads to an analogous criterion under less stringent conditions. It suggests that the orbits should be distorted as they re-orient themselves with respect to the potential trough. The fate of this adiabatic orbital instability is analysed and illustrated in section 4. Below a critical temperature, the system finds a non-axisymmetric configuration which rotates at a temperature dependent pattern speed. All the resonances are taken into account in section 5, where the interaction potential is derived exactly from Poisson’s equation. The exact integral equation accounting for the self consistency of the field is given in terms of angles and actions. The corresponding formulation is contrasted with that of Kalnajs (1971) [54] and Zang (1976) [114]. Various techniques to solve this integral equation are sketched and developed in the appendix.

2.2 The alignment instability model

The toy model described in the following section provides a simplified insight into the mechanism that may be involved in the production of bars in galaxies. It is constructed around three basic concepts introduced by Lynden-Bell (1979) [82] to account for bar formation via orbital instability. These address the following questions

- Which orbits form a bar?
  If a particular star has an orbit with radial period $2\pi/\kappa$ and an average period around the galaxy $2\pi/\Omega$, then when viewed from axes that rotate at the rate $\Omega - \kappa/2$, it will close in a figure not unlike a centred ellipse. If this whole orbit is populated by an orbital stream, it will look stationary in the rotating frame but will tumble over and over at the angular rate $\Omega - \kappa/2$ in the fixed frame. Numerical simulations suggest that the stationary pattern of such orbits form the skeleton of a bar.

- How do they interact?
  If two orbital streams tumble at the same rate, $\Omega$, but one has its lobes a little ahead of the other, then the gravitational interaction of the streams will produce a forwards torque on the hindmost stream and a backwards torque on the foremost stream. Since
the two tumbling rates are equal (resonant), this torque will not reverse but will continue until it modifies the orbits. If however the tumbling rates are significantly different, then the torques will reverse as the angle between the lobes changes, so the effects will cancel. Near-resonant orbits with a small difference of tumbling rates may nevertheless bring about significant changes before the difference can lead to any reversal of the torque.

- How does a backward torque affect a tumbling orbit?

The orbits follow a remarkable generic behaviour as described by Lynden-Bell and Kalnajs (1972) [76]: a given orbit accelerates when pulled backwards adiabatically and slows down when pulled forwards. The underlying physical property is that their adiabatic moment of inertia is generally negative. Therefore no collective alignment is to be expected from such “donkey” orbits.\(^2\) However, for the so called inner Lindblad resonance, the adiabatic moment of inertia is positive in the central parts of galaxies. The lobes of two orbital streams may then attract and cooperate, vibrating around exact alignment and so forming a deeper potential well that traps further orbits into that orientation, and so makes a bar.

The essence of these ideas are formalised in the following model.

### 2.2.1 Simplifying assumptions

Consider a set of ellipses or centred ovals rotating about their centre, which represents quasi-resonant inner Lindblad resonant orbits viewed from a rotating frame $\Omega_p$. They will all have the same characteristics (i.e. same geometry and same mass). Each ellipse will therefore be entirely specified by the orientation of its semi-major axis with respect to the rotating frame, $\varphi$, and the angular frequency $\Omega_\ell$.

The relative interaction of these ellipses will be assumed to derive from an effective alignment potential:

$$\psi_{12} = GA^2 \cos 2 (\varphi_1 - \varphi_2), \quad (2.2.1)$$

where $G$ is the gravitational constant, and $\varphi_1 - \varphi_2$ measures the relative azimuthal orientation of the two orbital lobes. In this model, $A^2$ is taken to be constant. The torque that ellipse “two” applies on ellipse “one” may then be written:

$$dh_1/dt = \partial\psi_{12}/\partial\varphi_1. \quad (2.2.2)$$

The response in angular velocity to such a torque can be derived via the moment of inertia $1/\alpha$ of the ellipse corresponding to the adiabatic moment of inertia of the inner Lindblad orbit this ellipse is supposed to represent, where $\alpha$ is given by

$$\alpha = \left(\frac{\partial\Omega_\ell}{\partial h}\right)_j, \quad (2.2.3)$$

so that

$$d\Omega_\ell/dt = \alpha \partial\psi_{12}/\partial\varphi_1. \quad (2.2.4)$$

Note that $\alpha$ may take both signs to account for the various orbital responses. Let $F^*(\varphi, \Omega_\ell, t)d\varphi d\Omega_\ell$ be the number of ellipses with orientation between $\varphi$ and $\varphi + d\varphi$ and with angular frequency

---

\(^2\) Collett, private communication nevertheless pointed out that a given orbit could possibly *anti-align* with the potential trough created by the other orbits, and widely oscillate, therefore spending most of its time at ninety degrees to the minimum of the potential, creating a pattern. Alternatively, he also pointed out that donkey orbits might be unstable with respect to the azimuthal equivalent of the two-stream instability in plasma physics.
Section 2.2: The Alignment Instability Model

Figure 2.2.1: The toy model problem: under what circumstances will an assembly of rotating centred ovals rotating about their centre evolve spontaneously towards a barred configuration?

between \( \Omega_\ell \) and \( \Omega_\ell + d\Omega_\ell \). The conservation of orbits implies that this distribution function satisfies a continuity equation in \((\varphi, \Omega_\ell)\) space

\[
\frac{\partial F^*}{\partial t} + \frac{\partial}{\partial \varphi} (\Omega_\ell F^*) + \frac{\partial}{\partial \Omega_\ell} \left( \frac{d\Omega_\ell}{dt} F^* \right) = 0.
\]

Assuming that \( \alpha \) is a constant of either sign, depending on the co-operative or “donkey” behaviour of the orbits, \( d\Omega_\ell/dt \) becomes independent of \( \Omega_\ell \), so \( F^* \) satisfies a Boltzmann equation:

\[
\frac{\partial F^*}{\partial t} + \Omega_\ell \frac{\partial F^*}{\partial \varphi} + \frac{\partial F^*}{\partial \Omega_\ell} \left( \alpha \frac{\partial \psi}{\partial t} \right) = 0.
\]

The density of ellipses which have orientation \( \varphi \) may then be defined by:

\[
\rho (\varphi, t) = \int F^* (\varphi, \Omega_\ell, t) d\Omega_\ell.
\]

The interaction potential generated by the “two” ellipses may be obtained by multiplying Eq. (2.2.1) with \( F^*_2 \) and integrating over their orientations

\[
\psi (\varphi, t) = GA^2 \int \rho (\varphi_2, t) \cos 2 (\varphi - \varphi_2) d\varphi_2.
\]

The stationary axisymmetric unperturbed state obeys \( F^* = F^*_0 (\Omega_\ell) \) and \( \rho = \rho_0 = \text{constant} \). Therefore \( \psi \) is identically zero and Eq. (2.2.5), Eq. (2.2.6) and Eq. (2.2.7) are trivially satisfied. What is the stability of such a configuration to bi-symmetrical instabilities?
Writing \( F^* = F_0^* + f^* \), and linearising the equation with respect to the perturbation \( f^* \) and \( \psi \), Eq. (2.2.6) becomes
\[
\frac{\partial f^*}{\partial t} + \Omega_f \frac{\partial f^*}{\partial \varphi} + \alpha \frac{\partial F_0^*}{\partial \Omega_f} \frac{\partial \psi}{\partial \varphi} = 0 .
\] (2.2.9)

### 2.2.2 Instability criterion

The Fourier expansion of \( f^* \) with respect to \( \varphi \) reads
\[
f^* = \sum_{m=-\infty}^{\infty} e^{im\varphi} f_m^* (\Omega_f, t) ,
\] (2.2.10)

Similarly
\[
\rho = \sum_{-\infty}^{\infty} e^{im\varphi} \rho_m , \quad \text{where} \quad \rho_m = \int f_m^* d\Omega_f .
\] (2.2.11)

Substituting Eq. (2.2.11) into (2.2.8) gives:
\[
\psi = \pi G A^2 \left( \rho_2 e^{i\varphi_2} + \rho_{-2} e^{-i\varphi_2} \right) .
\] (2.2.12)

When the perturbation has no \( |m| = 2 \) component, Eq. (2.2.8) becomes
\[
\frac{\partial f^*}{\partial t} + \frac{\partial f^*}{\partial \varphi} = 0 .
\] (2.2.13)

The general solution is \( f^* = g(\varphi - \Omega_f t) \) where \( g \) is an arbitrary function without an \( |m| = 2 \) component. As Eq. (2.2.9) is linear, its complete solution corresponds to the sum of this solution (which propagates in \( (\Omega, \varphi) \) space) and a solution for the resonant modes \( |m| = 2 \). These modes are responsible for the gravitational instability. For growing modes \(^3 \), \( f_m^\star \propto e^{i\omega t} \), where \( \omega \) has a negative imaginary part. From Eqs. (2.2.9), (2.2.10) and (2.2.12), it follows that
\[
i (\omega + m\Omega_f) f_m^* = -i m \pi G \rho_m A^2 \alpha \frac{\partial F_0^*}{\partial \Omega_f} , \quad \text{for} \quad |m| = 2 .
\] (2.2.14)

Dividing by \( i(\omega + m\Omega_f)/A \), and integrating over all \( \Omega_f \) so as to leave \( \rho_m \) on the r.h.s. leads to the identity
\[
1 = \pi G \int -\frac{\alpha A^2 \partial F_0^*}{\partial \Omega_f} \frac{\partial \Omega_f}{\Omega_f - \Omega_p} d\Omega_f ,
\] (2.2.15)

where \( \Omega_p = -\omega/m \) (which corresponds to the angular frequency at which the perturbation is propagating). Recall that \( \Omega_p \) has a positive imaginary part (for \( m = 2 \)) corresponding to the growth rate of the instability. Marginal stability is therefore reached as the imaginary part of \( \Omega_p \) vanishes. The integral in Eq. (2.2.15) may then be written as the sum of a Cauchy principal part, and a half residue corresponding to the pole at \( \Omega_f = \Omega_p^\star \):
\[
1 = \pi G \int -\frac{\alpha A^2 \partial F_0^*}{\partial \Omega_f} \frac{\partial \Omega_f}{\Omega_f - \Omega_p} d\Omega_f - i\pi^2 G \left[ \frac{\alpha A^2 \partial F_0^*}{\partial \Omega_f} \frac{\partial \Omega_f}{\Omega_f} \right]_{\Omega_f = \Omega_p^\star} ,
\] (2.2.16)

where \( \Omega_p \) is now real. Identifying real and imaginary parts, it follows that
\[
1 = \pi G \int -\frac{\alpha A^2 \partial F_0^*}{\partial \Omega_f} \frac{\partial \Omega_f}{\Omega_f - \Omega_p} d\Omega_f ,
\] (2.2.17a)
\[
\frac{\partial F_0^*}{\partial \Omega_f} = 0 \quad \text{for} \quad \Omega_f = \Omega_p .
\] (2.2.17b)

Eq. (2.2.17b) fixes \( \Omega_p \), the pattern speed of the bi-symmetrical instability.

---

\(^3\) Since this analysis concentrates on a dispersion relation for marginally stable modes, rather than transient behaviour, there is no need to Laplace transform Eq. (2.2.12), and so get the instability criterion by avoiding the poles, as is usually done for Landau damping.
Section 2.2: The alignment instability model

2.2.3 Discussion: azimuthal Jeans instability

Candidates for the angular frequency of the perturbation are given by the extrema of $F_0^\ast$. Eq. (2.2.17a) tells which (if any) will be dominant. The pattern speed of the perturbation as well as of its stability is therefore a characteristic of the unperturbed galaxy, as illustrated by Fig. 2.2.2\textsuperscript{4}. When $\alpha$ is positive\textsuperscript{5}, Eq. (2.2.16) may be re-arranged as

$$\sigma_\Omega^2 \leq \frac{1}{2} G M \alpha ,$$

(2.2.18)

where $\sigma_\Omega^2$ measures the weighted dispersion of the distribution function in the neighbourhood of the resonance

$$\frac{1}{\sigma_\Omega^2} = \int \left[ \frac{A^2 \alpha \partial F_0^\ast / \partial \Omega_\ell}{\Omega_p - \Omega_\ell} \right] d\Omega_\ell / \int A^2 \alpha F_0^\ast d\Omega_\ell .$$

(2.2.19)

Equation (2.2.18) is very similar to a Jeans instability criterion. It implies that the dispersion in $\Omega_\ell$ around $\Omega_p$ should be less than an effective $G$ density, built from the adiabatic moment of inertia $\alpha$, and $GA^2$, the amplitude of the effective interaction potential. The inverse adiabatic moment of inertia, $\alpha$, scales like [Length\textsuperscript{-2}] while $A^2$ scales like [Length\textsuperscript{-1}]. As $\sigma_\Omega^2$ measures the dispersion in angular velocity in the frame of the maximum of $F_0^\ast(\Omega_\ell)$, the criterion (2.2.19) implies that the lower the dispersion, the more efficient the instability.

\textsuperscript{4} This description was suggested to me by Collett who derived independently the instability criterion (2.2.17) in 1987. Earn [30] also derives (2.2.17).

\textsuperscript{5} here, the equivalent to the two stream instability corresponds to a negative $\alpha$ and a minimum of the distribution function, $\partial^2 F_0^\ast / \partial \Omega_\ell^2 \geq 0$, i.e. few repealing orbits widely librating so as to anti-align to the potential valley.
2.3 The inner Lindblad resonance model

2.3.1 Orbital streams instability criterion

Consider a galaxy made of orbital streams of inner Lindblad resonances. Each orbital stream is characterised by its orientation, $\varphi$, its specific angular momentum, $h$, and its circulation $J = 4\pi (J_R + 1/2 h)$. As such streams interact gravitationally, $\varphi$ and $h$ change, but $J$ is taken to be adiabatically invariant. Indeed, it was shown in section 2.2 that the onset of an orbital instability required a narrow dispersion in the angular velocity of near resonant orbits. Under this assumption, the period corresponding to the relative libration of these orbits around their mean pattern speed, $\Omega_p$, should be large as they almost resonate. Many orbits are found for which the ratio of the orbital period to the oscillation period is very small. It is therefore appropriate to construct the corresponding adiabatic invariant $J$. Strictly speaking, $J$ is only an invariant for the quasi-resonant orbits. On the other hand, non-resonant orbits are not expected to play any significant role in restructuring the galaxy. It is also assumed in this model that any two orbital streams interact through their mutual potential energy, the angular dependent part of which is approximated by $\Psi_{12}$ where

$$\Psi_{12} = GA_1A_2 \cos 2(\varphi_1 - \varphi_2). \quad (2.3.1)$$

Here, $\varphi_1 - \varphi_2$ is the relative azimuth of the orbital lobes. The amplitude of the alignment potential must vanish if either orbit is circular, and for small ellipticities it should be proportional to their product. The critical (and somewhat arbitrary) assumption in this section is that the amplitude depends on the orbits through a product $A_1A_2$ with $A_1$ depending on orbit one only and $A_2$ on orbit two. While such a split is unlikely to be strictly satisfied – and indeed it is shown in the appendix not to be true – it should give a good qualitative account for the interaction between similar resonant orbits which are responsible for the type of orbital instability described in section 2.2.

The specific torque on stream “one” due to unit mass on stream “two” then reads

$$dh_1/dt = \partial \Psi / \partial \varphi_1. \quad (2.3.2)$$

Let $F^*(\varphi, \phi, h, J, t)d\varphi d\phi dh dJ$ be the mass weighted distribution function of stars belonging to an orbital stream with lobe azimuths in the range $\varphi$ to $\varphi + d\varphi$, with orbital phase between $\phi$ and $\phi + d\phi$, with angular momenta in the range $h$ to $h + dh$ and with circulation between $J$ and $J + dJ$. This distribution function obeys a Boltzmann equation corresponding to the conservation of mass in phase space. Assuming that the only relevant resonance is the inner Lindblad one (i.e. all other resonances between orbits and in the relative orbital phases are disregarded), an averaging principle may then be applied to this equation in order to write a Boltzmann equation for the number of orbital streams. These are the relevant objects for this analysis according to the above assumptions. Calling $F = \int F^* d\phi$, it follows that:

$$\frac{\partial F}{\partial t} + \Omega_p \frac{\partial F}{\partial \varphi} + \frac{\partial \Psi}{\partial h} \frac{\partial F}{\partial h} = 0. \quad (2.3.3)$$

---

$^6$ See Appendix A & Lynden-Bell (1972) [76] for an introduction to angle and action variables.
With this definition, $F$ is a distribution function in phase-space since $(\varphi, h)$ are canonically conjugate variables. The orientation potential, $\Psi$, due to all the orbital streams is found by weighting Eq. (2.3.1) with $F_2$ and integrating over all streams “two”:

$$\Psi (\varphi, h, J, t) = G A(h, J) \frac{1}{2} \int \rho (\varphi, t) \cos 2(\varphi - \varphi_2) d\varphi_2,$$  \hspace{1cm} (2.3.4)

where $\rho$ is the “effective” density of streams at orientation $\varphi$, and is defined by

$$\rho (\varphi, t) = \int A(h, J) F (\varphi, h, J, t) dh dJ.$$ \hspace{1cm} (2.3.5)

Note that with this set of variables, $\Psi$ depends on $h$ and $J$ as well as orientation because the torque applied on a given orbit depends on the shape and size of that orbit. Unperturbed axially-symmetric steady states have $F = F_0(h, J)$. Writing $F = F_0 + f$ and linearising in the perturbed quantities $f$ and $\psi$ equation Eq. (2.3.3) becomes

$$\frac{\partial f}{\partial t} + \Omega \frac{\partial f}{\partial \varphi} + \frac{\partial \Psi}{\partial \varphi} \frac{\partial F_0}{\partial h} = 0.$$ \hspace{1cm} (2.3.6)

Note that

$$\frac{\partial F_0}{\partial h} = \frac{\partial F_0}{\partial \Omega} \cdot \alpha,$$ \hspace{1cm} (2.3.7)

Equation (2.3.6) may then be formally identified with Eq. (2.2.9). Equation (2.3.5) is the stellar equivalent of Eq. (2.2.7). The sum over $d\Omega$ becomes here a double sum over $dh dJ$. The assumption that the function $A$ is separable with respect to the characteristics of orbits “one” and “two” leads to a solution which is derived following precisely the same steps as that of section 2.2 and reads

$$1 = \pi G \int \int \frac{A^2(h, J) [-\alpha \frac{\partial F_0}{\partial \Omega}]}{\Omega - \Omega_p} dh dJ.$$ \hspace{1cm} (2.3.8)

### 2.3.2 Discussion: alignment and distortion

How does the criterion (2.3.8) relate to the toy model criterion, Eq. (2.2.17)?

Defining

$$F^\ell (\Omega, J) = F_0(J, h [\Omega, J]),$$ \hspace{1cm} (2.3.9)

the integrand in Eq. (2.3.8) may be integrated by parts

$$\int - (\frac{\partial F^\ell}{\partial \Omega}) A^\ell_J^2 dJ = - \frac{d}{d\Omega} \left( \int F^\ell A_J^2 dJ \right) - \int F^\ell (\partial A_J^2 / \partial \Omega) dJ.$$ \hspace{1cm} (2.3.10)

The square brackets correspond to the limits of integration, for which the argument vanishes because these are non-interacting orbits (circular orbits) $[F^\ell A_J^2] = 0$. Calling

$$F^\ell_0 (\Omega) = 1/\left< A^2 \alpha \right> \int F^\ell (\Omega, J) A_J^2 dJ,$$ \hspace{1cm} (2.3.10a)

$$\frac{d}{d\Omega} \delta F^\ell_0 (\Omega) = -1/\left< A^2 \alpha \right> \int F^\ell (\Omega, J) \partial A_J^2 / \partial \Omega dJ,$$ \hspace{1cm} (2.3.10b)

equation (2.3.8) then becomes

$$\frac{1}{\pi G} = \left( \int - \frac{\alpha A^2 > d/d\Omega (F^\ell_0 + \delta F^\ell_0)}{\Omega - \Omega_p} d\Omega \right),$$ \hspace{1cm} (2.3.11)
Orbital instabilities and bar formation Chapter 2

in direct analogy with Eq. (2.2.15). Here the constant $<A^2\alpha>$ corresponds to the average $\int F^t A^2_t d\Omega d\ell$ in order to identify Eq. (2.3.8) with Eqs. (2.2.17). Note that $F^t d\Omega dJ = \alpha F_0 dh dJ$. Therefore $F^t d\Omega dJ$ is the distribution of orbital streams with angular frequency in the range $\Omega_i + d\Omega$ and invariants between $J$ and $J + dJ$ weighted by the corresponding adiabatic moment of inertia. $F_0(\Omega_i)$ is therefore the “effective” mass density of resonant orbits at the frequency $\Omega_i$. As $\alpha$ generally changes sign, the above summation is algebraic. Recall that negative $\alpha$ correspond to those inner Lindblad orbits which fail to cooperate. The integrand in Eq. (2.3.10b) may be rewritten as: $\delta^* (\partial A^2 / \partial \Omega_i) = \delta^* (\partial A^2 / \partial h) (\partial h / \partial \Omega_i)$. But $(\partial A^2 / \partial h) (\partial h / \partial \Omega_i)$ measures the distortion – variation in size and eccentricity – of the resonant orbit when adiabatically pulled forward, and $(\partial h / \partial \Omega_i)$ cancels with the implicit weighting by $\alpha$ in Eq. (2.3.10). Therefore $\delta F^*$ is the correction to the effective mass density of orbits induced by the distortion of the resonant orbits. In summary, for a galaxy made of orbital streams of inner Lindblad resonant orbits, it was shown that as the orbits align, they are distorted by the growing potential.

2.4 Adiabatically relaxed bars

Once a bi-symmetric instability has occurred – triggered by an instability of orbital nature or indeed by other source of instability – it is worth analysing the most likely path of evolution. Numerical simulations (e.g. Sellwood 1993 [101] and references enclosed) predict a characteristic secular evolution of the amplitude and the pattern speed of the bar beyond the linear regime described in the previous section. Such a trend is investigated here using the framework of thermodynamics. The state of maximum entropy compatible with total energy and angular momentum conservation corresponds to uniform rotation. In practice, this state is not reached because of the preservation of other quasi-invariants such as the circulations of section 2.3 which are kept adiabatically constant. The distribution function corresponding to the extremum of entropy, given these supplementary constraints, is derived here. For that purpose, it is assumed that the underlying axisymmetric component of the galaxy is left unchanged as a bar grows. This hypothesis is only exact in the linear régime of perturbation theory. Here, it will nevertheless be assumed to hold beyond that régime.

2.4.1 The coupling energy

The total energy, $E$, is the sum of the kinetic energy, $K$, and the gravitational energy of interaction, $V$. When only one resonance is relevant, $E$ may be formally re-arranged as two components:

- the energy arising from the orbits distributed axisymmetrically:

$$E_{\text{axi}} = K + V_0 = \int \varepsilon(h, J) F(J, \varphi) dJ dh d\varphi - V_0,$$  \hspace{1cm} (2.4.1)

where $\varepsilon(h, J)$ is the energy of the orbit $(h, J) = J$ as defined by the axisymmetric component of the potential, and $V_0$ is the axisymmetric potential energy which is subtracted here to avoid counting its contribution twice;

---

7 this result is not surprising since a linear theory should include both the alignment and the distortion independently

8 In any circumstance, the adiabatic re-alignment of elongated $x_1$ orbits is believed to re-enforce existing bars.
Section 2.4 Adiabatically relaxed bars

- the non-axisymmetric component corresponding to the contribution from the interaction of quasi-resonant inner Lindblad orbits:

\[ V_2 = -\frac{1}{2} \int F(J, \phi) \Psi(J, \phi) \, d\phi \, d^2J, \tag{2.4.2} \]

where

\[ \Psi(J, \phi) = \int A_2(J, J') F(J', \phi') \cos 2(\phi - \phi') \, d\phi' \, d^2J'. \tag{2.4.3} \]

The function \( F(J, \varphi) = \int F(J, \varphi, \phi) \, d\phi \) is the distribution function of resonant orbital streams with adiabatic invariant \( J \), angular momentum \( h \), and orientation \( \varphi \) which are taking part in the instability. The expression \( \Delta E \), for the total energy measured with respect to the axisymmetric potential energy follows

\[ \Delta E = E - V_0 = \int \left[ \varepsilon(J) - \frac{1}{2} \Psi(J, \varphi) \right] F(J', \varphi') \, d^2J \, d\varphi, \tag{2.4.4} \]

Equation (2.4.4) together with Eq. (2.4.3) is assumed to represent an adequate description of the energy of the system. Hereafter \( \Delta E \) shall be referred to as the energy, and is assumed to be constant in this section. Its functional expression avoids the divergences usually found in the statistical mechanics of gravitating systems. The energy of interaction remains finite for overlapping orbits, and no orbit escapes the galaxy. In fact, the orbits become circular as the star reaches its maximum specific momentum at a given circulation, and so the orbit stops contributing to the instability.

### 2.4.2 States of extremum entropy

The Boltzmann entropy for an assembly of orbital streams reads

\[ S = -\int F(J, \varphi) \log F(J, \varphi) \, d^2J \, d\varphi, \tag{2.4.5} \]

in units of Boltzmann’s constant set to unity. Angular momentum is exchanged between resonant orbits, but strict invariance is assumed for the adiabatic action corresponding to the motion of a given star along its orbit. All possible variations of \( F \) are explored, subject to the constraint that \( \psi_{\text{axi}} \) is left unchanged. The total energy, \( \Delta E \), the total angular momentum, \( H \), and the mass density of orbits, \( n(J) \, dJ \), which have a circulation in the range \( J, J + dJ \) are all supposed to be conserved in the process. The latter hypothesis holds as long as the conditions of adiabatic invariance hold. Realistic values of \( n(J) \) may be derived from the axisymmetric distribution functions, \( F_0(\varepsilon, h) \), constructed in the next chapter. Its behaviour is illustrated on Fig. 2.4.4 for the Isochrone potential in the epicyclic approximation as described in the next subsection. Explicitly,

\[ n(J) = \int F_0(\varepsilon[J, h], h) \, dh, \tag{2.4.6} \]

where \( \varepsilon[J, h] \) is given implicitly by

\[ J = \frac{1}{2} \theta + \frac{1}{\pi} \int \sqrt{2(\psi_0(R) + \varepsilon) - \frac{h^2}{R^2}} \, dR. \tag{2.4.7} \]

---

9 for the Hamiltonian \( \varepsilon(J) - \Psi(J, \varphi) \), the canonical angle coupled to \( J \) is ignorable, therefore \( J \) is formally a strict invariant by construction.

10 In fact, for collisionless systems, the density in phase space described by \( F \) is also conserved following the motion; however, the coarse-grained distribution will evolve in time.
These conserved quantities put three constraints on the variation of $F$, namely:

\[ n(J) = \int F(h, J, \varphi) \, d\varphi \, dh, \quad \text{(2.4.8a)} \]
\[ H = \int F(h, J, \varphi) \, h \, d\tau, \quad \text{(2.4.8b)} \]
\[ \Delta E = \int F(h, J, \varphi) \left[ \varepsilon(h, J) - \frac{1}{2} \psi(h, J, \varphi) \right] \, d\tau, \quad \text{(2.4.8c)} \]

with \( d\tau = d\varphi \, dh \, dJ \), the phase space volume element. The variation of $V_2$ in Eq. (2.4.8c) via Eq. (2.4.3) may be carried out by switching indices in the double integration:

\[ \delta V_2 = -\frac{1}{2} \int \int A_2(J, J') \left( \delta F F' + F \delta F' \right) \cos \left[ 2(\varphi - \varphi') \right] \, d\tau \, d\tau', \quad \text{(2.4.9a)} \]
\[ = -\int \delta F(J, \varphi) \, \Psi(J, \varphi) \, d\tau. \quad \text{(2.4.9b)} \]

Note that the \( 1/2 \) factor of Eq. (2.4.9a) required to avoid counting all pairs of interaction twice yields the correct potential in Eq. (2.4.9b). It follows that

\[ \delta S = 0 = \int \delta F \left[ \ln F(J, \varphi) + 1 - \alpha(J) + \beta \varepsilon(J) - \beta \Psi(J, \varphi) - \beta \Omega_p h \right] \, d\tau, \quad \text{(2.4.10)} \]

where the Lagrange multipliers $\alpha(J)$, $\beta \Omega_p^{11}$ and $\beta$ account for the constraints Eqs. (2.4.8).

Solving for all variations of $\delta F$ gives

\[ F(J, \varphi) = C(J) \exp \left[ -\beta \{ \varepsilon(J) - \Psi(J, \varphi) - \Omega_p h \} \right]. \quad \text{(2.4.12)} \]

Equation (2.4.12) is not a surprise; the distribution function is a function of the Jacobi invariant, $\varepsilon(J) - \Psi(J, \varphi) - \Omega_p h$, which corresponds to the energy of an orbit measured in the frame which rotates at the frequency $\Omega_p$.

A solution must be found self-consistently as $F$ is a functional of $\Psi$ which in turn is a functional of $F$:

\[ \Psi(J, \varphi) = \int A_2(J, J') \, C(J') \exp \left[ -\beta \{ \varepsilon(J') - \Omega_p h - \Psi(J', \varphi') \} \right] \cos \left[ 2(\varphi - \varphi') \right] \, d\tau'. \quad \text{(2.4.13)} \]

\[ ^{11} \text{Collett and Lynden-Bell (1989) [18] have shown that } \Omega_p, \text{ the Lagrange multiplier associated with conservation of angular momentum can be identified with the pattern speed of the perturbation. The proof is derived while requiring that the energy of a barotropic flow is stationary with respect to all displacements preserving circulations, total angular momentum and number of particles. In fluid dynamics, the flow is characterised by its velocity field, } \mathbf{u}, \text{ its pressure, } p, \text{ and density, } \rho, \text{ responsible for the gravitational potential } \psi. \text{ The first order change in the energy then reads} \]

\[ \Delta E = \Omega_p \Delta H + \int \int \xi_1 \rho \cdot \nabla \cdot d^2r - \int \int \int \xi_1 \nabla \cdot \rho \, d^2r \]
\[ + \int \int \int \xi_2 \left[ -\mathbf{v} \times \nabla + \nabla \left( \frac{1}{2} \mathbf{v}^2 + \int \frac{dp}{\rho} - \psi + \frac{1}{2} (\Omega_p \times \mathbf{r})^2 \right) \right] \, d^2r, \quad \text{(2.4.11)} \]

where the velocity, \( \mathbf{v} = \mathbf{u} - \Omega_p \times \mathbf{r} \), corresponds to the velocity of the flow measured in a frame which rotates at the frequency $\Omega_p$, and $\omega = (\nabla \times u)$ is the vorticity of the flow. The displacements, $\xi_1$ and $\xi_2$ may be varied independently given the conservation of circulations. At constant total angular momentum, and constant total number of stars, the first three terms in Eq. (2.4.10) vanish for all displacements $\xi_1$. The integrand of the remaining term corresponds to the equation of motion of a steady flow in a coordinate frame rotating with angular velocity $\Omega_p$. Requiring that this term vanishes for all displacements $\xi_2$ implies that the Lagrange multiplier $\Omega_p$ is the frequency of the frame in which the flow is stationary, i.e. the pattern speed. Note that Eq. (2.4.10) also provides an value for the pattern speed directly from the adiabatic variation of $E$ with respect to $H$, namely $\Omega_p = (\partial E/\partial H)_J$. A similar proof holds for stellar dynamics.
This solution is easily found under the assumption that \( A_2(J, J') \) can be approximated as a product of \( A(J) \) by \( A(J') \) as in section 2.3. Indeed Eq. (2.4.3) then becomes:

\[
\Psi (J, \varphi) = A(J) \cos (2 \varphi) K, \tag{2.4.14}
\]

where \( K \), the “effective” bi-symmetric over-density, is defined to be

\[
K = \int F(J, \varphi) A(J) \cos (2 \varphi) \, d\tau. \tag{2.4.15}
\]

All angles are measured with respect to the maximum of the over density which is assumed to be symmetrical in \( \varphi \). Inserting Eq. (2.4.12) for \( F \) in Eq. (2.4.15) yields, self-consistently,

\[
K = \int AC \exp \left[ -\beta (\varepsilon - \Omega_p h - A \cos (2 \varphi) K) \right] \cos (2 \varphi) \, d^2J \, d\varphi, \tag{2.4.16a}
\]

\[
= 2\pi \int A(J) C(J) I_1 [\beta A(J) K] e^{-\beta \varepsilon(J)+\beta \Omega_p h} \, d^2J. \tag{2.4.16b}
\]

where \( I_n \) is the Bessel function of order \( n \). Equation (2.4.16) may be re-written formally as

\[
\frac{K}{N} = \frac{\langle A(J) I_1 [\beta A(J) K] \rangle}{\langle I_0 [\beta A(J) K] \rangle} \equiv \mathcal{G}(\beta K). \tag{2.4.17}
\]

The bracket \( \langle \rangle \) denotes the normalised average over phase space weighted by the distribution function \( F(J, \varphi) \). Non-trivial solutions of this implicit equation correspond to the relation between the amplitude of the bar and the temperature of the system. A critical temperature, \( 1/\beta_{\text{crit}} \), may be defined in terms of \( \mathcal{G} \) and its derivatives below which Eq. (2.4.17) has more than one trivial root \( (K \equiv 0) \) is always a solution given that \( I_1[0^+] / I_0[0^+] = 0^+; I_1[x] / I_0[x] \) tends asymptotically to 1 as \( x \to \infty \); therefore the two curves \( y = N^{-1}x/\beta \) and \( y = \mathcal{G}(x) \) must intersect for sufficiently high values of \( \beta \).

The actual values for the Lagrange multipliers \( C(J) \), \( \beta \) and \( \Omega_p \) may be derived from the constraints, Eqs. (2.4.18). These may be re-arranged as

\[
n(J) = 2\pi C(J) \int I_0 [\beta A(J) K] e^{-\beta \varepsilon(J)+\beta \Omega_p h} dh, \tag{2.4.18a}
\]

\[
\langle h \rangle \equiv \frac{H}{N} = \frac{\partial \log N}{\partial \gamma}, \quad \text{where} \quad \gamma = \Omega_p \beta, \tag{2.4.18b}
\]

\[
\langle \varepsilon \rangle \equiv \frac{\Delta E}{N} = -\frac{\partial \log N}{\partial \beta} + \frac{K^2}{2}, \tag{2.4.18c}
\]

where the auxiliary function, \( N \), corresponds to the total number of orbits:

\[
N = \int F(J, \varphi) \, d^2J \, d\varphi = 2\pi \int C(J) I_0 [\beta A(J) K] e^{-\beta \varepsilon(J)+\beta \Omega_p h} \, d^2J. \tag{2.4.19}
\]

In Eqs. (2.4.18), the partial derivation is done holding constant the other Lagrange multipliers, the boundaries of integration and \( K \), the effective number of orbits caught in the bar. Eqs. (2.4.16) and (2.4.18) yield the relationship between the pattern speed of the bar and its amplitude as a function of the temperature, which follows from the macroscopic properties of the disk (mean energy and angular momentum, initial distribution of orbital streams with a given circulation).
Chapter 2

Orbital instabilities and bar formation

0.1
0.2
0.3
0.4
0.5
0.6
1/β

0
0.2
0.4
0.6
0.8
1

-1.5
-1
-0.5
0
0.5
1

-1
-0.5
0
0.5
1

Figure 2.4.1: the amplitude of the bar $K/\bar{A}N$ (left panel) as a function of the temperature $(\beta N \bar{A}^2)^{-1}$. The width of the bar – defined here by the separation between the zeros of the reduced autocorrelation $\gamma(\phi) = \langle \rho(\phi + \varphi)\rho(\varphi) \rangle / \langle \rho(\varphi) \rangle^2$ and plotted on the right panel as a function of $\phi$ – decreases slowly as a function of $\beta \bar{A} K = 0.5, 1, 2$.

2.4.3 A simple example: rigid rods

Consider first the thermodynamics of identical rigid rods. For these objects, both the moments of inertia, $I$, and the amplitude of interaction, $\bar{A}^2$, are assumed to be constant. Condition Eq. (2.4.17) then becomes

$K = \frac{I_1}{I_0} (\beta \bar{A} K) \bar{A} N$.  

Equation (2.4.20) will have non-trivial solutions if

$\beta \bar{A}^2 N \geq \frac{1}{(I_1/I_0)(0)} = 2$.  

This equation gives the critical temperature, $\bar{A}^2 N/2$, below which the rods start aligning as illustrated on Fig. 2.4.1. The kinetic energy of each rod reads $I^{-1}h^2/2$, which implies that

$N = C \sqrt{\frac{2\pi I}{\beta}} \exp \left( \frac{1}{2} I \Omega_p^2 \right) I_0 (\beta \bar{A} K)$,  

therefore Eqs. (2.4.34) become $\Omega_p = \langle h \rangle/I$ and $\langle e \rangle = I \Omega_p^2/2 + \beta^{-1}/2 - N^{-1} K^2/2$. The pattern speed of the aligned rods corresponds to the mean angular frequency. The energy versus temperature variations are illustrated on Fig. 2.4.2. The specific heat at constant total momentum, $C_H = -\beta^2 \partial \langle e \rangle / \partial \beta$, is given by $C_H = 1/2 + \beta^2 K \partial K / \partial \beta$. It is positive everywhere, which is a property of one dimensional self-gravitating systems\textsuperscript{12}. Note that $C_H$ is discontinuous at the

\textsuperscript{12} This property is best understood by contrasting the qualitatively different behaviour of gravity in one and two dimensions; consider a test particle moving in the potential well $\psi$ of a central mass concentration. In two dimensions, the test particle describes a rosette while oscillating radially between perigee and apogee. In one dimension, the particle oscillates symmetrically between $-x_{\text{max}}$ and $x_{\text{max}} = \psi^{-1}(-\epsilon)$. If energy is retrieved from that particle, it readjusts its trajectory to a state of narrower oscillation, corresponding to lower mean kinetic energy. In two dimensions, the particle re-adjusts its trajectory to a more eccentric Lissajous figure which, contrary to the one dimensional case, has a higher mean kinetic energy. If this increase in energy more than counter balances the energy loss induced by the self-consistent change in potential energy, an ensemble of self-gravitating particles displays negative specific heat.
Section 2.4  Adiabatically relaxed bars

17

Figure 2.4.2: the energy of the bar, (left panel), and its specific heat, (right panel), as a function of the temperature \((\beta N A^2)^{-1}\). Note the discontinuity of \(C_H\) – inducing a kink in the energy – at the critical temperature.

critical temperature as shown on Fig. 2.4.2, which is characteristic of a phase transition described in the mean field approximation.

2.4.4 Low temperature expansion of the general case

The temperature, \(1/\beta\), measures the relative dispersion in the distribution of orbits with angular momentum \(h\). In the low temperature limit, the régime in which the bar forms, this dispersion is small and the integration over \(h\) is well estimated using a steepest descent approximation (cf Fig. 2.4.3). This in effect describes the integrand of Eq. (2.4.19) as a truncated Gaussian centred on the maximum of this integrand. Equation (2.4.19) then reads

\[
N = 2\pi \int C(J) e^{-\beta(\tilde{\epsilon} - \Omega_p \tilde{h})} I_0(\beta \tilde{A} K) \sqrt{\frac{2\pi I}{\beta}} dJ, \tag{2.4.23}
\]

where \((\tilde{\ })\) corresponds to \((\ )\) evaluated at \(\tilde{h}\), and \(\tilde{h}(J, \Omega_p, \beta)\) is the angular momentum corresponding to the extremum of \(F(h, J)\) at fixed \(J\). In Eq. (2.4.23), \(\tilde{I}\) corresponds to an effective – temperature dependent – moment of inertia given by

\[
\tilde{I} = \tilde{I}^* \left( \text{erf} \left( \sqrt{\frac{\beta}{2\tilde{I}^*}} (h_{\max}(J) - \tilde{h}) \right) \right)^2, \quad \text{where} \quad \tilde{I}^* = \left[ \frac{\partial^2 \tilde{\epsilon}}{\partial h^2} - \frac{1}{\beta} \frac{\partial^2 \log I_0}{\partial h^2} \right]^{-1}, \tag{2.4.24}
\]

and \(\text{erf}(x) = 1/\sqrt{\pi} \int_0^x \exp(-t^2) dt\). The erf function in Eq. (2.4.24) accounts for the correction of the moment of inertia \(\tilde{I}^*\) induced by a maximum for the angular momentum reached by the circular orbits rotating at the frequency \(\Omega_p\) with adiabatic action \(J\). In Eq. (2.4.24), \(h_{\max} = 2J\) is the angular momentum of these circular orbits with invariant \(J\). Note that \(A\) vanishes for these orbits (characterised by \(J_R = 0 = J - h_{\max}/2\)). In Eq. (2.4.24), \(1/\tilde{I}^*\) is the sum of two contributions: \(\alpha\), the donkey parameter, and \(\beta^{-1} \partial^2 \log I_0/\partial h^2\) which accounts for the fact that \(A\) varies with \(h\); this contribution scales like \(\beta\) at low temperature which implies that the adiabatic
distortion of the orbits dominates the last stage of the collapse. The value of $\bar{h}(\Omega, J)$ is given by the smaller\(^{13}\) root of

$$\Omega_\ell (\bar{h}) \equiv \left( \frac{\partial \bar{\epsilon}}{\partial h} \right) = \Omega_p + \frac{1}{\beta} \left( \frac{\partial \log I_0}{\partial h} \right) = \Omega_p + \left( \frac{\partial A}{\partial h} \right) \frac{I_1}{I_0} K, \quad (2.4.25)$$

which fixes the pattern of the bar, $\Omega_p$, for a given $\bar{h}$. Note the similarity between Eq. (2.4.25) and the value of $\Omega_p$ predicted by the linear theory of section 3: both analyses yield a pattern corresponding to the maximum of the effective number of orbits rotating at the frequency $\Omega_\ell$, once a correction has been made to account for the orbit adiabatic distortion by the bar. In fact, $(\partial A/\partial h)$ is generally negative which implies that patterns speed above the maximum of $\Omega_\ell$ may be reached while the orbits are adiabatically distorted.

Putting Eq. (2.4.23) into Eqs. (2.4.18) yields after some algebra\(^{14}\)

$$\langle h \rangle = \frac{1}{N} \int n(J) \bar{h} \left[ 1 + \frac{1}{2\beta \Omega_p h} \frac{\partial \log \bar{T}}{\partial \log \Omega_p} \right] dJ, \quad (2.4.26a)$$

$$\langle \epsilon \rangle = \frac{1}{N} \int n(J) \left[ \bar{\epsilon} + \frac{1}{2\beta} \left( 1 - \frac{\partial \log \bar{T}}{\partial \log \beta} \right) - \frac{K^2}{2N} \right] dJ. \quad (2.4.26b)$$

\(^{13}\) Equation (2.4.25) has generally two roots corresponding to the inner-inner Lindblad resonance and the outer-inner Lindblad resonance respectively. The latter is ignored in this analysis as it corresponds to donkey orbits which anti-align.

\(^{14}\) Equation (2.4.17) was used to write $\int (\partial \log I_0/\partial J) n(J) dJ = \int AK n(J) I_1/I_0 dJ = K^2/N$. Equation Eq. (2.4.25) is also taken in to account.
Equations (2.4.26), together with Eq. (2.4.25) fix the temperature and the pattern speed of a bar embedded in a galaxy with specific angular momentum $\langle h \rangle$, specific energy $\langle e \rangle$ and circulation distribution $n(J)$. All thermodynamic properties of the bar follow from Eqs. (2.4.26). For instance, the adiabatic specific heat at constant total momentum, $C_H = -\beta^2 \partial \langle e \rangle / \partial \beta$, is given by

$$C_H = \frac{1}{2N} \int n(J) \left[ \left( 1 + \frac{\partial \log \bar{T}}{\partial \log \beta} \right) - \left( \frac{\partial^2 \log \bar{T}}{\partial (\log \beta)^2} \right) - 2\beta^2 \frac{\partial \bar{\varepsilon}}{\partial \beta} + 2\beta^2 \frac{K}{N} \frac{\partial K}{\partial \beta} \right] dJ, \quad (2.4.27)$$

where $\partial K/\partial \beta \geq 0$ is given implicitly by Eq. (2.4.17). Here $\bar{\varepsilon}$ is a function of $\beta$ via $\Omega_p = \Omega_{p}[\beta, K(\beta)]$. The variation of the effective moment of inertia of the orbits is crucial in fixing the sign of $C_H$, and the values of $\beta$ and $\Omega_p$. Now

$$\frac{\partial \log \bar{T}}{\partial \log \Omega_p} = \frac{\partial \log \bar{T}^*}{\partial \log \Omega_p} (1 - \mathcal{L}[X]) + 2\mathcal{L}[X] \frac{\partial \log \left( h_{\text{max}} - \bar{h} \right)}{\partial \log \Omega_p}, \quad (2.4.28a)$$

where $\mathcal{L}(x) = \frac{\partial \log \text{erf}(x)}{\partial \log x}$, and $X = \left[ \sqrt{\frac{\beta}{2\Omega^*}} (h_{\text{max}} - \bar{h}) \right]$.

The sign of $\partial \log \bar{T}/\partial \log \Omega_p$ therefore depends on the detailed behaviour of $A(J)$ in Eq. (2.4.24), but will certainly be negative when $\mathcal{L}(X) \rightarrow 1$, i.e. when Eq. (2.4.25) fixes $\bar{h}$ sufficiently close to $h_{\text{max}}$ in units of $\sqrt{\beta_{\text{crit}}/\bar{T}^*}$. For a set of initial conditions $(n(J), H, E)$ yielding such a sub-critical temperature, Eq. (2.4.26a) implies that $\bar{h}$ should, on the average, be larger than $\langle h \rangle$, inducing a bar which rotates faster than the mean angular frequency. This result corresponds formally to the intuitive fact that a skewed distribution function biased towards the escape angular momentum has a maximum higher than its mean. As this distribution becomes narrower, the mean and the maximum converge. In the zero temperature limit, Eq. (2.4.26a) suggests indeed that $\Omega_p$ converges towards this mean rotation rate like $1/\beta$.

Equations (2.4.26), (2.4.27) and (2.4.28) are compatible with a scenario in which a galaxy forms with a mean energy, a mean angular momentum and a circulation distribution such that the corresponding temperature is below the critical temperature for bi-symmetric instability and such that Eqs. (2.4.28) yield a negative specific heat for some temperature range. The collective alignment of inner Lindblad orbits would then act as an energy sink; these orbits would exchange energy and momentum via other resonances, but without inducing any collective collapse. Weakly trapped orbits could then be thought of as a heat bath coupled to a bar displaying negative specific heat. If this specific heat were larger in modulus to that of the bath, the equilibrium would be secularly unstable. In effect, the heat flow would drive energy away from the bar, yielding a stronger, colder bar which would rotate at slower and slower angular speed until it reaches temperatures for which $\mathcal{L}[X] \rightarrow 0$. The bar may then stop displaying negative specific heat – leading the system, bar plus galaxy, towards a secular equilibrium, or the system may continue to cool down.

---

15 The low temperature value might not be reached, as $X$ scales like $\beta$, therefore $\mathcal{L}(X) \rightarrow 0$ when $\beta \rightarrow \infty$. In this régime, as mentioned earlier, the adiabatic distortion of the orbit dominates the variation of its moment of inertia, fixing the sign of $\partial \log \bar{T}/\partial \log \Omega_p$. 
Orbital instabilities and bar formation

Chapter 2

Figure 2.4.4: behaviour of $n(J) = \int F_Q(h, J) \, dh$ for different values of the Toomre $Q$ number in the tepid Isochrone disk described by Eqs. (2.4.6).

2.4.5 Illustration: the Isochrone disk

The above analysis is now applied to the Isochrone disk. An analytic expression for the energy as a function of actions is available for this disk (in units of $GM = b = 1$):

$$\varepsilon(h, J) = -\frac{1}{2} \left( J + \sqrt{h^2/4 + 1} \right)^{-2}.$$  \hspace{1cm} (2.4.29)

The angular frequency, $\Omega_\ell = \partial\varepsilon/\partial h$, and the inverse moment of inertia, $\alpha = \partial^2\varepsilon/\partial h^2$, follow readily from Eq. (2.4.29) and are illustrated on Fig. 2.4.5. The collective instability described in this chapter requires that $\alpha$ be positive, which in turn implies that only these orbits should be counted in Eqs. (2.4.8), yielding

$$\langle e \rangle = \frac{1}{N} \int_0^{J_{\max}} \int_0^{2J} \varepsilon(h, J) F_0(h, J) \, dh \, dJ, \quad \langle h \rangle = \frac{1}{N} \int_0^{J_{\max}} \int_0^{2J} F_0(h, J) \, h \, dh \, dJ,$$  \hspace{1cm} (2.4.30)

where $N = \int_0^{J_{\max}} n(J) \, dJ$, and $J_{\max} \approx 0.75$. Distribution functions for the Isochrone are derived in the next chapter. Here, for simplicity, tepid disks are constructed by the ansatz

$$F_Q(h, J_R) = \frac{\Sigma \sigma_R^2}{\kappa} \exp \left( -\frac{\kappa J_R}{\sigma_R^2} \right) \left( \frac{\partial R}{\partial h} \right), \quad \text{where} \quad \sigma_R = \frac{\pi Q \Sigma}{\kappa}. \hspace{1cm} (2.4.31)$$

The function $\sigma_R$ measures the radial velocity dispersion corresponding to a disk with Toomre number $Q$. The epicyclic frequency, $\kappa$, the surface density, $\Sigma$, and $\sigma_R$ are all functions of
Figure 2.4.5: the isocontours of $\Omega_\ell$ and $\alpha$ in $(h, J)$ space. The section of the plot below the line $J = 1/2h$ is unphysical. Therefore for each value of $J$, $h$ may vary between 0 and $h_{\text{max}} = 2J$. From right to left, the vertical contours correspond to $\alpha = 0, 0.01...0.25$, while the horizontal contours correspond from bottom to top to $\Omega_\ell = 0.11, 0.15,...,0.01$. Note the $\alpha = 0 = 1/I^* \rho$ contour, normal to the maxima of $\Omega_\ell$, which corresponds to orbits which are indifferent to torques. In this analysis, only orbits on the left of that contour have positive $\alpha$ and contribute to the collective instability. This yields a maximum value for $J \approx 0.75$.

The distribution of the number of orbits with adiabatic circulation $J$ given by Eq. (2.4.32) is illustrated on Fig. 2.4.4.

In the spirit of section 2.3, consider here an “idealised” Isochrone galaxy, for which a small subset of resonant orbits is responsible for the onset of the instability, in effect describing the distribution $n(J)$ shown on Fig. 2.4.4 as a delta function centred on $\mathcal{J} \propto Q$. Let Eq. (2.4.30) fix $\langle e \rangle$ and $\langle h \rangle$ for the corresponding $Q$ number. For simplicity, assume also that the adiabatic distortion of the orbits can be neglected. Therefore $A(h, \mathcal{J}) = \text{Const.} \equiv \bar{A}$. Condition Eq. (2.4.17) then reads again

$$K = \frac{I_1(\beta \bar{A} K)}{I_0(\beta \bar{A} K)} \bar{A} N,$$

and Eq. (2.4.21) gives the critical temperature, $\bar{A}^2 N/2$, below which a bi-symmetrical pattern is spontaneously generated within the disk, which corresponds essentially to a Jeans-like criterion.
Orbital instabilities and bar formation

Chapter 2

Figure 2.4.6: pattern speed, $\Omega_p$, versus momentum, $\langle h \rangle$, for different temperatures measured by $(\beta A^2 N)^{-1}$ and $\bar{J}$ in the tepid Isochrone disk. This figure illustrates the “temperature” behaviour of $(\beta A^2 N)^{-1}$. The different bundles, from top to bottom, correspond to $J = 0, 0.2, 0.4$. The lines in each bundle correspond to $5, 7, \ldots$, for $\beta A^2 N$ moving up the page.

Note that $K/\bar{A}$, which represents the number density of orbits trapped in the bar is bounded by $N$, the total number of orbits, as shown on Fig. 2.4.1. The pattern speed of the bar, $\Omega_p(\beta)$ corresponds to the solution of the system

$$\bar{h} = \langle h \rangle - \frac{1}{2\beta \Omega_p} \left( \frac{\partial \log \bar{I}}{\partial \log \Omega_p} \right) \quad \text{and} \quad \Omega_p = \Omega_p(\bar{h}) , \quad (2.4.34)$$

while the temperature, $1/\beta$, follows from $\Omega_p(\beta)$, Eq. (2.4.33) and Eq. (2.4.26b) which now reads

$$\langle e \rangle = \bar{e} + \frac{1}{2\beta} \left( 1 - \frac{\partial \log \bar{I}}{\partial \log \beta} \right) - \frac{K^2}{2N} . \quad (2.4.35)$$

The effective moment of inertia obeys

$$\frac{\partial \log \bar{I}}{\partial \log \Omega_p} = \Omega_p \frac{\partial \alpha^{-1}}{\partial h} (1 - \mathcal{L}[X]) - 2 \frac{\Omega_p \alpha^{-1}}{h_{\text{max}} - h} \mathcal{L}[X] , \quad (2.4.36)$$

where $\mathcal{L}[X]$ is given by Eq. (2.4.28b). For realistic galaxies, $\alpha \equiv \partial^2 \bar{e}/\partial h^2$, is a decreasing function of $h$. Its isocontours are illustrated on Fig. 2.4.5 for the Isochrone disk.

In practise, a given model is best implemented by choosing $\Omega_p$ and $\beta > \beta_{\text{crit}}$. Equation (2.4.33) yields the amplitude of the bar, while $\bar{h}$ is given by $\bar{h} = \Omega_\ell^{-1}(\Omega_p)$. The mean momentum and energy follow in turn from Eqs. (2.4.34) and (2.4.35) via Eq. (2.4.36). Fixing $\langle h \rangle$ yields $\Omega_p = \Omega_p(\beta)$ and therefore $\langle e \rangle$ as a function of $\beta$; its derivative corresponds to $C_H$. The above prescription is implemented on Fig. 2.4.7 for two disks with different global temperature $\bar{J} \propto Q$. Collett and Lynden-Bell (1989) [18] have shown that, on these diagrams, the bar will evolve along the lines obeying $\delta \langle e \rangle = \Omega_p \delta \langle h \rangle$. 

Figure 2.4.7: isocontours of $\langle h \rangle$ and $\langle e \rangle$ in temperature (horizontal axis) and pattern speed (vertical axis) space for a cold (right panel), $\bar{J} = 0.15$, and a warm (left panel), $\bar{J} = 0.5$, Isochrone disk. On the right panel from bottom to top, the horizontal contours correspond to $\langle h \rangle = 0.1, 0.2, ..., 2$, and from left to right, the vertical contours correspond to $\langle e \rangle = -0.825, -0.8125, ..., -0.50$. Similarly, on the left panel, from top to bottom, horizontal contours correspond to $\langle h \rangle = 0.5, 1, ..., 10$, while from left to right the vertical contours correspond to $\langle e \rangle = -0.65, -0.60, ... 0$. The intersection of the corresponding contours yields the temperature and the pattern speed of the instability. If this temperature is below the critical temperature, Eq. (2.4.33) yields the amplitude of that perturbation.

In the low temperature limit, $\partial \log \bar{I} / \partial \log \beta = -\partial \log \bar{I} / \partial \log \Omega_p$, and Eq. (2.4.27) gives for the specific heat

$$C_H = \frac{1}{2} \left( 1 + \frac{\partial \log \bar{I}}{\partial \log \Omega_p} \right) - \frac{1}{2} \left( \frac{\partial^2 \log \bar{I}}{\partial \log \beta \partial \log \Omega_p} \right) - \beta^2 \frac{\partial \bar{e}}{\partial \beta} + \beta^2 \frac{K \partial K}{N \partial \beta}, \quad (2.4.37a)$$

$$= \frac{1}{2} + \beta^2 \frac{K \partial K}{N \partial \beta}. \quad (2.4.37b)$$

When $\beta \to \infty$, the orbits start behaving similarly to rigid rods, the moment of inertia converges towards $\alpha^{-1}$, and $C_H$ is positive as illustrated on Fig. 2.4.2. Recall that for simplicity, the self-consistent distortion of orbits has not here been fully taken into account while taking $\bar{A}$ as a constant. A more detailed analysis accounting for both the dispersion in $n(J)$, the intermediate temperature regime, and the variations of $A(J)$ is required in order to be conclusive about the fate of real galaxies.

2.4.6 Discussion

The equations (2.4.17) and (2.4.33) may be understood both as a bar formation and also bar dissolution criteria. Indeed, numerical evidence suggests that the process of galaxy formation occurs in a highly non-symmetrical environment which is likely to trigger transient non-axisymmetric features within the forming disks. The viability of these grand design structures depends on the intrinsic properties of the disk such as its kinetic temperature. Criterion
Eq. (2.4.17) provides an estimate of the temperature above which a bar is not viable and will disappear adiabatically.

The formation of initially fast bars that slow down as they capture adiabatically more orbits corresponds clearly to the results of numerical simulations, and is compatible with observations. The mechanism presented in this section should therefore be implemented in details on realistic models such as those constructed in the next chapter to account for other observable trends such as aspect ratio evolution, etc.

2.5 Global azimuthal instabilities

The previous models were introduced with simplifications to present some physical insight in the phenomenon of bar formation. The instability is now investigated within the framework of the Boltzmann and the Poisson equations.

2.5.1 Boltzmann’s Equation: the dynamics

Boltzmann’s equation reads

$$\frac{\partial F}{\partial t} + \{H, F\} = 0,$$

where $H$ is the Hamiltonian for the motion of one star, $F$ is the mass-weighted distribution function in phase space, and the square bracket denotes the Poisson bracket

$$\{H, F\} = \frac{\partial H}{\partial p} \cdot \frac{\partial F}{\partial q} - \frac{\partial H}{\partial q} \cdot \frac{\partial F}{\partial p}. \quad (2.5.2)$$

Writing $F = F_0 + f$, and $\Psi = \psi_0 + \psi$ and linearising Eq. (2.5.1) in $f$ and $\psi$ yields

$$\frac{\partial f}{\partial t} + \{H_0, f\} - \{\psi, F_0\} = 0. \quad (2.5.3)$$

According to Jeans theorem, the unperturbed equation, $\{H_0, F_0\} = 0$, is solved if $F_0$ is any function of the energy, $E$, and the angular momentum, $h \ (= R^2 v_0)$.

Following Lynden-Bell and Kalnajs (1972) [76], angle and action variables of the unperturbed Hamiltonian $H_0$ are chosen here as canonical coordinates in phase-space. Indeed, the unperturbed Hamiltonian equations are quite trivial in these variables, which makes them suitable for perturbation theory in order to study quasi-resonant orbits\(^{16}\). In polar coordinates $(R, \phi)$, the unperturbed Hamiltonian reads

$$H_0 = \frac{1}{2} \left( p_R^2 + p_\phi^2 \right) - \psi_0(R), \quad \text{where} \quad p_R = v_R \quad \text{and} \quad p_\phi = h/R. \quad (2.5.4)$$

The actions are defined in terms of these variables by $J = (J_R, J_\phi) = (J_R, h)$, where

$$J_R = \frac{1}{2\pi} \oint [\hat{R}] \, dR, \quad J_\phi = h = R^2 \dot{\phi}. \quad (2.5.5)$$

Here $[\hat{R}]$ is a function of the radius $R$, the specific energy $E$, and the specific angular momentum, $h$, given by

$$[\hat{R}] = \sqrt{2E + 2\psi_0(R) - h^2/R^2}. \quad (2.5.6)$$

\(^{16}\) See Appendix A for more details on angle and action variables
The phase-angles conjugate to $J_R$ and $h$ are $\varphi = (\varphi_R, \varphi_\phi)$, where
\begin{align}
\varphi_R &= \kappa \int_R^R \left[ \ddot{R} \right]^{-1} dR, \\
\varphi_\phi &= \phi + \int_R^R (h/R^2 - \Omega) \left[ \ddot{R} \right]^{-1} dR. \tag{2.5.7a}
\end{align}

The generalised epicyclic frequency, $\kappa$, obeys
\begin{equation}
2\pi/\kappa = \oint \left[ \ddot{R} \right]^{-1} dR = P_R, \tag{2.5.8}
\end{equation}
where $P_R$ is the (radial) period between apocentres. The angle between apocentres is
\begin{equation}
\Phi = \oint h/R^2 \left[ \ddot{R} \right]^{-1} dR \quad \text{and} \quad \Omega = \Phi/P_R. \tag{2.5.9}
\end{equation}

Hamilton’s equation for the unperturbed orbits reads
\begin{equation}
\dot{\varphi} = \left( \partial H_0 / \partial \mathbf{J} \right) = \Omega (\mathbf{J}), \tag{2.5.10}
\end{equation}
where $\Omega$ stands for $(\kappa, \Omega)$, while the $\mathbf{J}$ are kept constant as the actions were designed so that $H_0$ is independent of the phases $\varphi$. \footnote{The actions are a combination of integrals of the unperturbed motion by construction. See the appendix for details.}

The stationary unperturbed Boltzmann equation Eq. (2.5.1) is solved by $F = F_0(\mathbf{J})$, since $[H_0, \mathbf{J}] = 0$. When expanded in Fourier series with respect to the angles $\varphi$, Eq. (2.5.2) becomes
\begin{equation}
\partial f_\mathbf{m} / \partial t + i \mathbf{m} \cdot \Omega f_\mathbf{m} + i \mathbf{m} \cdot \partial F_0 / \partial \mathbf{J} \psi_\mathbf{m} = 0, \tag{2.5.11}
\end{equation}
where $\mathbf{m}$ is an integer vector with components $(\ell, m)$, and $\psi_\mathbf{m}$ and $f_\mathbf{m}$ are the Fourier transforms of $\psi$ and $f$ with respect to $(\varphi_R, \varphi_\phi)$.

For growing instabilities, $f_\mathbf{m}$ and $\psi_\mathbf{m}$ are taken to both be proportional to $e^{i\omega t}$, $\omega$ having a negative imaginary part. Eq. (2.5.11) then becomes
\begin{equation}
f_\mathbf{m} = \frac{-m^{-1} \mathbf{m} \cdot \partial F_0 / \partial \mathbf{J}}{\Omega_\ell - \Omega_p} \psi_\mathbf{m}, \tag{2.5.12}
\end{equation}
where $\Omega_\ell = m^{-1} \mathbf{m} \cdot \Omega = \Omega + \ell \kappa/m$ and $\Omega_p = -\omega/m$. When $\ell = -1$ this coincides with the lobe tumbling rate of the orbits discussed in section 2.3. When $\ell = 1$ it gives the lobe tumbling rate of the type of orbits encountered at outer-Lindblad-resonance while $\ell = 0$ gives $\Omega_\ell = \Omega$, the tumbling rate of co-rotating epicycles.

There is evidently a resonant response from all those parts of phase-space for which $\Omega_\ell (\mathbf{J})$ is close to $\Re \{ \Omega_p \}$. For different $\ell$ values these are distinct parts of phase-space because $\Omega_\ell - \Omega_{\ell'} = (\ell' - \ell) \kappa/m \neq 0$, for $\ell \neq \ell'$. In dealing with each resonant term, new angle and action variables are introduced which are more appropriate and make the particular resonance labelled by $(\ell, m)$ especially simple. These “fast” and “slow” coordinates are obtained from the canonical transformation generated by the generating function
\begin{equation}
S(J_s, J_f, \varphi_R, \varphi_\phi) = J_s (\varphi_\phi - \Omega_p t - \ell \varphi_R/m) + J_f \varphi_R. \tag{2.5.13}
\end{equation}
The transformation to the new variables is given by
\begin{align}
\partial S / \partial \varphi_\phi &= h \quad \text{gives} \quad J_s = h, \tag{2.5.14a} \\
\partial S / \partial \varphi_R &= J_R \quad \text{gives} \quad J_f = J_R - \ell h/m \equiv J_t, \tag{2.5.14b} \\
\partial S / \partial J_s &= \varphi_s \quad \text{gives} \quad \varphi_s = \varphi_\phi + \ell/m \varphi_R - \Omega_p t, \tag{2.5.14c} \\
\partial S / \partial J_f &= \varphi_f \quad \text{gives} \quad \varphi_f = \varphi_R. \tag{2.5.14d}
\end{align}
Note that
\[ \Omega_s = \dot{\varphi}_s = \Omega - \ell \kappa / m - \Omega_p = \Omega - \Omega_p. \]  
(2.5.15)
In axes that rotate at \( \Omega + \ell \kappa / m \) the orbit closes with \( m \) lobes after \( |\ell| \) turns about the centre. From Eq. (2.5.15), \( \Omega_s \) is zero on resonance, and for all near-resonant orbits \( \varphi_s \) varies slowly. In fact \( \varphi_s \) is the azimuth of the lobe of an orbit in the frame that rotates with angular velocity \( \Omega_p \). It is convenient to define
\[ \varphi_\ell \equiv \varphi_\phi + \ell \varphi_R / m, \]  
(2.5.16)
which is the lobe’s direction in absolute space. Then \( \varphi_s = \varphi_\ell - \Omega_p t \). It is interesting to see that \((\partial H_0 / \partial h)_J = \Omega_\ell = \dot{\varphi}_\ell \) for every \( \ell \).

Returning to Eq. (2.5.12) and evaluating it using Eq. (2.5.14)
\[ m^{-1} \mathbf{m} \cdot \partial F_0 / \partial \mathbf{J} = \left( \frac{\partial F_0}{\partial h} \right)_{J_R} + \frac{\ell}{m} \frac{\partial F_0}{\partial J_R}, \]
\[ = \left( \frac{\partial F_0}{\partial h} \right)_{J_R} + \left( \frac{\partial F_0}{\partial J_R} \right) h \left( \frac{\partial J_R}{\partial h} \right)_{J_f}, \]
\[ = \left( \frac{\partial F_0}{\partial h} \right)_{J_f}. \]  
(2.5.17)
Thus Eq. (2.5.12) reads
\[ f_\ell = \frac{[-\partial F_0 / \partial h]_{J_f}}{\Omega_\ell - \Omega_p} \psi_\ell. \]  
(2.5.18)
The result (2.5.18) shows how strikingly simple the dynamics is in angle action variables.

### 2.5.2 Poisson’s equation

The Poisson equation relates the potential, \( \psi \), to the density perturbation:
\[ \psi (R', \phi') = \int \int \frac{f(R, \mathbf{v})}{|R - R'|} dR d\phi d\mathbf{v}_R d\mathbf{v}_\phi. \]  
(2.5.19)
This equation may be rewritten in order to make explicit the contribution from the interaction of orbits. Here again angle-action variables are useful, as a given unperturbed orbit is entirely specified by its actions. It is straightforward to identify in Poisson’s equation the contribution corresponding to the interaction of given orbits. Re-expressing this equation in terms of angle and actions \((\varphi, \mathbf{J})\) and taking its Fourier transform with respect to \( \varphi \) implies:
\[ \psi_m (\mathbf{J}) = \pi G \sum m' \int f_{m'} (\mathbf{J'}) A_{mm'} (\mathbf{J}, \mathbf{J'}) d^2 J', \]  
(2.5.20)
where
\[ A_{mm'} = \frac{1}{4\pi^3} \int \int \frac{\exp i (m' \cdot \varphi' - m \cdot \varphi)}{|R - R'|} d^2 \varphi' d^2 \varphi. \]  
(2.5.21)
The (double) sum in Eq. (2.5.20) extends for both \( \ell \) and \( m \) going from minus infinity to infinity. \( R = R(\varphi, \mathbf{J}) \) and \( R' = R(\varphi', J') \) are radii re-expressed as functions of these variables. Note that a \( m \)-fold symmetry family of orbits may contribute to the \( m' \) the component of the potential according to Eq. (2.5.20). Now \( |R - R'| \) depends on \( \varphi_\phi \) and \( \varphi'_\phi \) in the combination \( \varphi'_\phi - \varphi_\phi = \Delta \varphi \) only. Thus
\[ (m' \cdot \varphi' - m \cdot \varphi) = m' \Delta \varphi - (m - m') \varphi_\phi - \ell \varphi_R + \ell' \varphi_R. \]  
(2.5.22)
As \( |\partial \varphi_\phi \partial \varphi_\phi / \partial \Delta \varphi \partial \varphi_\phi| = 1 \), the order of integration may then be reversed doing the \( \varphi_\phi \) integration with \( \Delta \varphi \) fixed. This yields \( 2\pi \delta_{m,m'}/2 \), so \( m' \) becomes \( (l', m) \) in the surviving terms. This gives for Eq. (2.5.22)

\[
A_{mm'} = \frac{\delta_{m,m'}}{2\pi^2} \int \int \int \exp i(m\Delta \varphi - \ell' \varphi_R + \ell \varphi_R) d\Delta \varphi d\varphi_R d\varphi.
\] (2.5.23)

Now \( \varphi_\ell = \varphi_\phi + \ell \varphi_R/m \) and \( \varphi_f = \varphi_R \), so Eq. (2.5.23) becomes

\[
A_{mm'} = \frac{\delta_{m,m'}}{2\pi^2} \int \int \int \exp im(\varphi' - \varphi_\ell) d(\varphi' - \varphi_\ell) d\varphi' d\varphi
\]  

\[
= \frac{\delta_{m,m'}}{2\pi^2} \int \int d\varphi d\varphi' \exp im(\Delta \varphi_\ell) d(\Delta \varphi_\ell),
\] (2.5.24)

where \( |\mathbf{R} - \mathbf{R}'| \) is thought of as a function of \( \mathbf{J}, \mathbf{J}' \), the angle between the lobes of the two orbits, \( \Delta \varphi_\ell \), and the two phases around those orbits \( \varphi_f \) and \( \varphi_f' \). Equation (2.5.24) states an interesting result; it implies that two orbits with different \( m \)-fold azimuthal symmetry do not interact. Eq. (2.5.24) has a simple physical interpretation in terms of potential energy. Consider the energy of gravitational interaction, \( V \), given by

\[
V = -\frac{1}{2} \int \int (\Delta \varphi, \varphi) \psi(R, \varphi) dR d\varphi,
\] (2.5.25)

where the surface density reads in terms of the distribution function

\[
\Sigma(R, \phi) = \int f(R, \phi, v_R, v_\phi) dR d\phi d\varphi.
\] (2.5.26)

Therefore

\[
V = -\frac{1}{2} \int \int f(R, \phi, v_R, v_\phi) \psi(R, \phi) dR d\phi dR d\phi,
\] (2.5.27)

Angle and action variables of the axisymmetric Hamiltonian are introduced again to enlight the contribution of each orbit to the overall potential energy. The potential \( \psi \) expressed in terms of \( (\varphi, \mathbf{J}) \) is given by Eq. (2.5.20)

\[
\psi(\mathbf{J}, \varphi) = \sum_{m_m} \int A_{mm'}(\mathbf{J}, \mathbf{J}') e^{i(m \varphi - m' \varphi')} f(\mathbf{J}', \varphi') d\chi',
\] (2.5.28)

with \( d\chi = d^2 \mathbf{J}' d^2 \varphi' \), the elementary phase space volume. Inserting Eq. (2.5.28) into Eq. (2.5.27) leads to the formal expression for \( V \)

\[
V = \sum_{m, l, l'} V_{m, l, l'} = \frac{1}{2} \int \int f(\mathbf{J}, \varphi) A_{m, l, l'}(\mathbf{J}, \mathbf{J}') e^{i(m \varphi - m' \varphi')} f(\mathbf{J}', \varphi') d\chi d\chi'.
\] (2.5.29)

Switching the fast and slow angles, Eq. (2.5.29) becomes

\[
V_{m, l, l'} = 2\pi^2 \int \int f_\ell(\mathbf{J}, \varphi_\ell) A_{m, l, l'}(\mathbf{J}, \mathbf{J}') e^{i m \Delta \varphi_\ell} f_{l'}(\mathbf{J}', \varphi_{l'}) d^2 \mathbf{J} d^2 \mathbf{J}' d\Delta \varphi_\ell,
\] (2.5.30)

where

\[
f_\ell(\mathbf{J}, \varphi_\ell) = \frac{1}{2\pi} \int f(\mathbf{J}, \varphi_\ell, \varphi_R) e^{i \varphi_R} d\varphi_R.
\] (2.5.31)

Imagine a unit mass orbital stream on the orbit specified by \( m, \ell', \mathbf{J}', \varphi_{l'} \). In the gravitational potential of that stream, place another stream specified by \( m, \ell, \mathbf{J}, \varphi_\ell \); then their mutual gravitational potential energy is some function

\[
V_\Delta(m, \ell, \ell', \mathbf{J}, \varphi_\ell).
\]
If $V_{\Delta}$ is expanded as a Fourier series in $\Delta \varphi_{\ell}$, the angle between the lobes of the two orbits, then the first non-zero angular-dependent term is

$$A_{m,\ell,\ell'} \exp(m \Delta \varphi_{\ell}).$$

Formally this follows from Eq. (2.5.30) where delta functions are substituted for $f_{\ell}$ and $f_{\ell'}$. $V_{\Delta}$ is therefore the exact analogue of the orientation potential $\Psi$ introduced in section 2.3.

### 2.5.3 Instability criterion

Putting Eq. (2.5.18) into Eq. (2.5.20) leads to the integral equation

$$\psi_{\ell_1} \left( J_1 \right) = 2\pi^2 G \sum_{\ell_2} \int \int A_{m,m'} \left( J_1, J_2 \right) \frac{\left[ -\partial F_0 / \partial h \right]_{\ell_2}}{\Omega_{\ell_2} - \Omega_p} \psi_{\ell_2} \left( J_2 \right) d^2 J_2.$$  \hspace{1cm} (2.5.32)

Equation (2.5.32) is formally identical to the integral equation found in section 2.3 except for the summation over all $\ell$ resonances.

- The integral equation Eq. (2.5.32) could be tackled analytically under the assumption that there exists a way to split $A_{m,m'} \left( J_1, J_2 \right) = A_{\ell_1} \left( J_1 \right) A_{\ell_2} \left( J_2 \right)$. Eq. (2.5.32) then becomes

$$1 = 2\pi^2 G \sum_{\ell_2} \int \int A_{\ell_2}^2 \left( J_2 \right) \frac{\left[ -\partial F_0 / \partial h \right]_{\ell_2}}{\Omega_{\ell_2} - \Omega_p} d^2 J_2,$$  \hspace{1cm} (2.5.33)

which is similar to equation Eq. (2.3.8) apart for the summation over all orbits. Identifying both real and imaginary part of Eq. (2.5.33) yields

$$2\pi^2 G \sum_{\ell_2} \int \int A_{\ell_2}^2 \left( J_2 \right) \frac{\left[ -\partial F_0 / \partial h \right]_{\ell_2}}{\left( \Omega_{\ell_2} - \Omega_R \right)^2 + \eta^2} d^2 J_2 = 0,$$  \hspace{1cm} (2.5.34a)

$$2\pi^2 G \sum_{\ell_2} \int \int \left| \Omega_{\ell_2} - \Omega_R \right| A_{\ell_2}^2 \left( J_2 \right) \frac{\left[ -\partial F_0 / \partial h \right]_{\ell_2}}{\left( \Omega_{\ell_2} - \Omega_R \right)^2 + \eta^2} d^2 J_2 = 1,$$  \hspace{1cm} (2.5.34b)

where $\Omega_p = \Omega_R + i\eta$. When the main resonant contribution for bars is assumed to come from the inner Lindblad and co-rotation, $\left[ -\partial F_0 / \partial h \right]_{J_f}$ is generally positive for $J_f = J_R$, $\ell_2 = 0$ at co-rotation, and negative at the inner Lindblad resonance ($J_f = J_R + h/2$, $\ell_2 = -1$). As a consequence, Eq. (2.5.34a) may be satisfied, while both resonance contribute to Eq. (2.5.34b) positively, when $\Omega_0 > \Omega_p$ while $\Omega_{-1} < \Omega_p$.

- More accurately, solutions to a given integral equation may be found either by expanding the unknowns over a set of orthonormal functions that is likely to describe the type of perturbation correctly, or by searching for the eigen modes of the Kernel. An alternative prescription would, for instance, involve constructing a complete set of orthonormal distribution functions on which the perturbation is projected

$$f_{\ell} \left( J \right) = \sum_n \chi_{n,\ell} f_n \left( J \right).$$  \hspace{1cm} (2.5.35)

The system of equations which follows from inserting Eq. (2.5.35) into Eq. (2.5.32) has a non-trivial solution for the $\chi_{n,\ell}$’s only if

$$D \left( \Omega_p \right) = \det \left| 1 + M \left( \Omega_p \right) \right| = 0,$$  \hspace{1cm} (2.5.36)
where \( \mathbf{I} \) is the identity matrix, and \( \mathbf{M} \) is a matrix given by

\[
(M^{(n,n')})_{(\ell,\ell')} = \pi G \int \int f_n (\mathbf{J}) \left[ \frac{-\partial F_0/\partial h}{\Omega_\ell - \Omega_p} - \frac{\partial F_0/\partial h}{\Omega_{\ell'} - \Omega_p} \right] A_{\ell,\ell'} (\mathbf{J}, \mathbf{J}') \frac{-\partial F_0/\partial h}{\Omega_\ell - \Omega_p} f_{n'} (\mathbf{J}') \, d^2 \mathbf{J} \, d^2 \mathbf{J}'.
\]

Equation (2.5.36) gives the criterion for the existence of exponentially growing unstable modes. These are combinations of the eigenvectors of the null space of \( \mathbf{I} + \mathbf{M}(\Omega_p) \). The corresponding perturbed distribution function characterises the orbits which are first caught by the instability. If this distribution function is peaked in the neighbourhood of a given orbit caught in a given resonance, the instability is orbital in nature. If not, the method of potential density pairs developed by Kalnajš (1971) [54] gives a more direct route to the standing wave generated by the instability. Note that this wave is available from Eq. (2.5.35) by simple integration over velocity space. For the purposes of this method, a numerical evaluation of the coupling coefficients \( A_{\ell,\ell'}^m \) is given in the appendix. Distributions functions parameterised by their temperature are also constructed in the next chapter. In practice, the choice of an appropriate basis function, Eq. (2.5.35), is critical to ensure a fast convergence of the determinant in Eq. (2.5.36). Indeed, a good description of the natural mode of the system must be chosen differently for each galaxy.

- An alternative systematic inversion method for Eq. (2.5.32) is presented in the appendix. The eigen-modes \( (\lambda_n, u_n) \) of the Kernel \( A_{\ell,\ell'} (\mathbf{J}, \mathbf{J}) \) are constructed in the epicyclic approximation with softened gravity. Within the framework of that approximation, the integration over eccentricity and over the angles may be performed in Eq. (2.5.24), yielding a one dimensional integral equation for which simple analytic modes are derived. The projection of Eq. (2.5.32) onto these modes leads to the dispersion relation

\[
D(\Omega_p) = \det |\Lambda + \mathbf{H}(\Omega_p)| = 0, \tag{2.5.38}
\]

where

\[
H_{(\ell,\ell')}^{(n,n')} = \gamma_{(\ell,\ell')} \pi G \int \int A_n^{\ell} (\mathbf{J}) A_{n'}^{\ell'} (\mathbf{J}) \frac{-\partial F_0/\partial h}{\Omega_\ell - \Omega_p} d^2 \mathbf{J}, \tag{2.5.39}
\]

and the effective \( A_n^\ell \) are defined by:

\[
A_n^\ell (\mathbf{J}) = a_\ell [e (\mathbf{J})] u_n [\hat{R}(\mathbf{J})]. \tag{2.5.40}
\]

The eccentricity dependence, \( a_\ell(e) \), and the eigen-functions, \( u_n \), are given in the appendix. These functions are characteristic of the law of gravitation, and as such need not be tailored to each specific underlying galaxy.

The dispersion relations Eqs. (2.5.36) and (2.5.38) give the criterion for the existence of exponentially growing unstable modes. They are functions of a free parameter \( \Omega_p \) corresponding to the pattern velocity of the growing bar. Their interpretation is as follows: consider the complex \( \Omega_p \) plane and a contour that traverses along the real axis and then closes around the circle at \( \infty \) with \( \Im(\Omega_p) \) positive. The determinant \( D \) is a continuous function of the complex variable \( \Omega_p \); so, as \( \Omega_p \) traces out that closed contour, \( D(\Omega_p) \) traces out a closed contour in the complex \( D \) plane. If the \( D(\Omega_p) \) contour encircles the origin, then the system is unstable. An intuitive picture of how the criterion operates on Eq. (2.5.38) is provided with the following thought experience. Imagine turning up the strength of \( G \) for the perturbation (i.e. turn on self-gravity, which is physically equivalent to allowing for the gravitational interaction to play its role). Starting with small \( G \), all the \( H_{m,n,m'} \) are small, so for all \( \Omega_p \)’s the determinant \( D(\Omega_p) \) remains on a small contour close to \( \prod_i \lambda_i \neq 0 \). As \( G \) is increased to its full value, either the \( D \) contour crosses the origin to give a marginal instability or it does not. If it does not, then by
continuous change of $G$ does not modify the stability, and therefore the self-gravitating system has the same stability as in the zero $G$ case, i.e. is stable. If, however, it crosses, and remains circling the origin then it has passed beyond the marginally stable case and is unstable. A similar argument holds for Eq. (2.5.36).

### 2.5.4 Overview

The simple picture invoking the alignment of inner Lindblad orbits described in section 2.3 as the triggering mechanism for the instability in stellar disk is somewhat endangered both by the results of numerical N-body simulations (see Sellwood 1993 [101] for a review) and more crucially by linear analysis of the natural density modes of the disk carried by Kalnajs (1976) [58] Sawamura (1988) [99] and Hunter (1992) [50]. Indeed the orbital analysis should predict the same pattern speeds and growth rates and reproduce, via integration over velocity space, all features of the natural density modes. But these analyses predict pattern speeds which are significantly (by a factor of two) above the maximum pattern speed reached by inner Lindblad orbits. In fact the fastest inner Lindblad orbits correspond to circular orbits. This factor of two discrepancy is therefore quite difficult to reconcile with a model in which the root of the instability corresponds to the alignment of such orbits via relative torques! In other words, the alignment of inner Lindblad orbits described in section 2.3 does not seem to trigger the instability with the fastest growth rate in the models studied by Kalnajs [54] and others. Merritt [86] Polyachenko [94] and Allen 1992 [2] have shown that a similar type of instability involving the alignment of radial orbits may occurs in the core of elliptical galaxies.

The adiabatic alignment of inner Lindblad orbits may nevertheless play some role later in the evolution of the structure of the bar beyond the linear regime as described in section 2.4. The corresponding results are in effect more likely to account for the observations than those of the linear theories described in sections 2.2, 2.3 and 2.5. Section 2.4 demonstrated that the alignment of resonant orbits could not be treated independently of their distortion, which in effect reconciled the mechanisms of eccentricity forcing [76] and alignment instability [82]. The formalism presented in section 2.5 is likely to re-unify all theories as it involves both the wave and the orbital description.

It may well turn out that the type of standing (fluid) waves described by Toomre [105] and others corresponds to the fastest linearly growing modes of stellar disks, and that this description is indeed still accurate for the evolved structures observed in real galaxies. Yet this picture is unsatisfactory as the swing amplification mechanism – a leading density wave which is gravitationally enhanced at co-rotation while becoming trailing – does not really make any prediction for the non-existence of bars in a given model. Moreover, this scenario doesn’t account for the nature of the flow in the bars found in numerical simulations, in which the bulk of the bar appears to be made of closed elongated orbits. There is a significant gap between the physical mechanism of wave amplification on the one hand, and the corresponding growth rates and pattern speed found by the matrix projection method of Kalnajs [58] on the other hand. This gap is difficult to fill while resorting to a partially numerical inversion method which requires different implementation for each galaxy.

A detailed implementation of the analysis of section 2.5 is required to address these drawbacks and make any final conclusion on the relevance of the orbital description for galactic bars.

In the final analysis, the problem of bar formation remains an open question!
Aspects of orbital instabilities

A.1 Numerical evaluation of the $A_{m'm}$

For numerical purposes, it is desirable to re-express the coupling factor $A_{m'm}$ introduced in section 5 in terms of integration over positions rather than angles. It may be re-arranged from Eq. (2.5.24) as

$$A_{m'm} = \frac{1}{4\pi^3} \int \int \int \frac{\exp(i\mu \Delta \phi)}{|R - R'|} \frac{\exp(i\Delta \{\ell \varphi_R\})}{|R - R'|} d\Delta \varphi d\varphi_R d\varphi_{R'},$$  \hspace{1cm} (A.1.1)

where the $\Delta$ symbol stands for the difference of its argument primed minus unprimed. The denominator in Eq. (A.1.1) may be expanded as

$$|R - R'| = \left( R^2 + R'^2 \right)^{1/2} \left[ 1 + x \cos(\Delta \phi) \right]^{1/2},$$  \hspace{1cm} (A.1.2)

where $x = 2RR'/(R^2 + R'^2)$. Given that

$$\varphi_\phi = \phi - \kappa^{-1} \int \frac{h}{R^2} d\varphi_R + \kappa^{-1} \frac{\varphi_R}{2\pi} \oint \frac{h}{R^2} d\varphi_R,$$  \hspace{1cm} (A.1.3)

$A_{m'm}$ may be re-written as

$$A_{m'm} = \frac{1}{4\pi^3} \int \left[ \exp \left( i \Delta \{\ell \varphi_R + m\varphi_\phi - m\phi\} \right) \right] \left[ \frac{\exp(i\Delta \phi)}{1 + x \cos(\Delta \phi)^{1/2}} \right] d\varphi_R d\varphi_{R'}.$$  \hspace{1cm} (A.1.4)

Calling

$$C_m (x) = \frac{1}{2\pi} \int \frac{\exp(i\Delta \phi)}{[1 + x \cos(\Delta \phi)]^{1/2}} d\Delta \phi,$$  \hspace{1cm} (A.1.5)

and

$$\varphi^\ell [R] = \ell \varphi_R + m\varphi_\phi - m\phi = \ell \varphi_R - m\kappa^{-1} \int \frac{h}{R^2} d\varphi_R + m\kappa^{-1} \frac{\varphi_R}{2\pi} \oint \frac{h}{R^2} d\varphi_R,$$  \hspace{1cm} (A.1.6)

yields
Figure A.1.1: The angular frequency, $\Omega_\ell$, and shape of a resonant orbit as a function of angular momentum, $h$, radial action, $J_R$, and resonance $m$ – Inner Lindblad on the left panel, outer Lindblad on the right panel. Note the departure from simple elongated ellipses for the eccentric orbits ($J_R$ large). Here the axis are re-scaled for each plot.

$$A_{m'm} = \frac{1}{2\pi^2} \int_{R_p}^{R_p'} \int_{R'_p}^{R'_p} \exp \left[ i \varphi'_R [R] - i \varphi'_R [R'] \right] C_m \left( \frac{2R'R'}{R^2 + R'^2} \right) \frac{dR}{R} \frac{dR'}{R'} \ . \tag{A.1.7}$$

Here $[R]$ is a function of $R$ and of the actions $(J, h)$ via

$$[R] = \sqrt{2} \left[ \varepsilon (J, h) + \psi (R) - \frac{h^2}{2R^2} \right]^{1/2} \ ; \tag{A.1.8}$$

the radial angle $\varphi_R$ is in turn a function of $R$ via the identity $d\varphi_R = dR/[R]$; consequently $\varphi_R'[R]$ is also a function of $R, J$ and $h$. The integral Eq. (A.1.7) may therefore be numerically evaluated over $R$ and $R'$ space. Note the the real part of $C_2$ reads

$$\Re \left[ C_2 (x) \right] = -\frac{8}{3 \pi x^2 \sqrt{1 + x}} \left[ (1 + x) \ E \left( \frac{2x}{1 + x} \right) - K \left( \frac{2x}{1 + x} \right) \right] - \frac{2}{3 \pi \sqrt{1 + x}} K \left( \frac{2x}{1 + x} \right) \ .$$

The Kernel, $G_m$, in Eq. (A.1.7) presents an essential singularity near the line $R = R'$. Physically, it corresponds to the contribution from tangential orbits for which the energy of interaction is infinite in the infinitely thin disk description. This artifact of this description is removed by softening gravity. In effect, softening restores the third dimension by putting all orbits in ‘different’ planes. Fig. A.1.2 illustrates this calculation of the interaction of inner Lindblad orbits.

Equation (A.1.7) requires the knowledge of the radial action $J_R$ and its derivatives. A general semi-algebraic algorithm was implemented to construct these functions for any given potential as shown on Fig. A.1.1 for the Kuzmin disk.
Figure A.1.2: The iso-contours of the coupling coefficient $A_{\ell,\ell'}$ between two orbits trapped in the inner Lindblad resonance of a Kuzmin disk as a function of angular momentum, $h$, and fraction of the circular energy, $f = \varepsilon / \varepsilon_h$ of the second orbit. The first orbit corresponds to $f = 0.55, h = 0.85$, and softening was taken to be equal to 0.1. Note the elongation along $h \propto f$, which accounts for the similar perigee of these orbits, when the overlapping of the two trajectories corresponds to the largest contribution in Eq. (A.1.7).

A.2 Epicyclic Criterion

In this section, (A.1.7) is expanded using epicycles: this allows one to carry the integration over geometry and eccentricity, reducing the integration over the full dimensional phase-space in Eq. (2.5.32) to a simple one dimensional integral equation. This integral equation is in turn inverted by a simple ansatz which effectively soften gravity for quasi-identical orbits. A stability criterion similar to that of section two is recovered, with a detailed prescription for the coupling factor $A^f(J)$.

A.2.1 Epicyclic expansion and phase space reduction

In the epicyclic approximation, the coordinates of a star $i$ describing its orbit read

\begin{align}
R_i &= \bar{R}_i + a_i \cos (\varphi_{R_i}) \,, \\
\phi_i &= \varphi_{\phi_i} - 2 \frac{a_i}{\bar{R}_i \kappa_i} \sin (\varphi_{R_i}) \,,
\end{align}

where the indices refer to orbit “one” and “two”. Here $a_i$ is the epicyclic radius, $\bar{R}_i$ the guiding centre radius, and $\phi_i$ the orientation of the guiding centre of star $i = 1, 2$ measured from the rotating frame $\Omega_{\nu_i}$. Recall that $\kappa_i$ and $\Omega_i$ are respectively the epicyclic frequency and the angular
frequency of the star \(i\). In the epicyclic approximation, Eq. (A.1.6) for \(\varphi^e_i\) becomes

\[
\varphi^e_i [\varphi_{R_i}] = \ell_i \left[ \varphi_{R_i} - 2\eta_i \frac{a_i}{R_i} \sin (\varphi_{R_i}) \right] \quad \text{where} \quad \eta_i = 1 + \frac{\Omega_{p_i} m}{\kappa_i \ell_i}.
\]

(A.2.2)

Here \(\Omega_{p_i}, \; i = 1, 2\), are intermediate variables obeying \(\Omega_i - \Omega_{p_i} = \ell_i \kappa_i / m\); their purpose is described below. Substituting Eq. (A.2.2) and Eq. (A.2.1a) into Eq. (A.1.7) allow \(A_{m_1,m_2}\) to be re-arranged as

\[
A_{m_1,m_2} = \frac{\delta_{m_2}^{m_1}}{2\pi^2} \int \int \left\{ G_m \left[ \tilde{R}_1 \left( 1 + \frac{a_1}{R_1} \cos (\varphi_{R_1}) \right), \tilde{R}_2 \left( 1 + \frac{a_2}{R_2} \cos (\varphi_{R_2}) \right) \right] \exp \left( i \ell_2 \left[ \varphi_{R_2} - 2\eta_2 \frac{a_2}{R_2} \sin (\varphi_{R_2}) \right] \right) - \right\} d\varphi_{R_1} d\varphi_{R_2},
\]

(A.2.3)

where the function \(G_m\) stands for

\[
G_m (R_1, R_2) = \frac{1}{(R_1^2 + R_2^2)^{1/2}} C_m \left[ \frac{2R_1 R_2}{R_1^2 + R_2^2} \right],
\]

(A.2.4)

and \(C_m\) is given by Eq. (A.1.5). Equation (A.2.3) is now Taylor-expanded in power of \(e_i \equiv a_i / \tilde{R}_i^1\). The zeroth order term of the expansion is

\[
A_{m_1,m_2}^{(0)} = \frac{\delta_{m_2}^{m_1}}{2\pi^2} \int \int \tilde{G}_m \exp [i \ell_2 \varphi_{R_2} - i \ell_1 \varphi_{R_1}] d\varphi_{R_2} d\varphi_{R_1} = 2 \delta_{m_1}^{m_2} \delta_{0}^{\ell_1} \delta_{0}^{\ell_2} \tilde{G}_m,
\]

(A.2.5)

and \(\tilde{G}_m\) stands for \(G_m (\tilde{R}_1, \tilde{R}_2)\). Equation (A.2.5) corresponds to the interaction of two orbits at co-rotation. It does not involve any eccentricity, as the orbits need not be eccentric to interact at co-rotation.

The first order term in \(e_i\) of the expansion reads

\[
A_{m_1,m_2}^{(1)} = \frac{\delta_{m_2}^{m_1}}{2\pi^2} \int \int \exp i (\ell_2 \varphi_{R_2} - \ell_1 \varphi_{R_1}) \left[ \tilde{G}_m (i \sin (\varphi_{R_1}) e_1 \eta_1 \ell_1 - i \sin (\varphi_{R_2}) e_2 \eta_2 \ell_2) + \right] \left( \cos (\varphi_{R_2}) \frac{\partial \tilde{G}_m}{\partial R_2} + \cos (\varphi_{R_1}) \tilde{R}_1 e_1 \frac{\partial \tilde{G}_m}{\partial R_1} \right) d\varphi_{R_1} d\varphi_{R_2}
\]

(A.2.6)

where \(\tilde{G}_m\) and its derivatives are to be evaluated at \(R_i = \tilde{R}_i, \; i = 1, 2\). Performing the double integration over the angles \(d\varphi_1 d\varphi_2\) leads to

\[
A_{m_1,m_2}^{(1)} = \frac{\delta_{m_2}^{m_1}}{2} \left\{ \delta_{\ell_1}^{\ell_1} \delta_0^{\ell_2} \left[ \tilde{G}_m \eta_1 \ell_1 + \tilde{R}_1 \frac{\partial \tilde{G}_m}{\partial R_1} \right] e_1 + \right\}
\]

\[
\left. \delta_{\ell_2}^{\ell_2} \delta_0^{\ell_1} \left[ \tilde{G}_m \eta_2 \ell_2 + \tilde{R}_2 \frac{\partial \tilde{G}_m}{\partial R_2} \right] e_2 + \tilde{G}_m \eta_1 \ell_1 - \tilde{G}_m \eta_2 \ell_2 \right\}
\]

(A.2.7)

These contributions correspond to the cross-interaction of orbits trapped in the different resonances \(\ell = -1, 1\) with orbits at co-rotation. It only requires the inner or the outer Lindblad orbit to be eccentric. The expression corresponding to the interaction between inner Lindblad

\(^1\) In the epicyclic approximation, the expansion is expected to converge fast only if \(a_i = o(\tilde{R}_2 - \tilde{R}_1)\) because \(G_m\) diverges when \(R_1 = R_2\).
orbits described in the previous sections comes into the expansion as a second order term in \( e \), which is derived similarly:

\[
A^{(2)}_{m, m_2} = \frac{\delta_{m_1} m_2}{4} e_1 e_2 \left\{ \delta_{\ell_1} \delta_{\ell_2} \left[ G_m \eta_1 \eta_2 \ell_1 \ell_2 + \eta_1 \ell_1 R_2 \frac{\partial G_m}{\partial R_2} + \eta_2 \ell_2 R_1 \frac{\partial G_m}{\partial R_1} + R_1 R_2 \frac{\partial^2 G_m}{\partial R_1 \partial R_2} \right] + \right\} ,
\]

where all corrections to the lower order terms have been dropped. The first group in Eq. (A.2.8) corresponds to the interaction of inner Lindblad orbits and the second to outer Lindblad orbits. It requires both orbits to be eccentric.

Recognising in the integrand of Eq. (A.2.4) the generator for Legendre polynomials, it is a simple matter to expand \( G_m \) as

\[
G_m (\tilde{R}_1, \tilde{R}_2) = \begin{cases} 
\sum_n \Gamma_n \frac{\tilde{R}_2^n}{\tilde{R}_2^{n+1}} & \text{if } \tilde{R}_2 \geq \tilde{R}_1, \\
\sum_n \Gamma_n \frac{\tilde{R}_1^n}{\tilde{R}_2^{n+1}} & \text{if } \tilde{R}_1 \geq \tilde{R}_2,
\end{cases}
\]

But \( \tilde{R}_1^n/\tilde{R}_2^{n+1} \) and \( \tilde{R}_2^n/\tilde{R}_2^{n+1} \) are eigen-functions for the operators \( \tilde{R}_i \partial / \partial \tilde{R}_i, \quad i = 1, 2 \). It is therefore straightforward to find new \( \gamma_{l_1, l_2} \)'s corresponding to expressions like the brackets in Eq. (A.2.7) and Eq. (A.2.8) so that

\[
A_{l_1, l_2} (e_1, e_2, R_1, R_2) = a_{l_1} (e_1) a_{l_2} (e_2) \sum_n \gamma_{l_1, l_2} g_n (\tilde{R}_1, \tilde{R}_2) ,
\]

where the coupling term of this equation obeys

\[
g_n (R_1, R_2) = \begin{cases} 
g_1^n (R_1) g_2^n (R_2) = \frac{R_1^{2n}}{R_2^{2n+1}} & \text{if } R_2 \geq R_1, \\
g_1^n (R_2) g_2^n (R_1) = \frac{R_2^{2n}}{R_1^{2n+1}} & \text{if } R_1 \geq R_2,
\end{cases}
\]

and the dependency on the eccentricity is – to second order in \( e \) –

\[
a_{l_1} (e_1) = (\delta_{l_1} + \delta_{l_1}) e_1 + \delta_{l_1}.
\]

In principle, the \( \gamma_{l_1, l_2} \) are here functions of \( \tilde{R}_i \) via \( \eta_i \). In practice, it is desirable to restrict the emphasis on the quasi-resonant orbits which contribute significantly to Eq. (2.5.32) as it was argued in section 3. Under these circumstances, \( \Omega_{p_1} \sim \Omega_{p_2} \sim \Omega_p \). In simulations, the pattern speed of the bar, \( \Omega_p \), is usually found well below \( \kappa/2 \). Therefore, the variations of \( \eta_i \) above and below 1 will be neglected here-after.

Equation (A.2.10) is then directly separable in eccentricities and reference radii, which provides a unique opportunity to simplify the integral equation (2.5.32). Putting Eq. (A.2.10) into Eq. (2.5.32) yields

\[
\psi_{l_1} (J_1) = a_{l_1} (e_1) 2 \pi^2 G \sum_{l_2, n} \gamma_{n, l_1, l_2} \int \int a_{l_2} (e_2) g_n (\tilde{R}_1, \tilde{R}_2) \frac{[-\partial F_0 / \partial \eta]_{l_2}}{\Omega_{l_2} - \Omega_p} \psi_{l_2} d^2 J_2 .
\]

\(^2\text{The maximum of } \Omega - \kappa/2 \text{ certainly satisfies this constraint in the inner parts of galaxies.}\)
Switching to the new variables eccentricity and mean radius, \((e, \bar{R})\), and calling the core of the kernel
\[
K_{\ell_2}(\bar{R}_2,e_2) = 2\pi^2 G \left| \frac{\partial J_{\ell_2}}{\partial e \partial \bar{R}_2} \right| \frac{\left[ -\frac{\partial F_0}{\partial h} \right]_{\bar{R}_2}}{\Omega_{\ell_2} - \Omega_p},
\] (A.2.14)
equation (A.2.13) becomes
\[
\psi_{\ell_1}(\bar{R}_1,e_1) = a_{\ell_1}(e_1) \sum_{\ell_2,n} \gamma^{\ell_1,\ell_2}_{n} \int g_n(\bar{R}_1,\bar{R}_2) \left[ \int a_{\ell_2}(e_2) K_{\ell_2}(\bar{R}_2,e_2) \psi_{\ell_2}(\bar{R}_2,e_2) \, de_2 \right] \, d\bar{R}_2.
\] (A.2.15)
Multiplying Eq. (A.2.15) by \(a_{\ell_1}(e_1) K_{\ell_1}(\bar{R}_1,e_1)\) and integrating over \(e_1\) yields
\[
\Psi_{\ell_1}(\bar{R}_1) = H_{\ell_1}(\bar{R}_1) \sum_{\ell_2,n} \gamma^{\ell_1,\ell_2}_{n} \int g_n(\bar{R}_1,\bar{R}_2) \psi_{\ell_2}(\bar{R}_2) \, d\bar{R}_2,
\] (A.2.16)
where the new unknown vector reads
\[
\Psi_{\ell_1}(\bar{R}_1) = \int a_{\ell_1}(e_1) K_{\ell_1}(\bar{R}_1,e_1) \psi_{\ell_1}(\bar{R}_1,e_1) \, de_1, \quad i = 1, 2,
\] (A.2.17)
and the characteristics of the underlying galaxy are all taken care of in \(H_{\ell_1}(\bar{R}_1)\), given by
\[
H_{\ell_1}(\bar{R}_1) = \int a_{\ell_1}^2(e_1) K_{\ell_1}(\bar{R}_1,e_1) \, de_1,
\]
\[
= \int a_{\ell_1}^2(e_1) \left| \frac{\partial J_{\ell_1}}{\partial h} \right| \frac{\left[ -\frac{\partial F_0}{\partial h} \right]_{\bar{R}_1}}{\Omega_{\ell_1} - \Omega_p} \, de_1.
\] (A.2.18)

A.2.2 Softened gravity and Green kernels

At this stage, a supplementary simplifying assumption is taken by assuming that the summation in Eq. (A.2.10) reduces to one term
\[
A_{\ell_1,\ell_2}(e_1, e_2, R_1, R_2) = a_{\ell_1}(e_1) a_{\ell_2}(e_2) \sum_{\ell_2,n} \gamma^{\ell_1,\ell_2}_{n} g_n(\bar{R}_1,\bar{R}_2),
\] (A.2.19)
where the coupling term, \(g_n(\bar{R}_1,\bar{R}_2)\), is given by Eq. (A.2.11). This assumption in effect soften gravity for near resonant orbits of similar size. The precise exponent \(\tilde{n}\) chosen does not constrain what follows. In fact, the only compulsory feature of a model for the coupling factor \(g\) is a jump in the derivative at \(\bar{R}_1 = \bar{R}_2\) which is preserved by ansatz Eq. (A.2.19). This ansatz yields simple analytic eigen-functions corresponding to a “natural” set of functions (only dependent on the laws of gravity), on which the criteria Eq. (A.2.13) is expanded. A more realistic model for the core of gravity would lead to more complex sets of eigen-functions.

Eq. (A.2.16) may then be re-written more elegantly as:
\[
\Psi(R_1) = \int g_n(R_1, R_2) \mathcal{H}(R_2) \cdot \Psi(R_2) \, dR_2,
\] (A.2.20)
where \(\Psi\) is an unknown vector of dimension \(m\), where \(m\) is the order of the resonance, and \(\mathcal{H}\) is a \((2m) \times (2m)\) matrix, given by
\[
\mathcal{H} = \begin{pmatrix}
\gamma(-m,-m)H(-m) & \gamma(-m,-m+1)H(-m+1) & \cdots & \gamma(-m,m)H(m) \\
\gamma(-m+1,-m)H(-m) & \gamma(-m+1,-m+1)H(-m+1) & \cdots & \gamma(-m+1,m)H(m) \\
\vdots & \vdots & \ddots & \vdots \\
\gamma(m,-m)H(-m) & \gamma(m,-m+1)H(-m+1) & \cdots & \gamma(m,m)H(m)
\end{pmatrix}.
\] (A.2.21)
The $\gamma$ coefficients in Eq. (A.2.21) are constants depending on the geometry of the orbit which can be calculated for each resonance, and $H_\ell (\bar{R})$ is given by Eq. (A.2.18). The function $g_n$ in Eq. (A.2.19) has the properties of a Green function because of the discontinuity in its derivative. Hence there exists a differential operator, $\mathcal{L}$, corresponding to the integral equation Eq. (A.2.20) so that: $\mathcal{L} G(\bar{R}_1, \bar{R}_2) = \delta (R_1 - R_2)$. Applying $\mathcal{L}$ to Eq. (A.2.20) yields:

$$\mathcal{L} \Psi = \int \delta (\bar{R}_1 - \bar{R}_2) \mathcal{H}(\bar{R}_2) \cdot \Psi(\bar{R}_2) \, d\bar{R}_2 = \mathcal{H} \cdot \Psi.$$  

(A.2.22)

The operator $\mathcal{L}$ is calculated later. The existence of a complete set of orthonormal eigenfunctions, $u_n(\bar{R})$, and eigenvalues, $\lambda_n$, compatible with the integral equation is also discussed later. These eigenfunctions do not depend on the resonance $\ell$ under consideration. $\Psi_\ell$ may therefore expanded over the same set of eigenfunctions as follows

$$\Psi_\ell = \sum_n x_{\ell, n} u_n.$$  

(A.2.23)

Inserting Eq. (A.2.23) into Eq. (A.2.22) leads to $2m$ relations of the form:

$$\sum_{n'} x_{\ell, n'} \lambda_{n'} u_{n'} + \sum_{n'} \sum_{\ell'} H_{\ell, \ell'} x_{\ell', n'} u_{n'} = 0.$$  

(A.2.24)

Multiplying Eq. (A.2.24) by $u_n$ and integrating over $\bar{R}$ (given that the $u_n$ functions form an orthonormal set), implies

$$x_{\ell, n} \lambda_n + \sum_{n'} \sum_{\ell'} x_{\ell', n'} \int u_n (\bar{R}) H_{\ell, \ell'} (\bar{R}) u_{n'} (\bar{R}) \, d\bar{R} = 0.$$  

(A.2.25)

Calling

$$x = (x_{1, -m}, \ldots, x_{1, m}, \ldots, x_{n, -m}, \ldots, x_{n, m}, \ldots),$$  

(A.2.26a)

$$\Lambda = \text{Diag} (\text{Diag} (\lambda_1, \ldots, \lambda_1), \ldots, \text{Diag} (\lambda_n, \ldots, \lambda_n), \ldots),$$  

(A.2.26b)

$$\mathcal{H} (\Omega_p) = \left( \mathcal{H}^{n, n'} \right) = \left( \int u_n (\bar{R}) \mathcal{H} (\bar{R}) u_{n'} (\bar{R}) \, d\bar{R} \right) = \left( H^{(n, n')} \right),$$  

(A.2.26c)

yields for Eq. (A.2.25)

$$\Lambda \cdot x + \mathcal{H} (\Omega_p) \cdot x = 0.$$  

(A.2.27)

Here $\mathcal{H}$ is given by Eq. (A.2.22). The coefficients $H^{(n, n')}_{(\ell, \ell')}$ are functions of $\Omega_p$ given by

$$H^{(n, n')}_{(\ell, \ell')} = \gamma_{(\ell, \ell')} \pi G \int u_n (\bar{R}) u_{n'} (\bar{R}) \left\{ \int a_\ell^2 (e) \frac{\partial J_\ell}{\partial \Omega} \left[ a_{\ell/2} (\ell) \frac{\partial F_0}{\partial \Omega} \right]^{\ell \frac{\partial F_0}{\partial \Omega}} \, de \right\} d\bar{R}.$$  

(A.2.28)

Equation (A.2.28) may be rearranged as

$$H^{(n, n')}_{(\ell, \ell')} = \gamma_{(\ell, \ell')} \pi G \int \int A_\ell (\mathbf{J}) A_{\ell'} (\mathbf{J}) \left[ a_{\ell/2} (\ell) \frac{\partial F_0}{\partial \Omega} \right]^{\ell \frac{\partial F_0}{\partial \Omega}} d^2 \mathbf{J},$$  

(A.2.29)

where the effective $A_\ell$ are defined by:

$$A_\ell (\mathbf{J}) = a_\ell (e (\mathbf{J})) u_n [\bar{R} (\mathbf{J})].$$  

(A.2.30)

The system of equations Eq. (A.2.27) has a non-trivial solution for the $x_{n, \ell}$’s only if

$$D (\Omega_p) = \det |\Lambda + \mathcal{H} (\Omega_p)| = 0.$$  

(A.2.31)

Eq. (A.2.31) gives the criterion for the existence of exponentially growing unstable modes in the epicyclic approximation. The corresponding results are discussed in the main text.
A.2.3 Reduced eigenvalue problem

The eigenfunctions and eigenvalues of the Sturm Liouville equation corresponding to the integral equation Eq. (A.2.20) are now derived. The r.h.s. of Eq. (A.2.20) reduces to the integral equation

\[ u (\bar{R}) = \int_{\bar{R}}\! g_n (\bar{R}_1, \bar{R}_2) u (\bar{R}_2) d\bar{R}_2, \]  

(A.2.32)

where \( u(\bar{R}) \) is the unknown. Given Eq. (A.2.11), Eq. (A.2.32) may be re-written as:

\[ u (\bar{R}) = g_n^2 (\bar{R}_1) U_1 (\bar{R}_1) + g_n^1 (\bar{R}_1) U_2 (\bar{R}_1), \]  

(A.2.33)

where

\[ U_1 (\bar{R}) = \int_{\bar{R}}\! g_n^1 (\bar{R}) u (\bar{R}) d\bar{R}, \]  

(A.2.34a)

\[ U_2 (\bar{R}) = \int_{\bar{R}}\! g_n^2 (\bar{R}) u (\bar{R}) d\bar{R}. \]  

(A.2.34b)

The \( g_n^i \)'s are given by Eq. (A.2.11). Double differentiation of this equation leads to the system

\[ u = g_n^2 U_1 + g_n^1 U_2, \]  

(A.2.35a)

\[ u' = g_n^2' U_1 + g_n^1' U_2, \]  

(A.2.35b)

\[ u'' = g_n^2'' U_1 + g_n^1'' U_2 + (g_n^2' g_n^1 - g_n^1' g_n^2) u. \]  

(A.2.35c)

Eliminating \( U_1 \) and \( U_2 \) between the first two equations yields the Sturm Liouville eigenvalue problem

\[ \mathcal{L} u \equiv (p u)' + qu = \lambda u, \]  

(A.2.36)

with

\[ p (\bar{R}) = g_n^{3'} g_n^2 - g_n^{2'} g_n^1 = \frac{1 + 4 \bar{n}}{\bar{R}^2}, \]  

(A.2.37a)

\[ q (\bar{R}) = g_n^{1''} g_n^2 - g_n^{2''} g_n^1 - p (\bar{R})^2 = \frac{-1 + 10 \bar{n} + 28 \bar{n}^2 + 16 \bar{n}^3}{\bar{R}^4}. \]  

(A.2.37b)

This differential equation is equivalent to the integral equation Eq. (A.2.32), when associated to the initial conditions which follows from Eq. (A.2.32), namely that \( U_1 = [u g_n^1 - u' g_n^1] \), and \( U_2 = [u g_n^2 - u' g_n^2] \) vanish at the endpoints. The integration of Eq. (A.2.36) yields

\[ u_\lambda (\bar{R}) = A \mathcal{R} J_{-\alpha} (2 \mathcal{R}) + B \mathcal{R} J_{\alpha} (2 \mathcal{R}) \],

(A.2.38)

where \( \mathcal{R} = \sqrt{\frac{-\lambda}{16 (1 + 4 \bar{n})}} \bar{R}^2 \) and \( \alpha = \frac{(5 + 24 \bar{n} + 16 \bar{n}^2)^{1/2}}{4} \).

(A.2.39)

Given the mixed boundary conditions

\[ [u g_n^{1'} - u' g_n^1] (\bar{R}_{\text{min}}) = 0 \quad \text{and} \quad [u g_n^{2'} - u' g_n^2] (\bar{R}_{\text{max}}) = 0, \]  

(A.2.40)

and requiring that Eq. (A.2.38) is normalised reduces the variation in \( \lambda \) to a discrete set of possible eigenvalues. As \( \mathcal{L} \) is hermitian (by construction), the eigenvalues are real. It is therefore legitimate to expand \( \mathcal{H} \) over the eigenfunctions of \( \mathcal{L} \). It can be checked that Eq. (A.2.36) does not allow for a solution compatible with the initial conditions Eq. (A.2.40) when \( \lambda = 0 \). Therefore \( \prod_i \lambda_i \neq 0 \). The convergence of the determinant Eq. (A.2.31) should follow from the asymptotic behaviour of the solution to the differential equation. The asymptotic solutions vary like \( n \cos(n R^2) \). Therefore \( H_{\ell \nu}^{\text{asym}} = o(n^{-1/2}) \) according to Lebesgue-Riemann Lemma.
A brief introduction to angle action formalism

A fair fraction of the content of this subsection was inspired from D. Earn’s [30] PhD. Its purpose is to introduce Angle Action variables and generating functions, and their relevance to galactic dynamics (See Born (27) [16] for references).

Let $q$ and $p$ be some known coordinates and momentum for Hamilton’s equation. They obey the equation of motion:

$$\dot{q} = \frac{\partial H}{\partial p} \tag{A.3.1a}$$

$$\dot{p} = -\frac{\partial H}{\partial q} \tag{A.3.1b}$$

where $H$ is the Hamiltonian for this set of variables. This can be expressed as a variational principle, writing

$$\delta \int_{t_1}^{t_2} (p \cdot \dot{q} - H) \, dt = \int_{t_1}^{t_2} \left\{ \delta p \left( \frac{\partial H}{\partial p} \right) - \delta q \left( \frac{\partial H}{\partial q} \right) \right\} \, dt + [p \cdot \delta q]_{t_1}^{t_2}. \tag{A.3.2}$$

The r.h.s of Eq. (A.3.2) vanishes because the variation are assumed to vanish at the end points, and $q$ and $p$ obey Hamilton’s equation.

A new set of variables, say $P$ and $Q$ with $\mathcal{H}$ the new Hamiltonian expressed in these variables, will satisfy formally the same variational principle.

$$\delta \int_{t_1}^{t_2} (P \cdot \dot{Q} - \mathcal{H}) \, dt = 0. \tag{A.3.3}$$
This will have the same physical content as A1 only if there exists a function $S$ so that
\[
P \cdot \dot{Q} - \mathcal{H} = p \cdot \dot{q} - H + \frac{dS}{dt}, \quad (A.3.4)
\]
where $d/dt$ is a total derivative with respect to time.
Restricting possible change of variables to new sets for which $S(q, P) = S_q(q) + S_P(P) - P \cdot Q$, Eq. (A.3.4) is satisfied only if
\[
p = \frac{\partial S}{\partial q}, \quad Q = \frac{\partial S}{\partial P} \quad \text{and} \quad \mathcal{H} = H, \quad (A.3.5)
\]
which can be written in a more compact form as:
\[
H \left( q, \frac{\partial S_q}{\partial q} \right) = \mathcal{H} \left( \frac{\partial S_P}{\partial P}, P \right). \quad (A.3.6)
\]
Notice that $\mathcal{H}$ is a function of $P$ only, by choice of $S$. The new equations of motion therefore read:
\[
\dot{P} = -\frac{\partial \mathcal{H}}{\partial Q} \equiv 0 \quad \text{and} \quad \dot{Q} = \frac{\partial \mathcal{H}}{\partial P}, \quad (A.3.7)
\]
which leads to the immediate solution:
\[
P = \text{constant}, \quad \text{and} \quad Q = \frac{\partial \mathcal{H}}{\partial P} t + Q_0. \quad (A.3.8)
\]
Therefore $S$ is a generating function to a set of variables for which the momenta are integral of the motions.

- Angle and action in galactic dynamics.

Restricting this analysis to multiply periodic bound motion, some new momenta for which the associated coordinates are the phase of the oscillation of each degree of freedom may be constructed. Calling this new set of variables $(\varphi, J)$, it follows
\[
2\pi = \oint d\varphi_i = \oint \left( \frac{\partial \varphi_i}{\partial q} \cdot dq + \frac{\partial \varphi_i}{\partial P} \cdot dp \right) = \oint \left( \frac{\partial}{\partial q} \frac{\partial S}{\partial q_i} \cdot dq + \frac{\partial}{\partial P} \frac{\partial S}{\partial q_i} \cdot dp \right)
\]
\[
= \frac{\partial}{\partial J_i} \oint \frac{\partial S}{\partial q} \cdot dq = \frac{\partial}{\partial J_i} \oint p \cdot dq, \quad (A.3.9)
\]
where the loop integral refers to integration over one complete period of oscillation for the corresponding degree of freedom. This relation will be satisfied if
\[
J_i = \frac{1}{2\pi} \oint p_i dq_i \quad (A.3.10)
\]
These momenta are called actions. They correspond to the circulation of each degree of freedom. In section 3, new momenta were built in the context of galactic dynamics for the radial and azimuthal motion of a star in the disc of the axisymmetric galaxy whose potential is $\Psi_0$. Namely $J = (J_R, J_\varphi) = (J_R, h)$, where
\[
J_R = \frac{1}{2\pi} \oint \left[ \dot{R} \right] dR, J_\varphi = h = R^2 \dot{\varphi}.
\]
Here $[\dot{R}]$ is the function of $R$, the specific energy $E$, and the specific angular momentum $h$ of the star given by
\[
[\dot{R}] = \sqrt{2E + 2\Psi_0 (R) - h^2/R^2}
\]
Actions are relevant to galactic dynamics for the following reasons

(i) the equations of motion are trivial in these variables as the unperturbed actions are by construction integrals of the motion. The perturbation theory is therefore easy to implement;

(ii) the actions label unperturbed orbits;

(iii) the actions can be adiabatically invariant.

Two other concepts central to the analysis of the main sections are the averaging principle and the adiabatic invariance

• The averaging principle

Let $F$ be a distribution function in phase space parameterised by angles and actions. Expressing this function as the sum over its Fourier components with respect to the phase $\varphi$

$$F(J, \varphi, t) = \sum_m F_m(J, t) e^{im\varphi} \quad (A.3.11)$$

it follows that

$$\frac{dF}{dt} = \frac{\partial F}{\partial t} + \frac{\partial H}{\partial J} \cdot \frac{\partial F}{\partial \varphi} - \frac{\partial H}{\partial \varphi} \cdot \frac{\partial F}{\partial J} = 0 \quad (A.3.12)$$

This implies in turn that for each $m$

$$F_m(J, t) = F_m(J) \exp \left( -im \cdot \frac{\partial H}{\partial J} t \right) \quad (A.3.13)$$

If $m \cdot \partial H/\partial J \neq 0$ the time average of $F_m(J, t)$ will be zero. If there exists $J$ so that $m \cdot \partial H/\partial J \equiv m \cdot \Omega = 0$ for a given $m$, i.e. if there is a resonance between the two degrees of freedom, the time average of the corresponding $F_m$ will not vanish. The averaging principle as applied in Section 3 of chapter 1 assumed that the distribution function does not have any resonant component but the inner Lindblad, for all $J$. Strictly speaking, this statement implies that $F \equiv 0$. Nevertheless, it may in practise be assumed that for a sufficiently long time (of the order of $1/|\Omega_m|$), the contribution from the higher $m$’s is negligible. For finite time, $F(J, \varphi, t)$ is therefore well described by $F_0 = \langle F(J, \varphi, t) \rangle$ when restricting to the other low order resonances.

• Adiabatic Invariance

The adiabatic invariance implies that when a system is slowly perturbed, its actions still behave as a first integral to second order in the ratio of the natural frequency of the corresponding phase to the characteristic time scale of the perturbation. This relies on the averaging principle, and therefore assumes no low order resonances are involved between these motions. In the main text, the libration of a given orbit in the potential well of the bar is taken to be an external slowly varying perturbation for the stream motion of a star along its orbit. The circulation corresponding to the latter motion is therefore assumed to be adiabatically invariant. This reads:

$$\frac{\dot{j}}{J} = o \left[ \left( \frac{\Omega_L - \Omega_p}{\kappa} \right)^2 \right]$$
Figure A.4.1: Trajectories of four initially resonant inner Lindblad orbits in an Isochrone disk while self gravity is slowly switched on. Four particles are required to fix the centre of mass frame, but do not prevent the mean angular drift of that frame.

A.4 Numerical N-body and restricted 3-body simulations

Two numerical experiments where carried out to illustrate and clarify the analysis of chapter 2.

The purpose of the first experiment was to verify the donkey or cooperative behaviour of resonant inner Lindblad orbits. A restricted four bodies code was written to account for the interaction between stars describing resonant orbits as prescribed by the mean field. Fig. A.4.1 gives the trajectory of two initially resonant orbits – represented by four stars – in the mean field description, when self gravity is switched on smoothly. The generic behaviour of the orientation of resonant orbits given by $\ddot{\varphi}_2 - \ddot{\varphi}_1 \propto -(\alpha_1 + \alpha_2) \sin (2\varphi_2 - 2\varphi_1)$ is recovered qualitatively, until a close encounter invalidates the description of the orbit in terms of a unique star.

The other experiment corresponds to an N-body simulation of a fully self gravitating disk. The integrator is a tree code developed by Richardson. The qualitative evolution is given on Fig. A.4.2 and Fig. A.4.3. The disk displays spontaneously non-axisymmetric features and rapidly form a strong bar, once some fraction of the angular momentum has been carried away by transient spiral waves.

The stars describe elongated $x1$ orbits oriented along the direction of the bar, in agreement with the models of section 3-5. The pattern speed is found below the maximum pattern speed of inner Lindblad resonances, but this apparent disagreement with Sellwood’s results probably accounts for the different régime investigated and the crudeness of this model$^3$. A more quantitative analysis would undoubtedly require using a better Poisson kernel with symmetrised initial conditions.

---

$^3$ This model explores the non-linear régime of bar formation with a relatively noisy code – a direct force integrator, starting with a set of initial conditions which does not correspond to an equilibrium.
Section A.4 Numerical N-body and restricted 3-body simulations

Figure A.4.2: Iso-density contours in log scale of a self-gravitating Kuzmin disk made of 32,000 particles. The initial conditions correspond to an isotropic pressure disk with a Toomre number, \( Q = 3.36G\Sigma_0/\sigma_R\kappa \), equal to unity.

(so called quiet stars as introduced by Sellwood [101]), along the lines of Earn’s [30] work, whose numerical integrator allows a natural extension of the linear analysis in the non-linear régime.

\[ \bullet \circ \bigcirc \circ \bullet \]
Figure A.4.3: Velocity field of the same self-gravitating Kuzmin disk for the time step displayed on the top right corner of Fig. A.4.2 (Only 8 000 particles are shown here). The stars in the bar describe elongated trajectories along the major axis of the bar.
Stellar dynamics
of round galactic disks

3.1 Introduction

Over the next decade, ample and accurate observational data on the detailed kinematics of nearby disk galaxies will become available. It would be of great interest to link these observations with theoretical models for the underlying dynamics. Independent measurement of the radial and azimuthal velocity distribution functions could, for instance, be contrasted with predictions arising from the gravitational nature of the interaction. Indeed the laws of the motion and the associated conserved quantities together with the assumption that the system is stationary put strong constraints on the possible velocity distributions. This is formally expressed by the existence of an underlying distribution function which characterises the dynamics completely. The determination of realistic distribution functions which could account for observed line profiles is therefore an important project vis a vis the understanding of galactic structure. Producing theoretical models that accounts globally for the observed line profiles of a given galaxy provides a unique opportunity to inspect the current understanding of the dynamics of S0 galaxies. It should then be possible to study quantitatively all departures from the flat axisymmetric stellar models. Indeed, axisymmetric distribution functions are the building blocks of all sophisticated stability analyses, and a good phase space portrait of the unperturbed configuration is often needed in order to assess the stability of a given equilibrium state as shown in chapter II. Numerical N-body simulations also require sets of initial conditions which should reflect the nature of the equilibrium.

In this chapter inversion methods are constructed and implemented. These lead to classical distribution functions compatible with a given surface density (section 2), or a given surface density and a given pressure profile (section 3). This latter technique yields distribution functions that correspond to given line profiles (or rather, to the dispersion in the line profiles induced by the velocity distributions which will loosely be called line profiles in this chapter) which may either be postulated or chosen to match the observations. In fact, only a sub-sample of the observation is required to construct a fully self-consistent model which accounts, in turn, for all the observed kinematics. The corresponding redundancy in the data may be exploited (as sketched in section 4) to address the limitations of the description.
3.2 Distribution functions for a given surface density

The problem of finding the distribution function for an axially symmetrical system may formally be solved by using Laplace transforms (Lynden-Bell 1960 [70]) or by using power series (Fricke 1952 [37]). These methods have been further developed for applications to flat galaxies by Kalnajs (1976) [58] and by Miyamoto (1974) [88] and Hunter & Qian (1992) [96] who chose specific forms of distribution function because the flat problem has no unique solution. In fact there is a whole functional freedom left in the distribution function. Here this freedom is exploited to choose special forms of distribution function which make the solution of the resulting integral equation remarkably simple. Explicit inversions are given for distribution functions whose even parts are of the form

\[ f^+ (\varepsilon, h) = (-\varepsilon)^{n+\frac{1}{2}} G_n(h^2/2) \]

where \( h \) is the specific angular momentum and \( \varepsilon \) the specific energy of a given star. Inversions are also provided when \( f^+ (\varepsilon, h) = h^{2n} F_n(\varepsilon) \). Examples are given and illustrated graphically. The part of \( f \) which is antisymmetric in \( h = Rv_\phi \) cannot be determined from the surface density alone unless it is assumed that all the stars rotate in the same sense about the galaxy, in which case \( f^- = \text{Sign}(h) f^+ \). More generally, if the mean velocity \( \langle v_\phi \rangle \) is specified – say by the azimuthal drift prescription – a similar ansatz for \( f^- \) allows its explicit determination too. The inversion formulae require expressing powers of \( R \) times the surface density \( \Sigma(R) \) as functions of either the potential \( \psi(R) \) or of the related function \( Z(R) = R^2 \psi \). These formulae can only be applied to those density distributions for which the potential in the plane of the matter is known (either self-consistently or not). The other characteristics of the disk follow from the knowledge of \( f \) and cannot be specified \( \text{a-priori.} \)

3.2.1 Derivation

For a flat galaxy all the stellar orbits are confined to a plane and by Jeans’ Theorem the steady state mass-weighted distribution function must be of the form \( f = f(\varepsilon, h) \), where the specific energy \( \varepsilon \) is given by

\[ \varepsilon = \frac{1}{2} \left( v_R^2 + v_\phi^2 \right) - \psi, \]  

and the specific angular momentum, \( h \), reads:

\[ h = R v_\phi. \]  

The surface density, \( \Sigma(R) \), arises from this distribution of stellar orbits provided that the integral of \( f \) overall bound velocities is \( \Sigma(R) \), i.e.

\[ \Sigma(R) = \int \int f(\varepsilon, h) \, dv_R dv_\phi, \]  

where the integral is over the region \( \frac{1}{2}(v_R^2 + v_\phi^2) < \psi \).

The following inversion method assumes that \( \Sigma(R) \) is given and its potential on the plane, \( \psi(R) \) is known. This is achieved either by requiring self-consistency via Poisson’s equation

\[ \nabla^2 \psi = -4 \pi \delta(z) \Sigma(R), \]  

or alternatively by assuming that the disk is embedded in a halo and choosing both \( \Sigma \) and \( \psi \) independently. At constant \( v_\phi \) and \( R \), \( dv_R = d\varepsilon \) and hence, using Eq. (3.2.1) and Eq. (3.2.2)

\[ dv_R = d\varepsilon \left[ 2(\varepsilon + \psi) - h^2 R^{-2} \right]^{-1/2}. \]  

1 The convention throughout this chapter the minus sign in the definition of the potential.
Furthermore, at constant $R$, $dh = R \, dv_\psi$, and therefore Eq. (3.2.3) reads
\[
\Sigma(R) = \int_{+\sqrt{2R^2}}^{-\sqrt{2R^2}} \int_0^{\frac{\psi}{\sqrt{2}}} 2 \int_0^\infty f(\varepsilon, h) \, d\varepsilon \, dh \, \frac{\sqrt{2} (\varepsilon + \psi)}{R^2 - h^2}. \tag{3.2.6}
\]

The factor of 2 arises because $v_R$ takes both positive and negative values which are mapped into the same range of $\varepsilon$ which depends only on $v_R^2$.

Splitting $f$ into its odd and even parts:
\[
f_+(\varepsilon, h) = \frac{1}{2} [f(\varepsilon, h) + f(\varepsilon, -h)] \]
\[
f_-(\varepsilon, h) = \frac{1}{2} [f(\varepsilon, h) - f(\varepsilon, -h)] \tag{3.2.7}
\]
leads to
\[
\Sigma(R) = 4 \int_0^{\sqrt{2R^2}} \int_0^h \frac{f_+(\varepsilon, h) \, d\varepsilon \, dh}{\sqrt{2} (\varepsilon + \psi) R^2 - h^2}, \tag{3.2.8}
\]
as only the even part of $f$ contributes to Eq. (3.2.6).

- Even component of $f$: first Ansatz

A first ansatz is to look for solutions for $f_+$ which take the form
\[
I \quad f_+(\varepsilon, h) = (-\varepsilon)^{n+\frac{1}{2}} |h| G_n \left(\frac{h^2}{2}\right), \tag{3.2.9}
\]
where $n$ is first assumed to be an integer. Here the $|h|$ is taken out for convenience; its appearance in no way implies that $f_n(\varepsilon, 0) = 0$ because $G_n(h^2/2)$ may behave as $|h|^{-1}$ for small $h$. The quantity $n$ measures the level of anisotropy in the disk. Large $n$ correspond to cold, centrifugally supported disks; small $n$ correspond to isotropic, pressure supported disks. Inserting the ansatz into Eq. (3.2.8) gives
\[
\Sigma(R) = \frac{2^{\frac{3}{2}}}{R} \int_0^{R^2} G_n \left(\frac{h^2}{2}\right) \left[ \int_0^Y \frac{(-\varepsilon)^{n+\frac{1}{2}} d(-\varepsilon)}{\sqrt{Y - (-\varepsilon)}} \right] d \left(\frac{h^2}{2}\right), \tag{3.2.10}
\]
where the effective potential, $Y$, is given by
\[
Y = \psi - \frac{h^2}{2R^2}. \tag{3.2.11}
\]
The inner integral can be transformed to $Y^{n+1} I_n$ where
\[
I_n = \int_0^1 x^{n+\frac{1}{2}} \, dx = \frac{(2n + 1)(2n - 1)...1}{2^{n+1} (n+1)!} \pi, \tag{3.2.12}
\]
and $x$ has been written for $(-\varepsilon/Y)$.

Re-expressing $Y^{n+1}$ via Eq. (3.2.11),Eq. (3.2.10) becomes
\[
\Sigma(R) = 2^{\frac{3}{2}} I_n R^{-1} \int_0^{R^2} \left(\psi - \frac{h^2}{2} R^{-2}\right)^{n+1} G_n \left(\frac{h^2}{2}\right) d \left(\frac{h^2}{2}\right). \tag{3.2.13}
\]
Writing $H$ for $h^2/2$ and $Z$ for $R^2 \psi$, this expression becomes on multiplication by $R^{2n+3}$
\[
S_n \equiv R^{2n+3} \Sigma(R) = 2^{\frac{3}{2}} I_n \int_0^{Z} (Z - H)^{n+1} G_n(H) \, dH. \tag{3.2.14}
\]
Z \equiv R^2 \psi \text{ is known, so } R^{2n+3} \Sigma (R) \text{ can be re-expressed as a function } S_n(Z). \text{ Differentiating Eq. (3.2.14) } n + 2 \text{ times with respect to } Z, \text{ gives:}

\left( \frac{d}{dZ} \right)^{n+2} S_n(Z) = 2^\frac{n}{2} I_n(n+1)! G_n(Z) . \quad (3.2.15)

Hence, substituting \( H \) for \( Z \) and using Eq. (3.2.12) yields

\[ G_n(H) = \frac{1}{\sqrt{2} \left[ (n+\frac{1}{2}) (n-\frac{1}{2}) \ldots \frac{1}{2} \right]} \pi \left( \frac{d}{dH} \right)^{n+2} S_n(H) . \quad (3.2.16) \]

The ansatz Eq. (3.2.9) therefore leads to the solution of the integral equation Eq. (3.2.8) in the form

\[ f_+ = (-\varepsilon)^{n+\frac{1}{2}} |h| G_n \left( \frac{h^2}{2} \right) , \quad (3.2.17) \]

where \( G_n \) is given by Eq. (3.2.16). Eq. (3.2.16) specifies the part of the distribution function which is even in \( h \). Demanding no counter-rotating stars would imply, for instance, \( f = f_+ + f_- = 0 \) for \( h < 0 \). In that case \( f_- = -f_+ \) for \( h < 0 \) which leads to \( f_- = \text{Sgn}(h) f_+(\varepsilon, h) \). Hence

\[ f = f_+ + f_- = \begin{cases} 2f_+(\varepsilon, h) & h > 0 , \\ f_+(\varepsilon, h) & h = 0 , \\ 0 & h < 0 . \end{cases} \quad (3.2.18) \]

These solutions are called maximally rotating disks. The formal solution, Eq. (3.2.16), must be non-negative for all \( h \) so that Eq. (3.2.17) corresponds to a realizable distribution of stars. Within that constraint, the choice of \( n \) is, in principle, free. There are many distribution functions which give the same surface density distribution because there are two functional freedoms in \( f(\varepsilon, h) \) and only one function \( \Sigma(R) \) is given. Here it has been shown that even within the special functional form of ansatz I there is potentially a solution for any chosen integer \( n \). When \( n \) is not an integer, \( n = n_0 - \alpha, \ 0 < \alpha < 1 \), Eq. (3.2.10) may still be inverted using Abel transforms (see appendix B for details), though the final solution involves an integral which might require numerical evaluation

\[ G_n(H) = \frac{\sin(\pi\alpha) 2^{-3/2} \pi^{-1} I_{\alpha}^{-1}}{[(n_0 + 1 - \alpha) (n_0 - \alpha) \ldots (1 - \alpha)]} \left[ \int_{0}^{H} \left( \frac{d}{dZ} S_n \right) \frac{dZ}{(H-Z)^{1-\alpha}} \right] + \left( \frac{d^{n_0+1}}{dZ} S_n \right) \left. \frac{1}{H^{1-\alpha}} \right|_{Z=0} , \quad (3.2.19) \]

where

\[ I_{\alpha} = \int_{0}^{1} \frac{x^{n+\frac{1}{2}} dx}{\sqrt{1-x}} = \frac{\sqrt{\pi} \Gamma(n+\frac{1}{2})}{2 \Gamma(n+3)} . \quad (3.2.20) \]

Alternatively, there are many ways in which it is possible to re-express \( \Sigma(R) \) in the general form

\[ \Sigma(R) = \sum_{n} R^{-2n-3} S_n^* (R^2 \psi) . \quad (3.2.21) \]

For any particular sum, a distribution function \( G_n^*(H) \) can be found which gives rise to that part of the density given by the \( n \)th term \( S_n^* \). Indeed the relationship between \( G_n^* \) and \( S_n^* \) is just that given by Eq. (3.2.16) between \( G_n \) and \( S_n \). Since the original integral equation is linear, it follows that each expression for \( \Sigma(R) \) in the form Eq. (3.2.21) yields a distribution function in the form

\[ f_+ (\varepsilon, h) = \sum_{n} (-\varepsilon)^{n+\frac{1}{2}} |h| G_n^* \left( \frac{h^2}{2} \right) . \quad (3.2.22) \]
Section 3.2 Distribution functions for a given surface density

There is no requirement that each \( G^*_n \) should be positive provided that the sum \( f_+ \) is positive for all \( \varepsilon \) and \( h \). Thus the expression for the density Eq. (3.2.21) with the distribution function Eq. (3.2.22) gives a very considerable extension of the original ansatz Eq. (3.2.9).

- Odd Component of \( f \) and azimuthal drift

When the average azimuthal velocity field is known, a similar inversion procedure is available in order to specify the odd part in \( h \) of the distribution function. This field may be assumed to be given by the azimuthal drift equation which takes the form

\[
\Sigma \langle v_\phi \rangle^2 = p_\phi - p_R \left( \frac{\kappa^2 R^2}{4V_c^2} \right),
\]

where the azimuthal pressure, \( p_\phi \), and the radial pressure, \( p_R \), are derived from the even component of \( f_- \), and the epicyclic frequency and the circular velocity \( \kappa \) and \( V_c \) follow from \( \psi \). Formally, the only difference from the previous analysis is that in place of Eq. (3.2.3), the integral equation relating \( f_- \) and \( \langle v_\phi \rangle \) becomes

\[
\Sigma \langle v_\phi \rangle = \int \int f(\varepsilon, h) v_\phi dv_R dv_\phi.
\]

In place of Eq. (3.2.8), this leads to the equation

\[
R\Sigma \langle v_\phi \rangle = 4 \int_0^{\sqrt{2R^2\phi}} \int_{\frac{h^2}{2h^2}}^{\phi} \frac{f_- (\varepsilon, h)}{\sqrt{2(\varepsilon + \psi)} R^2 - h^2} d\varepsilon dh.
\]

Replacing ansatz I by ansatz II:

\[
I' \quad f_- (\varepsilon, h) = (-\varepsilon)^{n+\frac{1}{2}} \tilde{G}_n \left( \frac{h^2}{2} \right) \text{Sgn}(h),
\]

yields a solution for \( \tilde{G}_n \) which is formally equivalent to Eq. (3.2.16), but with \( \tilde{S}_n \) defined by

\[
\tilde{S}_n (Z) = R^{2n+4}\Sigma \langle v_\phi \rangle.
\]

The extension to solutions for \( f_- \) when

\[
\Sigma \langle v_\phi \rangle = \sum_n R^{-2n-4}\tilde{S}_n^*(Z),
\]

is, of course,

\[
f_- (\varepsilon, h) = \sum_n (-\varepsilon)^{n+\frac{1}{2}} \tilde{G}_n^* \left( \frac{h^2}{2} \right) \text{Sgn}(h).
\]

Equation (3.2.29) yields that part of the distribution function which is antisymmetric in \( h \) and is compatible with the mean azimuthal flow \( \langle v_\phi \rangle \) given ansatz II.

- Alternative Ansatz

In place of ansatz I, consider a different ansatz:

\[
II \quad f_+ (\varepsilon, h) = h^{2n} F_n (\varepsilon),
\]

where \( n \) is also first assumed to be an integer. Inserting Eq. (3.2.30) into Eq. (3.2.8), and reversing the order of the integrations, yields

\[
\Sigma (R) = 4 \int_0^\phi F_n (\varepsilon) \left[ \int_0^{\psi^{1/2}} \frac{h^{2n} dh}{\sqrt{X - h^2}} \right] d\varepsilon,
\]
where

\[ X = 2(\varepsilon + \psi) R^2. \]  

(3.2.32)

The inner integral is \( X^n J_n \), where

\[ J_n = \int_0^1 \frac{x^{2n} dx}{\sqrt{1-x^2}} = \left[ \left( n - \frac{1}{2} \right) \left( n - \frac{3}{2} \right) \ldots \frac{1}{2} \right] \pi / (2n!) \, , \]  

(3.2.33)

and \( x = h/X^{3/2} \). Thus Eq. (3.2.31) becomes, using Eq. (3.2.32) for \( X^n \)

\[ \Sigma(R) = 2^{n+2} J_n R^{2n} \int_{-\psi}^0 (\varepsilon + \psi)^n F_n(\varepsilon) \, d\varepsilon . \]  

(3.2.34)

Re-expressing \( R^{-2n} \Sigma(R) \equiv S_n(\psi) \) yields

\[ \left( \frac{d}{d\psi} \right)^{n+1} S_n(\psi) = 2^{n+2} n! J_n F_n(-\psi) . \]  

(3.2.35)

Writing \( \varepsilon \) for \(-\psi \) leads to a solution for \( F_n(\varepsilon) \)

\[ F_n(\varepsilon) = \frac{2^{-n-1}}{\left[ \left( n - \frac{1}{2} \right) \left( n - \frac{3}{2} \right) \ldots \frac{1}{2} \right] \pi} \left( \frac{d}{d(\varepsilon - \psi)} \right)^{n+1} [S_n(\varepsilon - \psi)] ; \]  

(3.2.36)

hence

\[ f_+ = h^{2n} F_n . \]  

(3.2.37)

Again, Eq. (3.2.18) is used to determine \( f_- \) when no retrograde orbits are allowed. Equation Eq. (3.2.37) was derived independently by Sawamura (1987) [99].

When \( n \) is not an integer, the solution to Eq. (3.2.34) can still be found, though with a supplementary integration (with \( n = n_0 - \alpha, n_0 \) the closest upper integer), to give

\[ F_n(-\varepsilon) = \frac{2^{-n-2}\pi^{-1} J_n \sin \pi \alpha}{[n_0 + 1 - \alpha)(n_0 - \alpha) \ldots (1 - \alpha)]} \left[ \int_0^\varepsilon \left( \frac{d^{n+1}}{d\psi} S_n \right) \frac{d\psi}{\varepsilon - \psi} \right]^{1-\alpha} \left[ + \frac{1}{\varepsilon^{1-\alpha}} \left( \frac{d^n}{d\psi} S_n \right) \right]_{\psi=0} , \]  

(3.2.38)

where

\[ J_n = \int_0^1 \frac{x^{2n} dx}{\sqrt{1-x^2}} = \frac{\sqrt{\pi} \Gamma \left( \frac{1}{2} + n \right)}{2 \Gamma (1 + n)} . \]  

(3.2.39)

The ansatz in Eq. (3.2.30) may be also generalised when \( \Sigma(R) \) is expressed in the form

\[ \Sigma(R) = \sum_n R^{2n} S_n^*(\psi) ; \]  

(3.2.40)

each such decomposition corresponds to a distribution function

\[ f_+(\varepsilon, h) = \sum_n h^{2n} F_n^*(\varepsilon) , \]  

(3.2.41)

where the \( F_n^* \) are given in terms of the \( S_n^* \) by Eq. (3.2.36) with stars inserted on \( F_n \) and \( S_n \).

The above procedure is also straightforward to implement in order to constrain the antisymmetric component of the distribution function. Indeed ansatz II:

\[ II' \quad f_-(\varepsilon, h) = h^{2n-1} \tilde{F}_n(\varepsilon) , \]  

(3.2.42)
may be chosen in place of ansatz I. Then Eq. (3.2.31) becomes

\[ R \Sigma \langle v_\phi \rangle = 4 \int_{-\psi}^{0} \tilde{F}_n (\varepsilon) \int_{0}^{\chi^{1/2}} \frac{h^{2n} dh}{\sqrt{X - h^2}} d\varepsilon. \]  

(3.2.43)

Hence the solution for \( F_n \) is formally identical to Eq. (3.2.36) but with \( \tilde{S}_n \) substituted for \( S_n \), where

\[ \tilde{S}_n (\psi) = R^{1-2n} \Sigma \langle v_\phi \rangle. \]  

(3.2.44)

Again, the integral equation is linear, so if \( \Sigma \langle v_\phi \rangle \) is expressed as

\[ \Sigma \langle v_\phi \rangle = \sum_n R^{2n-1} \tilde{S}_n^* (\psi), \]  

(3.2.45)

then

\[ f_+ (\varepsilon, h) = \sum h^{2n-1} \tilde{F}_n^* (\varepsilon), \]  

(3.2.46)

where the relationship of \( \tilde{F}_n^* \) to \( \tilde{S}_n^* \) is the same as in Eq. (3.2.36).

### 3.2.2 Disk properties

- **Radial velocity dispersion**

  The average radial velocity dispersion, \( \sigma_R^2 \), is an important quantity for the local stability of the disk. It satisfies

  \[ \Sigma \sigma_R^2 = \int \int f (\varepsilon, h) v_R^2 dv_R dv_\phi, \]  

(3.2.47a)

\[ = \frac{4}{R} \int_0^{R^2} \int_0^{\chi^{1/2}} f_+ (\varepsilon, h) \sqrt{2 (\varepsilon + \psi) - \frac{h^2}{R^2}} d\varepsilon dh, \]  

(3.2.47b)

where the integration is carried out over the region \( \frac{1}{2} (v_R^2 + v_\phi^2) < \psi \). Given Eq. (3.2.9) and using ansatz I, Eq. (3.2.47) may be rearranged as

\[ \Sigma \sigma_R^2 = \frac{2 \psi}{R} \int_0^{R^2} G_n \left( \frac{h^2}{2} \right) \left[ \int_0^{Y} (-\varepsilon)^{n+\frac{3}{2}} \sqrt{Y - (-\varepsilon)} d(-\varepsilon) \right] d \left( \frac{h^2}{2} \right), \]  

(3.2.48)

where the effective potential \( Y = \psi - \frac{h^2}{R^2} \). The inner integral is

\[ \int_0^{Y} (-\varepsilon)^{n+\frac{3}{2}} \sqrt{Y - (-\varepsilon)} d(-\varepsilon) = Y^{n+2} I_n, \]  

(3.2.49)

where

\[ I_n = \int_0^1 x^{n+\frac{3}{2}} \sqrt{1-x} dx = \frac{\pi (2n+1)!!}{2^{n+2} (n+2)!}. \]  

(3.2.50)

Re-expressing \( Y^{n+2} \), Eq. (3.2.48) then becomes

\[ \Sigma \sigma_R^2 = \frac{2 \psi I_n}{R^{2n+5}} \int_0^Z (Z - H)^{n+2} G_n (H) dH, \]  

(3.2.51)
where \( H = \frac{h^2}{2} \) and \( Z = R^2 \psi \). The calculation of \( \sigma_{2n}^2(R) \) is therefore straightforward once the function \( G_n \) has been found. Note the similarity between Eq. (3.2.51) and Eq. (3.2.14). In fact this similarity provides a check for self-consistency since one should have

\[
\frac{\partial}{\partial Z} (\sigma_{2n}^2 \Sigma R^{2n+5}) = 2(n+2) \mathcal{I}_n/I_n R^{2n+3} \Sigma = R^{2n+3} \Sigma. 
\]  

(3.2.52)

Putting Eq. (3.2.30) into Eq. (3.2.47) yields, for Ansatz II:

\[
R^2 \Sigma \sigma_R^2 = 4 \int_{-\psi}^{0} F_n(\varepsilon) \left[ \int_{0}^{X^{1/2}} h^2 \sqrt{X-h^2} dh \right] d\varepsilon, \tag{3.2.53}
\]

where \( X = 2(\varepsilon + \psi) R^2 \). The inner integral is \( X^{n+1} \mathcal{J}_n \), where

\[
\mathcal{J}_n = \int_{0}^{1} x^{2n} \sqrt{1-x^2} dx = \frac{\pi}{2n+1} (2n-1)!! + \frac{1}{2n+1} (2n+1)!, \tag{3.2.54}
\]

and \( x = h/X^{1/2} \). Thus Eq. (3.2.53) becomes, using Eq. (3.2.32) for \( X^n \)

\[
\Sigma \sigma_R^2 = 2^{n+3} \mathcal{J}_n R^{2n} \int_{-\psi}^{0} (\varepsilon + \psi)^{n+1} F_n(\varepsilon) d\varepsilon. \tag{3.2.55}
\]

The similarity between Eqs. (3.2.34) and (3.2.55) yields the identity:

\[
\frac{\partial}{\partial \psi} (\Sigma \sigma_R^2 R^{-2n}) = 2(n+1) \Sigma R^{-2n} \mathcal{J}_n/J_n = \Sigma R^{-2n}. \tag{3.2.56}
\]

**Azimuthal velocity dispersion**

The average azimuthal velocity dispersion, \( \sigma_\phi \), is an observable constraint on models which satisfies

\[
\Sigma \left( \sigma_\phi^2 + \langle v_\phi \rangle^2 \right) = \Sigma \left( \langle v_\phi^2 \rangle \right) = \int \int f(\varepsilon, h) v_\phi^2 dv_R dv_\phi, \tag{3.2.57a}
\]

\[
= \frac{4}{R^3} \int_{0}^{\sqrt{2R^2\psi}} \int_{h^2/2}^{\sqrt{2(\varepsilon + \psi) - h^2}} f_+ (\varepsilon, h) h^2 d\varepsilon dh, \tag{3.2.57b}
\]

where the mean azimuthal velocity \( \langle v_\phi \rangle \) is given in terms of \( f_- \) by linear combinations of Eqs. (3.2.25) and (3.2.43).

Putting Eq. (3.2.9) into Eq. (3.2.57b) and following the substitutions Eq. (3.2.6)-(3.2.16) yields, for the first Ansatz,

\[
R^{2n+5} \Sigma \left( \sigma_\phi^2 + \langle v_\phi \rangle^2 \right) = 2^2 I_n \int_{0}^{Z} (Z-H)^{n+1} H G_n(H) dH, \tag{3.2.58}
\]

where \( H = \frac{h^2}{2} \) and \( Z = R^2 \psi \). \( I_n \) and \( G_n \) are given by Eq. (3.2.12) and (3.2.16).

Putting Eq. (3.2.30) into Eq. (3.2.57b) gives, for the second ansatz,

\[
\Sigma \left( \langle v_\phi^2 \rangle \right) = \frac{4}{R^2} \int_{-\psi}^{0} F_n(\varepsilon) \left[ \int_{0}^{X^{1/2}} \frac{h^2 dh}{\sqrt{X-h^2}} \right] d\varepsilon, \tag{3.2.59}
\]
where \( X = 2 (\varepsilon + \psi) R^2 \), and \( F_n \) is given by Eq. (3.2.36). The inner integral is \( X^{n+1} J_{n+1} \), where \( J_n \) is given by Eq. (3.2.33). Therefore

\[
\Sigma \langle v_\phi^2 \rangle = 2^{n+1} J_{n+1} R^{2n} \int_{-\psi}^{0} (\varepsilon + \psi)^{n+1} F_n (\varepsilon) \, d\varepsilon.
\] (3.2.60)

Note that \( \sigma_R, \psi, \Sigma \) and \( \langle v_\phi^2 \rangle \) are related via the equation of radial support:

\[
\langle v_\phi^2 \rangle - R \frac{\partial \psi}{\partial R} = \frac{\partial (R \Sigma \sigma_R^2)}{\Sigma \partial R}.
\] (3.2.61)

The properties of the disks constructed with ansatz I & II are illustrated in Figs. 3.2.2 and Fig. 3.2.1 for the examples described below.

### 3.2.3 Examples

- Isochrone disks

Distribution functions for the flat Isochrone are simplest with the assumptions of Ansatz II. In the equatorial plane, the Isochrone potential reads

\[
\psi = GM/ (b + r_b) \quad \text{where} \quad r_b^2 = R^2 + b^2.
\] (3.2.62)

The corresponding surface density is

\[
\Sigma = \frac{Mb}{2\pi R^3} \ln [(R + r_b)/b] - R/r_b = \frac{Mb}{2\pi R^3} L.
\] (3.2.63)

But \( R^2 \psi = Z = (r_b^2 - b^2) GM/(r_b + b) = GMb(s - 1) \), where \( s \) is the dimensionless variable \( r_b/b \). Therefore

\[
\Sigma R^{2n+3} = \frac{Mb}{2\pi} R^{2n} L(s) = \frac{Mb^{n+1}}{2\pi} (s^2 - 1)^n L(s),
\] (3.2.64)

where \( L(s) = \ln(\sqrt{s^2 - 1} + s) + \sqrt{s^2 - 1}/s \). Note that \( L'(s) \) is quite simple: \( L'(s) = \sqrt{s^2 - 1}/s^2 \). With this notation,

\[
\left( \frac{\partial}{\partial Z} \right)^{n+2} (R^{2n+3} \Sigma) = \frac{Mb^{n+1}}{2\pi (GMb)^{n+2}} \left( \frac{\partial}{\partial s} \right)^{n+2} [(s^2 - 1)^n L(s)],
\] (3.2.65)

which implies, in turn, that

\[
G_n \left( \frac{h^2}{2} \right) = \frac{M^n b^{n-1} \Gamma(n + 1)}{2^{n/2} \pi GMb} \left( \frac{\partial}{\partial s} \right)^{n+2} [(s^2 - 1)^n L(s)] \bigg|_{s=1+h^2/2}.
\] (3.2.66)

Therefore, the final distribution function (or rather its symmetric part, \( f_+ \)) therefore reads

\[
f_+ (\varepsilon, h) = \frac{-\varepsilon/(GM/b)^{n+1/2}}{4\pi^2 Gb (n + 1/2)!!} \left( \frac{h^2}{2 GMb} \right)^{1/2} \left( \frac{\partial}{\partial s} \right)^{n+2} [(s^2 - 1)^n L(s)].
\] (3.2.67)

with \( s = 1 + h^2/(2GMb) \). For example, the first two distribution functions are given by (in units of \( GM = b = 1 \)):

\[
f_{+,0} (\varepsilon, h) = \frac{\sqrt{2} (-\varepsilon)^{1/2} (-4 + 4 h^2 + h^4)}{(2 + h^2)^3 \sqrt{4 + h^2} \pi^2},
\] (3.2.68a)

\[
f_{+,1} (\varepsilon, h) = \frac{2 \sqrt{2} (-\varepsilon)^{3/2} (40 + 8 h^2 + 18 h^4 + 8 h^6 + h^8)}{3(2 + h^2)^4 \sqrt{4 + h^2} \pi^2}.
\] (3.2.68b)
Stellar dynamics of round galactic disks Chapter 3

Figure 3.2.1: the “Toomre Q number” $Q = \frac{\sigma R \kappa}{(3.36 \Sigma_0)}$ against radius $R$ for the Isochrone model described by Eqs. (3.2.68) and (3.2.70). The parameter $n$ in Eq. (3.2.67) corresponds to a measure of the temperature of the disk.

Using these distribution functions, Toomre’s Q criterion for radial stability ($Q = \frac{\sigma R \kappa}{3.36 G \Sigma_0}$) may be calculated via Eq. (3.2.51). Note that for the Isochrone potential

$$\sigma^2(R) = \frac{2^{5/2} T_n}{R^{2n+3} \Sigma(R)} \int_0^Z (Z-H)^{n+2} \left( \frac{\partial}{\partial s} \right)^{n+2} L(s) (s^2 - 1)^n \left|_{s=H+1} \right. \left. dH. \right) \tag{3.2.69}$$

For instance

$$\Sigma \sigma^2_{R,0} = \frac{1}{2 \pi R^6} \left[ -2 R \arctan(R) + r \log(r + R) \right], \tag{3.2.70a}$$

$$\Sigma \sigma^2_{R,1} = \frac{1}{12 \pi R^7} \left[ 10 R - \frac{8 R^3}{3} - 6 \arctan(R) + 2 r (-2 + R^2) \log(r + R) \right], \tag{3.2.70b}$$

where $r = \sqrt{1 + R^2}$.

- Toomre-Kuzmin disks

  Ansatz II is also appropriate when seeking distribution functions for the Toomre-Kuzmin disks. In the equatorial plane, the potential of the disk reads

$$\psi = -\frac{GM}{r_b}, \quad \text{with} \quad r_b = \sqrt{b^2 + R^2}, \tag{3.2.71}$$

while the surface density is

$$\Sigma = \frac{1}{2\pi} \frac{GM b}{r_b^3}. \tag{3.2.72}$$
Ansatz II yields the distribution (in units of $b$ and $GM/b$)

$$f^\infty_+ (\varepsilon, h) = \frac{2^{n-2} h^{2n}}{\left(n - \frac{1}{2}\right) \left(n - \frac{3}{2}\right) \ldots \left(n - \frac{2n + 1}{2}\right) \pi^2} \left[\frac{s^{2n+3}}{(1 - s^2)^n}\right]^{(n+1)}_{s = -\varepsilon}.$$  

(3.2.73)

For instance,

$$f^\infty_{+, 0} (\varepsilon, h) = \frac{3(-\varepsilon)^2}{4 \pi^2},$$

$$f^\infty_{+, 1} (\varepsilon, h) = h^2 \left[-5 + (\varepsilon)^2 \right] \left[7 + (\varepsilon)^4 \right]$$

$$f^\infty_{+, 2} (\varepsilon, h) = h^4 \left[-5 + (\varepsilon)^2 \right] \left[7 + (\varepsilon)^4 \right]$$

$$f^\infty_{+, 3} (\varepsilon, h) = h^6 \left[-5 + (\varepsilon)^2 \right] \left[7 + (\varepsilon)^4 \right]$$

(3.2.74a, 3.2.74b, 3.2.74c)

This example illustrates the drawbacks of Ansatz II. The factor $h^{2n}$ in Eq. (3.2.30) removes all zero angular momentum orbits. Self-consistency yields a solution with circular orbits near the angular momentum origin ($f$ scales like powers of $h^2/[1 - (-\varepsilon)^2]^{3/2}$). In position space, this implies that the inner core of the galaxy is made of circular orbits! This qualitative feature induces rising radial pressure curves given by

$$\Sigma \sigma^2_{R, 0} = \frac{\psi^4}{8 \pi},$$

$$\Sigma \sigma^2_{R, 1} = \frac{-2 + \psi^2 + \psi^4}{8 \pi} + \frac{(-1 + \psi^2) \log (1 - \psi^2)}{4 \pi \psi^2},$$

$$\Sigma \sigma^2_{R, 2} = \frac{6 - 9 \psi^2 + 2 \psi^4 + \psi^6}{8 \pi \psi^2} + \frac{3 (-1 + \psi^2)^2 \log (1 - \psi^2)}{4 \pi \psi^4},$$

(3.2.75a, 3.2.75b, 3.2.75c)

for the examples above.

Realistic distribution functions will therefore require superpositions of solutions such as (3.2.74), in the spirit of Eqs. (3.2.41) or (3.2.22) as illustrated by Fig. 3.2.2. It is also clear from Eqs. (3.2.75) that $\Sigma \sigma^2_R$ does not fall off fast enough to yield cold disks (it ought to scale like $\psi^6$ to lead to constant $Q$ profiles).

• Power law disks

The simplicity of power law disks leads to solutions for the inversion problem which may be extended to models with a continuous free “temperature” parameter as described above. Recall that this parameter is a measure of the “temperature” of the disk. This temperature may therefore be varied continuously between the two extremes corresponding to the isotropic and cold disk.

The potential and surface density for the power law disk read

$$\psi = R^{-\beta} \quad \text{and} \quad \Sigma = \frac{S_{\beta}}{2\pi} R^{-\beta-1}, \quad (0 < \beta < 1),$$

(3.2.76)

where $S_{\beta}$ is given by:

$$S_{\beta} = \beta \left(\frac{\Gamma[1/2 + \beta/2]}{\Gamma[1/2 - \beta/2]}\right) \left(\frac{\Gamma[1 - \beta/2]}{\Gamma[1 + \beta/2]}\right).$$

(3.2.77)
Figure 3.2.2: radial (solid line) and azimuthal (dotted line) pressure for a Kuzmin disk constructed by linear superposition of Eqs. (3.2.75) corresponding to the distribution functions given by Eqs. (3.2.74).

It is assumed here that distances are expressed in terms of $R_0$, the reference radius, and energies in terms of $\psi_0 = \psi(R_0)$. (in units of $G = 1$). The inversion method corresponding to Ansatz I may be carried out and, using (3.2.17), yields the distribution function

$$f_+ (\varepsilon, h) = S_\beta (-\varepsilon)^{n+1/2} |h| \left[ \frac{2^{-5/2} \pi^{-3/2} \Gamma [\alpha + 1]}{\Gamma [n + 3/2] \Gamma [\alpha - n - 1]} \right] \left( \frac{h^2}{2} \right)^{\alpha - n - 2}$$

(3.2.78)

where $\alpha = 2n + 2 - \beta/(2 - \beta)$. The parameter $n$ is varied continuously in the range $[(3\beta - 2)/(4 - 2\beta), \infty]$ to account for the relative amount of pressure support in the disk. The velocity dispersions follow from Eqs. (3.2.19) and Eq. (3.2.39):

$$\sigma_\phi^2 + \langle v_\phi \rangle^2 = \frac{n \beta}{(2 - \beta + n)} R^{-\beta},$$

(3.2.79a)

$$\sigma_R^2 = \frac{2 - \beta}{2 (2 - \beta + n)} R^{-\beta}.$$  

(3.2.79b)

By construction, these dispersions satisfy the equation of radial equilibrium:

$$\langle v_\phi^2 \rangle - R \frac{\partial \psi}{\partial R} = \frac{\beta (2 - \beta)}{(2 - \beta + n)} R^{-\beta} = \frac{\partial (R \Sigma \sigma_R^2)}{\Sigma \partial R}.$$  

(3.2.80)

Toomre’s number follows from Eqs. (3.2.76) and (3.2.79b)

$$Q = \frac{\sigma_\phi \kappa}{\pi G \Sigma} = \frac{2 - \beta}{S_\beta} \left[ \frac{2\beta}{2 - \beta + n} \right]^{1/2},$$

(3.2.81)
Section 3.3 Distribution functions for given kinematics.

where \( S_\beta \) is given by Eq. (3.2.77). Here, the effect of the parameter \( n \) on temperature is obvious. These results for the power-law disks were also derived by Evans [33] who used the ansatz II for the inversion.

- Generalised Mestel disk

The generalised Mestel disk has more realistic features such as a very flat rotation curve. Its potential and surface density read (in natural length and energy units):

\[
\psi(R) = \frac{1}{R} \log [r + R] \quad \text{and} \quad \Sigma(R) = \frac{1}{2 \pi r} \frac{1}{(r + 1)},
\]

where \( r = \sqrt{1 + R^2} \). Ansatz I yields the following distribution:

\[
f_{+,0}(\varepsilon, h) = \left( -\varepsilon \right)^{3/2} \left| h \right| \left[ \frac{3 R (4 + 4 r + 3 R^2 + r R^2)}{(1 + r)^3 (1 + R^2) (R + r \log [r + R])^3} \log [r + R] \right. \\
\left. - \frac{R^4 (9 + 9 r + 12 R^2 + 8 r R^2 + 3 R^4 + r R^4)}{(1 + r)^3 (1 + R^2)^2 (R + r \log [r + R])^3} \right] R \log [R + r] = h^2/2 \quad (3.2.83)
\]

Note that for these disks, \( G_n \) is a function of \( h \) via the implicit relation \( R \log [R + r] = h^2/2 \).

Figure 3.2.3: the function \( G_n(h) \) – defined by Eq. (3.2.16), \( n = 0, 1, ..4 \) from bottom to top, for the generalised Mestel disk defined by Eq. (3.2.82). The \( n = 0 \) distribution function is given by Eq. (3.2.83).
### 3.3 Distribution functions for given kinematics.

The above Ansatz I & II lead to recursive expressions for \( f \) which start at the hot disk end. This artifact of the construction scheme is unfortunate because observed disks are fairly cold. Mathematical simplicity was achieved at the expense of a priori realism. For instance it was unclear in advance that any of the above inversions would lead to a positive distribution function. It is desirable to use the functional freedom left in \( f \) in order to construct a distribution function which accounts for all the kinematics, either observed or desired (i.e. which accounts for the line profiles observed, or which are marginally stable to radial modes). Here a different technique is suggested which addresses most of these drawbacks.

The philosophy behind this global inversion method is to introduce an intermediate observable, \( F_\phi(R, v_\phi) \), the distribution function for the number of stars which have azimuthal velocity \( v_\phi \) at radius \( R \) (or alternatively \( G_\varepsilon(\varepsilon, R) \) the distribution function for the number of stars which have specific energy \( \varepsilon \) at radius \( R \)). This technique may be applied to the reconstruction of distribution functions accounting for observed line profiles. In fact, given some symmetry assumptions about the shape of an observed galaxy, it is shown that only a subset of the available line profiles – say the azimuthal velocity distribution – is required to re-derive the complete kinematics. Therefore this method provides a general procedure for constructing self-consistent models for the complete dynamics of disk galaxies. These models predict radial velocity distributions which may, in turn, be compared with observations as discussed in section 4.

Alternatively, a ‘natural’ functional form for \( F_\phi \) or \( G_\varepsilon \) may be postulated and parameterised so that it is compatible with imposing the surface density, the average azimuthal velocity and the azimuthal pressure profiles (or equivalently the Toomre number \( Q \), as \( Q \) follows from the equation of radial support). The distribution function \( f(\varepsilon, h) \) of this disk follows in turn from \( F_\phi \) or \( G_\varepsilon \) via simple Abel transforms. This prescription is very general and especially useful when setting the initial conditions of numerical N-body simulations such as those described in the appendix of the previous chapter.

#### 3.3.1 Inversion via line profiles

The number of stars which have azimuthal velocity \( v_\phi \) within \( dv_\phi \) at radius \( R \) reads

\[
F_\phi(R, v_\phi) = \int f(\varepsilon, h) \, dv_\phi = \sqrt{2} \int _{-Y}^{0} \frac{f(\varepsilon, h)}{\sqrt{\varepsilon + Y}} \, d\varepsilon ,
\]

(3.3.1)

where the effective potential, \( Y \), is given by \( Y = \psi - \frac{h^2 R^2}{2} = \psi - \frac{v_\phi^2}{2} \). The line profile \( F_\phi(R, v_\phi) \) may be expressed in terms of \((h, Y)\). Indeed the identity \( Y = \psi - \frac{h^2 R^2}{2} \) may be solved for \( R \) which yields \( R(h, Y) \). The azimuthal velocity \( v_\phi \) becomes in turn a function of \( h, Y \) defined by \( v_\phi = h/R(h, Y) \).

Calling \( \tilde{F}_\phi(h, Y) \equiv F_\phi(R, v_\phi) \) allows the inversion of Eq. (3.3.1) by an Abel transform (cf appendix B):

\[
f(\varepsilon, h) = \frac{1}{\sqrt{2\pi}} \int_0^\varepsilon \frac{\partial \tilde{F}_\phi}{\partial Y} \frac{dY}{\sqrt{(-\varepsilon - Y)}} ,
\]

(3.3.2)

where the partial derivative is taken at constant \( h \). It was assumed here that the distribution \( F_\phi(R, v_\phi) \) vanishes at the escape velocity. Note that \( \tilde{F} \) yields both the symmetric and the antisymmetric parts of the distribution function. The r.h.s of Eq. (3.3.2) is advantageously re-expressed in terms of the variables \((R, h)\) given that (for monotonic integration)

\[
\left( \frac{\partial \tilde{F}_\phi}{\partial Y} \right)_h \, dY = \left( \frac{\partial F_\phi}{\partial R} \right)_h \left( \frac{\partial R}{\partial Y} \right)_h \, dY = \left( \frac{\partial F_\phi}{\partial R} \right)_h \text{Sgn} \left[ \left( \frac{\partial R}{\partial Y} \right)_h \right] dR .
\]

(3.3.3)
Section 3.3 Distribution functions for given kinematics.

Figure 3.3.1: the relationship between integration in terms of the effective potential $Y$ and $R$ integration. The equation $Y = -\varepsilon$ has two roots corresponding to the perigee and the apogee of the star, while $Y = 0$ has two roots corresponding to infinity and $R_e$, the inner bound of an orbit with zero energy and angular momentum $h$. The sign of the slope of $Y(R)$ gives the sign of the contribution of each branch.

This yields two contributions for Eq. (3.3.2)

$$f(\varepsilon, h) = \frac{1}{\pi} \int_{R_e}^{R_p} \frac{\partial F_\phi}{\partial R} \frac{dR}{\sqrt{h^2/R^2 - 2\psi(R) - 2\varepsilon}} - \frac{1}{\pi} \int_{R_e}^{\infty} \frac{\partial F_\phi}{\partial R} \frac{dR}{\sqrt{h^2/R^2 - 2\psi(R) - 2\varepsilon}},$$

where $R_p(h, \varepsilon)$, and $R_a(h, \varepsilon)$ are, respectively, the apogee and perigee of the star with invariants $h$ and $\varepsilon$, and $R_e(h)$ is the inner radius of a star on a “parabolic” (zero energy) orbit with momentum $h$. Note that the derivative in Eq. (3.3.4) is performed keeping $h = R v_\phi$ constant.

Equation (3.3.4) yields the unique distribution function compatible with an observed azimuthal velocity distribution. All macroscopic properties of the flow follow from $f$. For instance the surface density reads

$$\Sigma \equiv \int \left[ \int f_+(\varepsilon, h) \, dv_\phi \right] \, dv_\phi = 2 \int_0^{v_e} F_\phi(R, v_\phi) \, dv_\phi,$$

while the radial velocity distribution, $F_R(R, v_R)$, obeys

$$F_R(R, v_R) = \frac{2}{R} R^{(2\psi - v_e^2)^{1/2}} \int_0^{v_e} f \left( \frac{v_R^2}{2} + \frac{h^2}{2R^2} - \psi(R), h \right) \, dh.$$

Both $F_R$ and $\Sigma$ may in turn be compared with the corresponding observed quantities.
3.3.2 Disks with given \( Q \) profiles

From a theoretical point of view, \( F_\phi \) may be chosen to match given constraints such as a specified potential and temperature profile of a disk. The azimuthal pressure \( p_\phi \) is then fixed via the equation of radial support and by the temperature of the disk defined by Toomre’s \( Q \) number:

\[
p_\phi = \Sigma V_c^2 + \frac{\partial (Rp_R) }{\partial R}, \quad \text{and} \quad p_R = Q^2 \frac{\Sigma^3}{\kappa^2}, \tag{3.3.7}
\]

where the circular velocity, the epicyclic frequency and the surface density\(^2\) are given by

\[
V_c^2 = -\left[ \frac{1}{R} \frac{\partial \psi}{\partial R} \right], \quad \kappa^2 = \frac{1}{R^3} \left[ \frac{\partial (RV_\psi)^2}{\partial R} \right], \quad \text{and} \quad \Sigma = \frac{1}{2\pi} \left[ \frac{\partial \psi}{\partial z} \right]. \tag{3.3.8}
\]

The surface density, \( \Sigma \), and the azimuthal pressure, \( \Sigma \langle v_\phi^2 \rangle \), may in turn be expressed in terms of \( F_\phi (R, v_\phi) \) via

\[
\Sigma \equiv \int \left[ \int f_+ (\varepsilon, h) d v_R \right] dv_\phi = 2 \int_{v_\phi}^{v_c} F_\phi (R, v_\phi) dv_\phi, \tag{3.3.9a}
\]

\[
\Sigma \langle v_\phi^2 \rangle \equiv \int \left[ \int f_+ (\varepsilon, h) d v_R \right] v_\phi^2 dv_\phi = 2 \int_{v_\phi}^{v_c} F_\phi (R, v_\phi) v_\phi^2 dv_\phi, \tag{3.3.9b}
\]

where the circular escape velocity, \( v_c \), is equal to \( \sqrt{2\psi} \). The function \( f_+ \) stands for the even component of the distribution function in \( h \). Any function \( F_\phi \) satisfying these moment equations corresponds to a state of equilibrium stable against ring formation when \( Q > 1 \). Realistic choices for \( F_\phi \) are presented in the next sections.

3.3.3 Alternative inversion method

Another intermediate observable, \( G_\varepsilon (R, \varepsilon) \), the distribution function for the number of stars which have specific energy \( \varepsilon \), may also be parameterised to fix the surface density and the average energy density profiles. For external galaxies \( G_\varepsilon \) is not directly observed. For our Galaxy, \( G_\varepsilon \) is measurable indirectly – via the distribution of stars with given radial and azimuthal velocity given by spectroscopy and proper motions – but only yields the even component in \( h \) of the distribution function. However, the corresponding inversion method still provides a route for the construction of models with specified temperature profiles.

The number of stars which have specific energy \( \varepsilon \) within \( d\varepsilon \) at radius \( R \) reads

\[
G_\varepsilon (R, \varepsilon) = \int \frac{f_+ (\varepsilon, h)}{v_R} dv_\varphi = 2 \int_0^X \frac{g (\varepsilon, h^2/2)}{\sqrt{X - h^2/2}} d \left( \frac{h^2}{2} \right), \tag{3.3.10}
\]

where \( X \) was defined in Eq. (3.2.32) as

\[
X = R^2 (\psi + \varepsilon). \tag{3.3.11}
\]

Here the auxiliary function \( g \) is given by\(^3\)

\[
g (\varepsilon, h^2/2) = f_+ (\varepsilon, h) / |h|. \tag{3.3.12}
\]

\(^2\) It is assumed here that the field is self-consistent and obeys Eq. (3.2.4); alternatively \( \Sigma \) may be given independently of \( \psi \) and the method described here still applies.

\(^3\) Note that only the symmetric component of the distribution function follows from Eq. (3.3.12).
Gaussian velocity distributions are desirable both as building blocks to fit measured line profiles and as “realistic” choices for $F_\phi$ in the construction scheme of disks parameterised by their temperature. The former point is discussed in the next section.

The construction of Gaussian line profiles compatible with a given temperature requires a supplementary assumption for the mean azimuthal velocity of the flow, $\langle v_\phi \rangle$, on which the Gaussian should be centred. In addition to the two constraints, Eqs. (3.3.9), this puts a third constraint on $F_\phi$, namely

$$
\Sigma \langle v_\phi \rangle \equiv \int \left[ \int f_-(\varepsilon, h) d v_\phi \right] v_\phi d v_\phi = 2 \int_0^{v_e} F_\phi(R, v_\phi) v_\phi d v_\phi,
$$

where $f_-$ is the odd component in $h$ of the distribution function. Suppose the following functional form for $F$

$$
F_\phi(R, v_\phi) = S(R) W_n(R, v_\phi) \exp \left(-\frac{(v_\phi - v(R))^2}{2\sigma^2(R)}\right),
$$

where the window function $W_n$ is intended to damp $F_\phi(R, v_\phi)$ near the escape velocity $v_e$. A possible expression for $W_n$ is

$$
W_n(R, v_\phi) = \begin{cases} 
\exp \left[-\frac{v_\phi^2}{n^2 (v_e^2(R) - v_\phi^2)} \right], & |v_\phi| < v_e, \\
0 & \text{elsewhere},
\end{cases}
$$

The local mean energy density $\langle \varepsilon \rangle = \Sigma^{-1}(p_R/2 + p_\phi/2) - \psi(R)$ is fixed by the temperature of the disk defined by Toomre’s $Q$ number:

$$
\langle \varepsilon \rangle = \frac{1}{2} \left[ \Sigma V_e^2 + \frac{\partial (R p_R)}{\partial R} + p_R \right] - \psi(R) \quad \text{given} \quad p_R = Q^2 \frac{\pi^2 \Sigma^3}{\kappa^2},
$$

where $\kappa$ and $V_e$ are given by Eq. (3.3.8). The function $G(R, \varepsilon)$ may be expressed in terms of $\varepsilon, X$ via Eq. (3.3.11) and $R = R(\psi)$. Calling $\tilde{G}_\varepsilon(R, X) \equiv G_\varepsilon(R, \varepsilon)$ leads to the inversion of Eq. (3.3.10) by an Abel transform:

$$
g(\varepsilon, h^2/2) = \frac{1}{\sqrt{2}\pi} \frac{h^2/2}{\sqrt{h^2/2 - X}} \int_0^{\psi} \frac{\partial \tilde{G}_\varepsilon}{\partial X} dX.
$$

$f_+(\varepsilon, h)$ follows from $g(\varepsilon, h^2/2)$ via Eq. (3.3.12).

Any function $G_\varepsilon$ satisfying the moment constraint Eqs. (3.3.13) yields, via Eq. (3.3.15) and (3.3.14), a maximally rotating disk stable against ring formation.

### 3.3.4 Implementation: Gaussian line profiles

Gaussian velocity distributions are desirable both as building blocks to fit measured line profiles and as “realistic” choices for $F_\phi$ in the construction scheme of disks parameterised by their temperature. The former point is discussed in the next section.

The construction of Gaussian line profiles compatible with a given temperature requires a supplementary assumption for the mean azimuthal velocity of the flow, $\langle v_\phi \rangle$, on which the Gaussian should be centred. In addition to the two constraints, Eqs. (3.3.9), this puts a third constraint on $F_\phi$, namely

$$
\Sigma \langle v_\phi \rangle \equiv \int \left[ \int f_-(\varepsilon, h) d v_\phi \right] v_\phi d v_\phi = 2 \int_0^{v_e} F_\phi(R, v_\phi) v_\phi d v_\phi,
$$

where $f_-$ is the odd component in $h$ of the distribution function. Suppose the following functional form for $F$

$$
F_\phi(R, v_\phi) = S(R) W_n(R, v_\phi) \exp \left(-\frac{(v_\phi - v(R))^2}{2\sigma^2(R)}\right),
$$

where the window function $W_n$ is intended to damp $F_\phi(R, v_\phi)$ near the escape velocity $v_e$. A possible expression for $W_n$ is

$$
W_n(R, v_\phi) = \begin{cases} 
\exp \left[-\frac{v_\phi^2}{n^2 (v_e^2(R) - v_\phi^2)} \right], & |v_\phi| < v_e, \\
0 & \text{elsewhere},
\end{cases}
$$

The local mean energy density $\langle \varepsilon \rangle = \Sigma^{-1}(p_R/2 + p_\phi/2) - \psi(R)$ is fixed by the temperature of the disk defined by Toomre’s $Q$ number:

$$
\langle \varepsilon \rangle = \frac{1}{2} \left[ \Sigma V_e^2 + \frac{\partial (R p_R)}{\partial R} + p_R \right] - \psi(R) \quad \text{given} \quad p_R = Q^2 \frac{\pi^2 \Sigma^3}{\kappa^2},
$$

where $\kappa$ and $V_e$ are given by Eq. (3.3.8). The function $G(R, \varepsilon)$ may be expressed in terms of $\varepsilon, X$ via Eq. (3.3.11) and $R = R(\psi)$. Calling $\tilde{G}_\varepsilon(R, X) \equiv G_\varepsilon(R, \varepsilon)$ leads to the inversion of Eq. (3.3.10) by an Abel transform:

$$
g(\varepsilon, h^2/2) = \frac{1}{\sqrt{2}\pi} \frac{h^2/2}{\sqrt{h^2/2 - X}} \int_0^{\psi} \frac{\partial \tilde{G}_\varepsilon}{\partial X} dX.
$$

$f_+(\varepsilon, h)$ follows from $g(\varepsilon, h^2/2)$ via Eq. (3.3.12).

Any function $G_\varepsilon$ satisfying the moment constraint Eqs. (3.3.13) yields, via Eq. (3.3.15) and (3.3.14), a maximally rotating disk stable against ring formation.
Self-consistency requires that the unknown functions \((v, \sigma, S)\) of Eq. (3.3.17) are solved in terms of \((v_\phi), \Sigma, p_\phi\) via Eqs. (3.3.9) and Eq. (3.3.16). For practical purposes, the temperature range explored in realistic disk models is such that \((v_c - v)^2/2\sigma^2\) is quite large at all radii; if \(n\) is also chosen so that at temperature \(\sigma\)

\[
n \gg \frac{v_c \langle v_\phi \rangle}{(v_c^2 - \langle v_\phi \rangle^2)(v_c + \langle v_\phi \rangle)},
\]

(3.3.19)

then the tail of the Gaussian function need not be taken explicitly into account. The line profile \(F\) then reads directly in terms of \((v_\phi), \Sigma, p_\phi\):

\[
F_\phi (R, v_\phi) = \frac{\Sigma (R)}{\sqrt{2\pi} \sigma_\phi} \exp \left( -\frac{(v_\phi - \langle v_\phi \rangle)^2}{2\sigma_\phi^2} \right),
\]

(3.3.20)

where \(\sigma_\phi\) is the azimuthal velocity dispersion

\[
\sigma_\phi^2 = p_\phi / \Sigma - \langle v_\phi \rangle^2.
\]

(3.3.21)

The azimuthal pressure \(p_\phi\) follows from the equation of radial support and the temperature of the disk defined by Toomre’s \(Q\) number given by Eq. (3.3.7). The expression of the average azimuthal velocity, \(\langle v_\phi \rangle\), may be taken to be that which leads to an exact asymmetric drift equation:

\[
\Sigma \langle v_\phi \rangle^2 = p_\phi - p_R \left( \frac{k^2 R^2}{4 V_c^2} \right).
\]

(3.3.22)

Equations (3.3.20)-(3.3.22), together with Eqs. (3.3.7) provide a prescription for the Gaussian azimuthal line profile \(F_\phi\) which yield the distribution function \(f(\varepsilon, h)\) of a disk at given temperature profile via Eq. (3.3.4).

### 3.3.5 Example: constant temperature Gaussian Kuzmin disks

For the Kuzmin disk, the above prescription for the parameters of the line profile yields

\[
\sigma_\phi = \frac{Q}{4 (1 + R^2)^{3/4}} \quad \text{and} \quad \langle v_\phi \rangle = \frac{R (256 - 60 Q^2 + 128 R^2 - 21 Q^2 R^2 + 16 R^4)^{1/2}}{4 (1 + R^2)^{3/4} (4 + R^2)},
\]

where \(Q\) is Toomre’s number. From Eq. (3.3.17), \(F_\phi\) becomes as a function of \(h\)

\[
F [h, R] = \left( \frac{2}{\pi} \right)^{1/2} Q^{-1} \exp \left[ -\frac{8 (1 + R^2)^{3/2}}{Q^2} \left( \frac{h}{R} - \frac{R (256 - 60 Q^2 + 128 R^2 - 21 Q^2 R^2 + 16 R^4)^{1/2}}{4 (1 + R^2)^{3/4} (4 + R^2)} \right)^2 \right].
\]

Differentiating with respect to \(R\) gives

\[
\frac{\partial \log F}{\partial R} = (4 + R^2) \left[ \frac{8 h^2 (2 - R^2)}{Q^2 R^3 (1 + R^2)^{1/4}} + \frac{4 R P_6}{3 R} \right] - \frac{2 (1 + R^2)^{7/4}}{12 h F_6 R} + \frac{Q^2 (1 + R^2)^{3/4} (4 + R^2)^3}{6 h F_6 R} \left[ \frac{8 R^2}{Q^2 (1 + R^2)} + \frac{4 R}{2 (1 + R^2)^{7/4}} \right],
\]

(3.3.23)

with

\[
F_6 = \left( \frac{-1024 + 216 Q^2 - 768 R^2 + 106 Q^2 R^2 - 192 R^4 + 7 Q^2 R^4 - 16 R^6}{(256 - 60 Q^2 + 128 R^2 - 21 Q^2 R^2 + 16 R^4)^{1/2}} \right),
\]

(3.3.24a)

and

\[
P_6 = -256 + 60 Q^2 - 192 R^2 + 27 Q^2 R^2 - 48 R^4 - 4 R^6.
\]

(3.3.24b)

The equations (3.3.23) and (3.3.24) together with Eq. (3.3.9) fully characterise a complete family of distribution functions parameterised by their temperature via Toomre’s number.
Distribution functions for given kinematics.

3.3.6 Example: Algebraic Kuzmin Disks

A polynomial in $v_\phi$ for the line profile $F_\phi$ yields analytic distribution functions; the line profile must be parameterised in order to satisfy given constraints on its first two moments. A possible parameterisation reads, as a function of $h = R v_\phi$ and $R$,

$$F_n(R, h) = \left( \alpha + \beta \left( \frac{h}{R} \right)^{2n+2} \right) \left( \frac{h}{R} - v_c^2(R) \right)^n,$$

(3.3.25)

where $\alpha$ and $\beta$ are parameters, and the integer $n$ is a free index which should be varied to reach disks of different temperature. Putting Eq. (3.3.25) into Eqs. (3.3.9) and solving for $\alpha$ and $\beta$ leads to a formal solution which is a function of $h/R$, $v_c(R)$ and $p_\phi/(\Sigma v_c^2)$. Now, for the Kuzmin disks with $Q < 2$, $p_\phi/(\Sigma v_c^2)$ varies between $Q^2/32$ and $1/2$. Indeed $p_\phi$ and $\Sigma v_c^2$ read, in terms of $\psi$,

$$p_\phi(\psi) = \frac{\psi^4 (1 - \psi^2)}{2 \pi} + \frac{Q^2 \psi^6 (-5 - 3 \psi^2 + 12 \psi^4)}{8 \pi (1 + 3 \psi^2)^2}, \quad \text{and} \quad \Sigma v_c^2 = \pi^{-1} \psi^4.$$

(3.3.26)

Requiring the line profile $F_n$ to remain positive in that interval puts constraints the minimum power $n$ compatible with a given $Q$. The previous section showed that distribution functions constructed from polynomials in $h$ lead to disks with temperatures scaling like the inverse of the degree of the polynomial. Indeed, low order polynomials are not likely to produce a good fit to $\langle v_\phi \rangle$. Here too, high values of $n$ are required to reach low values of $Q$. For instance, $Q = 1$ requires $n > 14$. For example, the line profile $F_n$ corresponding to $n = 6$ ($\sqrt{32/11} < Q < 2$)

![Figure 3.3.2: isocontours of the Gaussian distribution functions for the Kuzmin disk with $Q = 1.5$ (bottom panel), $Q = 2$ (top panel). A hotter component is apparent at larger momentum for the $Q = 2$ disk where the contours are more widely spaced.](image)}
when written as a function of $\psi$ and $h$ reads

$$F_6(\psi, h) = \frac{\gamma}{\sqrt{2\pi} \psi^{3/2}} \left( \frac{h^2 \psi^2}{1 - \psi^2} - 2\psi \right)^4 \left[ 1056 \psi^5 - \frac{4199 h^{10} \psi^{10}}{(1 - \psi^2)^5} + \left( \frac{46189 h^{10} \psi^6}{2 (1 - \psi^2)^5} - 1008 \psi \right) \right]$$

where $\gamma = 231/5242880$. The integrand of Eq. (3.3.4) becomes in terms of $\psi$

$$\left( \frac{\partial F_6}{\partial R} \right) \frac{dR}{\sqrt{h^2/R^2 - 2(\psi + \varepsilon)}} = \frac{(\partial F_6/\partial \psi) d\psi}{\sqrt{h^2 R^2 / (1 - \psi^2) - 2(\psi + \varepsilon)}}, \quad (3.3.27)$$

Integrating Eq. (3.3.27) along the path described in Fig. 3.3.1 yields the analytic, though complicated, distribution function corresponding to a disk with Toomre number in the range $\sqrt{32/\Pi} < Q < 2$ (roughly the value measured in the Galaxy).

$$\cdot \circ \circ \circ \cdot \cdot$$

### 3.4 Observational implications

The inversion method described in section 3 may be applied on observational data. The observations may be carried out as follows: consider a disk galaxy seen almost edge on. It should be chosen so that it looks approximately axisymmetric and flat. It may contain a fraction of gas in order to derive the mean potential $\psi$ from the observed H II velocity curve (Alternatively, the potential may be derived directly from the kinematics if the asymmetric drift assumption holds). Putting a slit on its major axis and cross correlating the derived absorption lines with template stellar spectra yields an estimate of the projected velocity distribution which should essentially correspond to the azimuthal line profile $F_\phi(R, v_{\phi})^4$. An estimate of the radial line profile $F_R(R, v_R)$ arises when the slit is placed along the the minor axis. But $F_R(R, v_R)$, given by Eq. (3.3.6), reads

$$F_R(R, v_R) = \frac{2}{R} \int_0^{R(2\psi - \bar{v}_R)^{1/2}} f \left( \frac{v_R^2}{2} + \frac{h^2}{2R^2} - \psi(R), h \right) dh, \quad (3.4.1)$$

while Eq. (3.3.4) gave $f$ expressed as a function of $F_\phi(v_{\phi}, R)$, namely:

$$f(\varepsilon, h) = \frac{1}{\pi} \int_{R_a}^{R_p} \frac{(\partial F_\phi/\partial R)_h dR}{\sqrt{h^2/R^2 - 2\psi(R) - 2\varepsilon}} - \frac{1}{\pi} \int_{R_a}^{\infty} \frac{(\partial F_\phi/\partial R)_h dR}{\sqrt{h^2/R^2 - 2\psi(R) - 2\varepsilon}}, \quad (3.4.2)$$

Therefore, putting Eq. (3.4.2) into Eq. (3.4.1) provides a simple way of predicting the radial line profiles if the azimuthal line profiles are given. The surface density $\Sigma$ follows from both $F_R(v_R, R)$ and $F_\phi(v_{\phi}, R)$ and could also be compared with the photometry of that galaxy. The likely discrepancy between the predicted and the residual data may be used to assess the limitations of the reconstruction scheme. Indeed, the above prescription for the inversion relies on a set of hypotheses for the nature of the flow (axial symmetry, thin disk approximation etc..) and the

---

4 In practice, projection effects ought to be taken into account...
quality of measurements. The following features of an observed galaxy may constrain the scope of this analysis.

- Thick disk, with random motions perpendicular to the plane of the galaxy.

The measured line profile yields a mean value corresponding to the integrated emission through the width of the disk; as the galaxy is not edge on, the result of the cross correlation of the spectra does not give $F_\phi$ and $F_R$ exactly, but rather gives the projection onto the line of sight of the full velocity distributions $F_{\phi,z}(v_\phi, v_z, R)$ and $F_{R,z}(v_R, v_z, R)$. However, it is consistent with the thin disk approximation to assume that the motion in the plane is decoupled from that perpendicular to the plane. The observed line profile, $F_v(R, v)$, then corresponds to the convolution of these functions with the component of the velocity distribution perpendicular to the plane of the galaxy, $F_z(R, v_z)$. Formally, for the line profile measured along the major axis, this reads

$$ F_v(R, v) = \int_{-\infty}^{\infty} F_z[R, v \sin (i) + u \cos (i)] F_\phi[R, v \cos (i) - u \sin (i)] \, du $$  \hspace{1cm} (3.4.3) 

where $v$ and $u$ are the velocities along the line of sight, and perpendicular to the line of sight respectively. A similar expression for the line profile measured along the minor axis involving $F_R$ follows. Here, the angle $i$ measures the inclination of the plane of the galaxy with respect to the line of sight. At this level of approximation the velocity distribution normal to the plane is accurately described by a centred Gaussian distribution; its variance follows roughly from the equation of radial support and is used to deconvolve Eq. (3.4.3) (or practically to convolve the predicted line profile pairs and compare them to the data). Note that a rough estimate of the error induced is $\sigma_z^2 / \sigma_\phi^2 \tan^2 (i) \sim 1\%$ in our Galaxy when $i = 20$ degrees.

- Non-axisymmetric halo or disk, with dust and absorption.

The departure from axisymmetry reduces the number of invariants – angular momentum is not conserved, which invalidates this method. Some fraction of the population of stars may then follow chaotic orbits which cannot be characterised by a distribution function. Some real galaxies either present intrinsic non axisymmetric features such as bars, spiral or lopsided arms. Triaxial halos may induce warps within the disk, hence compromising the evaluation of the kinematics. Non-axisymmetrically distributed dust or molecular clouds may also affect measurement of the light from the stellar disk. On the positive side, non-axisymmetric features often appear as small perturbations which may only contribute weakly to the overall mass distribution. Under such circumstances, an estimate of the underlying distribution function for the axisymmetric component of the disk may be extracted from the data. Sets of initial conditions for an N-body simulation may also be extracted from the distribution function. The time evolution of the code would yield constraints on the stability of the observed galaxy. The above analysis could be carried out on an isolated S0 galaxy and on an Sb galaxy showing a grand design spiral pattern (or alternatively on an Sa presenting an apparent lopsided HI component to isolate plausible different instability modes). The relative properties of the two galaxies and their supposedly distinct fates when evolved forward in time should provide an unprecedented link between the theory of galactic disks and the detailed observation of the kinematics of these objects.

- Data quality.

The method described in section 3 involves a deconvolution of the measured line profiles which yields the velocity distributions which, in turns, are related by Abel transforms. Both transformations are very sensitive to the unavoidable noise in the data. Sources of noise are poor seeing, photon noise, detector noise, thermal and mechanical drift of the spectrograph,
poor telescope tracking, etc... It is therefore desirable to construct from Eqs. (3.4.2), (3.4.1) and (3.3.17) a set of pairs of Gaussian velocity distributions \((F_\phi, F_R)\) on which to project the observed quantities since this will help to assess the errors induced by this noise. Indeed, Eq. (3.4.2) is linear; therefore any superposition of Gaussian profiles corresponding to a good fit of the observed line profile will lead to a distribution function expressed in terms of sums of solutions of Eq. (3.4.2). These Gaussians may in turn be identified with populations at different temperatures – say corresponding to a young component and an older population of stars. Formally this translates as

\[
F_\phi(R, v_\phi) = \sum_i S_i(R) \exp \left( -\frac{[v_\phi - v_i(R)]^2}{2\sigma_i^2(R)} \right),
\]

where the functions \(S_i(R), \sigma_i(R)\) and \(v_i(R)\) should be fitted to the observed azimuthal line profiles. Each component in Eq. (3.4.4) may then be identified with those in Eq. (3.3.20), yielding the temperature profile of that population. This approach should yield the best compromise between fitting the radial and the azimuthal profile while avoiding the deconvolution of noisy data. It also provides directly a least square estimate of the errors.

The observations described in this section may be achieved with today’s technology, as illustrated in Fig. 3.4.1 and Fig. 3.5.1.

**Figure 3.4.1:** Contour plot of 300 s \(R_c\) band CCD image of NGC 7332 (Fisher et al. 1994 [35]). Lowest contour level is 22 \(R_c\) mag arcsec\(^{-2}\) with successive contours at 1 mag intervals. Orientation of the six slit illustrated in Fig. 3.5.1 positions are shown.
3.5 Conclusions

The ansatz presented in section 2 and 3 yield direct and general methods for the construction of distribution functions compatible with a given surface density for the purpose of theoretical modeling. These distribution functions describe stable models with realistic velocity distributions for power law disks, and also the Isochrone and the Kuzmin disks. Simple general inversion formulae to construct distribution functions for flat systems whose surface density and Toomre’s Q number profiles are given in section 4 and illustrated. The purpose of these functions is to provide plausible galactic models and assess their critical stability with respect to global non-axisymmetric perturbations. The inversion is carried out for a given azimuthal velocity distribution (or a given specific energy distribution) which may either be observed or chosen accordingly. When the azimuthal velocity distribution is measured from data, only a subset of the observationally available line profiles, namely the line profiles measured on the major axis, is required to re-derive the complete kinematics from which the line profile measured along the minor axis is predicted. This prediction may then be compared with the observed minor axis line profile.
Figure 3.5.1: NGC 7332 stellar velocity dispersions $\sigma$ (upper panel) and derived velocities (lower panel) plotted against radius in arcseconds along the position angles indicated.
B

B.1 Distribution functions for counter-rotating relativistic disks

In the next chapters, static solutions of Einstein’s equations corresponding to flat disks made of two equal counter rotating streams of stars are presented. The pressures and energy densities of the disks are those induced by a known vacuum field; these may therefore not be chosen independently. The inversion method presented in the previous section – which produces distribution function for a given surface density and pressure profile – deserves a generalisation to the relativistic régime for these objects.

The relativistic flow is characterised by its stress energy tensor, $T^{(\alpha\beta)}$, which, in Vierbein components, reads

$$T^{(\alpha\beta)} = \int\int f(\varepsilon, h) \frac{P^{(\alpha)}}{P^{(t)}} dP^{(R)} dP^{(\phi)}.$$  \hfill (B.1.1)

From the geodesic equation it follows that

$$P^\mu P_\mu = -1 \quad \Rightarrow \quad P^{(R)} = \frac{1}{\gamma_{tt}} \left[ \varepsilon^2 - h^2 \gamma_{tt}/\gamma_{\phi\phi} - g_{tt} \right]^{1/2},$$ \hfill (B.1.2)

where the line element in the disk is taken to be

$$ds^2 = -\gamma_{tt} dt^2 + \gamma_{RR} dR^2 + \gamma_{\phi\phi} d\phi^2.$$ \hfill (B.1.3)

Differentiating Eq. (B.1.2) implies

$$\frac{dP^{(R)}}{P^{(t)}} = \frac{d\varepsilon}{[\varepsilon^2 - h^2 \gamma_{tt}/\gamma_{\phi\phi} - \gamma_{tt}]^{1/2}}, \quad \text{while} \quad dP^{(\phi)} = \frac{dh}{\sqrt{\gamma_{\phi\phi}}},$$ \hfill (B.1.4)

where the differentiation is done at constant $R$ and $h$, and $R$ and $\varepsilon$ respectively. Combinations of Eqs. (B.1.1) lead to

$$\Lambda = \gamma_{\phi\phi}^{1/2} \left( T^{(tt)} - T^{(\phi\phi)} - T^{(RR)} \right) = \int\int \frac{f(\varepsilon, h) dh d\varepsilon}{\sqrt{\varepsilon^2 - h^2 \gamma_{tt}/\gamma_{\phi\phi} - \gamma_{tt}}},$$ \hfill (B.1.5a)

$$\Delta = \gamma_{\phi\phi}^{3/2} T^{(\phi\phi)} = \int\int \frac{h^2 f(\varepsilon, h) dh d\varepsilon}{\sqrt{\varepsilon^2 - h^2 \gamma_{tt}/\gamma_{\phi\phi} - \gamma_{tt}}}.$$ \hfill (B.1.5b)

Note that Eq. (B.1.5b) is precisely Tolmann’s formula corresponding to the total gravitating mass of the disk. Introducing

$$\mathcal{R}^2 = \gamma_{\phi\phi}/\gamma_{tt}, \quad e = \frac{1}{2} (\varepsilon^2 - 1), \quad \Psi = \frac{1}{2} (1 - \gamma_{tt}) \quad \text{and} \quad \hat{f}(e, h) = \frac{f(\varepsilon, h)}{\varepsilon},$$ \hfill (B.1.6)
Suppose that a function \( B.2 \) Abel transform density, pressure law, and rotation law. Frames requires to fix three moments of the velocity distribution to account for the given energy inversion method is also generalised to disks with mean rotation, where the dragging of inertial essentially all static solutions of Einstein equations describing these disks are derived. The above for counter-rotating disks characterised by their stress-energy tensor in the plane. In chapter 4, \( h \), \( e \), Eq. \( 3 \) provides direct and systematic means to construct families of distribution function for counter-rotating disks characterised by their stress-energy tensor in the plane. In fact \( h, e \), Eq. \( 6 \) may be expressed formally in terms of \( f, \hat{f}, R, \phi, \), and \( \bar{f}, R, e, \psi \). In fact \( \Lambda, \Delta, \hat{f}, R, e, \Psi \) tends to \( (R \Sigma, R^3 p_\phi, f, R, e, \psi) \) in the classical regime. Introducing the intermediate functions \( \hat{F}(h, R) \), which is chosen so that its moments satisfy Eqs. \( B.1 \) and following the steps of Eqs. \( 3 \) yields

\[
\hat{f}(e, h) = \frac{1}{\pi} \int_{R_p}^{\infty} \frac{\left(\partial \hat{F} / \partial R\right)_{h}}{\sqrt{h^2/R^2 - 2\Psi - 2e}} dR - \frac{1}{\pi} \int_{R_a}^{\infty} \frac{\left(\partial \hat{F} / \partial R\right)_{h}}{\sqrt{h^2/R^2 - 2\Psi - 2e}} dR,
\]

where \( R_p(h, e) \), and \( R_a(h, e) \) are respectively the apogee and perigee of the star with invariants \( (h, e) \), and \( R_a(h) \) is the inner radius of a star on a “parabolic” (zero energy) orbit with momentum \( h \). Eq. \( B.1 \) provides direct and systematic means to construct families of distribution function for counter-rotating disks characterised by their stress-energy tensor in the plane. In chapter 4, essentially all static solutions of Einstein equations describing these disks are derived. The above inversion method is also generalised to disks with mean rotation, where the dragging of inertial frames requires to fix three moments of the velocity distribution to account for the given energy density, pressure law, and rotation law.

**B.2 Abel transform**

Suppose that a function \( g(t) \) obeys the integral equation

\[
\int_{0}^{\pi} g(t) \frac{dt}{(x-t)^{\alpha}} = f(x), \quad \text{where} \quad 0 < \alpha < 1,
\]

then \( g \) may be expressed formally in terms of \( f \) by

\[
g(t) = \frac{\sin(\pi \alpha)}{\pi} \frac{d}{dt} \left[ \int_{0}^{t} \frac{f(x) \, dx}{(t-x)^{1-\alpha}} \right] = \frac{\sin(\pi \alpha)}{\pi} \left[ \int_{0}^{t} \frac{df}{dx} \frac{dx}{(t-x)^{1-\alpha}} + \frac{f(0)}{t^{1-\alpha}} \right].
\]

This classical result is demonstrated in Binney & Tremaine [14] while substituting Eq. \( B.2 \) in the first part of Eq. \( B.2 \) and integrating by part having interchanged the order of integration and given the identity

\[
\frac{\pi}{\sin(\pi \alpha)} = \int_{0}^{1} \frac{du}{u^{\alpha}(1-u)^{1-\alpha}}.
\]

The transformations Eqs. \( B.2 \) and \( B.2 \) are called Abel transforms and are used throughout chapter 3 and 5, and in the above section.

\[\bullet \circ \bigcirc \bullet \]
4

New solutions of Einstein’s equations:
Relativistic disks

4.1 Introduction

Thin relativistic disks are sources of axisymmetric metrics with gradient discontinuities across the disk. Recent progress in Newtonian potential theory enables one to generate the potential-density pairs of classical axisymmetric disks in closed forms. These results are transposed to construct complete sequences of new solutions of Einstein’s equations which describe super massive disks. To avoid the general-relativistic complications caused by inertial frame dragging (gravo-magnetic forces), the relativistic counterparts of Newtonian disks must have vanishing net angular momentum. Following Morgan & Morgan (1969) [90], one may consider disks to be made of two equal streams of collisionless particles (stars) that circulate in opposite directions around the barycentre. Those disks are described as counter-rotating. Counter rotating disks have recently become more attractive since the observations of NGC 4550 by V.C Rubin et al. (1992) [98], Rix et al. (1992) [97] and NGC 7217 by Merrifield & Kuijken (1993) [85] where the kinematics suggest the existence of such streams.

4.1.1 Background

In Newtonian theory, once a solution for the vacuum field of Laplace’s equation is known, it is straightforward to build solutions to Poisson’s equation for disks by using the method of mirror images. Indeed Kuzmin constructed his potential by considering a point mass placed at a distance $b$ below the centre $R = 0$ of the plane $z = 0$ and it gravitational field above that plane. By reflecting this potential in the plane $z = 0$, he obtained a symmetrical solution of Poisson’s equation everywhere, which has a discontinuous normal derivative in the plane. This jump he identified with the surface density of the disk.

Indeed, the potential of the Kuzmin disks in cylindrical coordinates $(R, z)$ is given by

$$\nu = -M/r_b, \quad \text{where} \quad r_b^2 = R^2 + (|z| + b)^2,$$

and $M$ is the total mass of the disk. The corresponding classical surface density of mass is

$$\Sigma_0 (R) = (4\pi)^{-1} \left[ \frac{\partial \nu}{\partial z} \right]_{0^+}^{0^-} = (2\pi)^{-1} M b / (R^2 + b^2)^{3/2}. \quad (4.1.2)$$
Evans & de Zeeuw [33] pointed out that the linearity of Poisson’s equation allows one to generate general potential density pairs for disks in a closed and compact form via fictitious line density distributions of mass along the negative $z$-axis. The potential is then given by

$$\nu = -\int_{b_1}^{b_2} \frac{W(b) \, db}{\left[R^2 + (|z| + b)^2\right]^{\frac{3}{2}}},$$

where the limits of integration are determined by the interval $<b_1, b_2>$ along the negative $z$-axis, in which the line density $W(b)$ – or “the weight function” – is non-zero. The corresponding surface density of mass on the disk becomes

$$\Sigma_0(R) = \left(2\pi\right)^{-1} \int_{b_1}^{b_2} \frac{W(b) \, b}{(R^2 + b^2)^{\frac{3}{2}}} \, db,$$

The total mass is $\int_{b_1}^{b_2} W(b) \, db$.

For Kuzmin’s disk, $W$ is a delta function, and for the family of Kuzmin-Toomre’s disks it is given by combinations of derivatives of the delta function. In Kuzmin’s picture this corresponds to placing multipoles at a distance $b$ below the centre $R = 0$. In the case of the Kalnajs-Mestel family (Mestel 1963 [87], Kalnajs 1976 [58]),

$$W(b) = \mathcal{M} \, b^{-m}, \quad \text{for} \quad b \geq b_1,$$

$\mathcal{M} = \text{const}$, and $m = 0, 1, \ldots$ in the intervals $<b_1, b_2 = \infty>$. The resulting potential (Eq. (4.1.3)) diverges logarithmically at infinity for $m = 0$, and the total mass of the disk
is infinite for both \( m = 0 \) and \( m = 1 \). However, “truncated” Kalnajs-Mestel disks may be constructed by considering line densities (weight functions) of the form Eq. (4.1.5) only in the intervals \( b_1 < b_2 < \infty \) yielding well-behaved potentials and masses even when \( m = 0 \) or \( 1^1 \).

In the present chapter, static axisymmetric metrics which are solutions of Einstein’s equations are constructed representing the relativistic generalisation of the Kuzmin-Toomre and the Kalnajs-Mestel families of disks. These solutions provide physical sources (interior solutions) for static axisymmetric gravitational fields. The general properties of counter-rotating disks are presented in section 2. The Kalnajs-Mestel disks are analysed in Section 3. The generating function for the second constituent that determines the complete relativistic metric for any member of the Kalnajs-Mestel family is constructed. The explicit forms of the metrics are written down for the truncated Kalnajs-Mestel disks with \( m = 0, m = 1 \) and for the relativistic version of the Isochrone disks \( (m = 2) \). Section 4 gives solutions representing relativistic Kuzmin-Toomre disks. In Section 5, the physical properties of the disks, their velocity curves, the graphs of the specific angular momentum, and both surface mass-energy and rest-mass densities are discussed for typical cases.

4.2 General characteristics of counter-rotating disks

Weyl (1917) \[111\] has shown that axially symmetric solutions of Laplace’s equation in flat space may be used to construct static solutions of Einstein’s equations (cf. Synge (1971) \[100\], for an extensive exposition). If the energy-momentum tensor of a source satisfies the condition \( T^R_R + T^z_z = 0 \), which, hereafter, will be assumed, then the metric in the ‘cylindrical’ coordinates \((t, R, z, \phi)\) can be written in the form

\[
ds^2 = -e^{2\nu} dt^2 + e^{2\zeta-2\nu} (dR^2 + dz^2) + R^2 e^{-2\nu} d\phi^2 ,
\]

(4.2.1)

where, as a consequence of Einstein’s equations, the functions \( \nu(R, z) \) and \( \zeta(R, z) \) satisfy

\[
\nabla^2 \nu = 4\pi e^{2\zeta-2\nu} \left( T^R_R - T^z_z \right) ,
\]

(4.2.2)

Here \( \nabla^2 \) is the standard flat space Laplace operator in cylindrical coordinates. Hence Eq. (4.2.2) implies that outside the matter, \( \nu \) satisfies the flat-space Laplace equation. The potential of any classical disk will therefore serve as a possible solution for \( \nu \). Thus to each classical disk there corresponds a relativistic disk for which \( \nu \) has exactly the same mathematical form when expressed in Weyl’s coordinates. When the parameters are such that the potential is much less than \( c^2 \), the physical properties of the classical and relativistic disks will coincide. However, when the parameters are such that the disks are truly relativistic, this correspondence generates one of many possible relativistic generalisations of the classical disk. Others would be obtained by demanding that the circular velocities should have the same functional form when expressed in terms of some suitably defined radial coordinate (such as circumferential radius) or that the angular velocity or surface density profiles should coincide. However this generalisation leads to exactly tractable mathematics. The matter tensor of these objects are confined to an axially symmetric surface which divides space into two reflectionally symmetric parts. The metric functions are finite everywhere; only their second derivatives have the required singularity on

---

\[1\] In fact, in Einstein’s theory, the vacuum solution generated by the potential of a rod of length \( b_2 - b_1 \) with constant line density equal to \( 1/2 \) is the standard Schwarzschild metric with mass \( M = \frac{1}{2}(b_2 - b_1) \) (Kramer et. al. 1980) \[63\]; corresponding disks are thus of special interest in general relativity.
the disk.
The remaining Einstein equations yield

\[
\frac{\partial \zeta}{\partial R} = R \left[ \left( \frac{\partial \nu}{\partial R} \right)^2 - \left( \frac{\partial \nu}{\partial z} \right)^2 \right], \tag{4.2.3}
\]

\[
\frac{\partial \zeta}{\partial z} = 2R \frac{\partial \nu}{\partial R} \frac{\partial \nu}{\partial z}, \tag{4.2.4}
\]

and

\[
\frac{\partial^2 \zeta}{\partial R^2} + \frac{\partial^2 \zeta}{\partial z^2} - \nabla^2 \nu + \left( \frac{\partial \nu}{\partial R} \right)^2 + \left( \frac{\partial \nu}{\partial z} \right)^2 = 4\pi \epsilon \frac{e^{2\zeta - 2\nu}}{T^t_t + T^z_z}. \tag{4.2.5}
\]

The Laplace equation (4.2.2) is the integrability condition for Eq. (4.2.3) and (4.2.4). When \( R = 0 \) and \( |\nabla \nu| \) has no singularities above \( z = 0 \), then \( \zeta \) is constant along the axis by Eq. (4.2.4). Hence the condition that \( \zeta = 0 \) at infinity ensures that \( \zeta = 0 \) on the axis. Furthermore Eq. (4.2.3) then implies that \( \zeta \) is \( 0(R^2) \) close to the axis which ensures elementary flatness of the metric on the axis.

For infinitesimally thin disks the stress energy tensor components along \( z = 0 \) are of the form \( S(R)\delta(z) \). Integration through the disk gives surface stress tensor

\[
\tau_{\phi} = \int T_{\phi}^\phi e^{\zeta - \nu} dz = \epsilon V^2, \tag{4.2.6}
\]

\[
\tau_t = \int T_t^t e^{\zeta - \nu} dz = -\epsilon, \tag{4.2.7}
\]

where \( V \) is the velocity of the stream as measured by a static observer \( V = \sqrt{g_{\phi\phi} d\phi/d\tau_{\text{loc}}} \) and

\[
\epsilon = \frac{\sigma}{1 - V^2} = \frac{2\varepsilon_0}{1 - V^2}; \tag{4.2.8}
\]

here \( \varepsilon \) and \( \sigma \) are the surface densities of energy and rest mass measured by the static observer, \( \varepsilon_0 \) is the surface proper rest mass density of one stream measured by a co-moving observer. Integrating Eq. (4.2.1) - (4.2.5) across \( z = 0 \) yields

\[
\left[ \frac{\partial \nu}{\partial z} \right]_{0^+}^{0^-} = 4\pi \epsilon \frac{e^{\zeta - \nu}}{T_{\phi}^\phi - \tau_t^t} \equiv 4\pi \epsilon \frac{e^{\zeta - \nu}}{V^2}, \tag{4.2.9}
\]

\[
\left[ \frac{\partial \zeta}{\partial z} \right]_{0^+}^{0^-} = 2R \frac{\partial \nu}{\partial R} \left[ \frac{\partial \nu}{\partial z} \right]_{0^+}^{0^-}, \tag{4.2.10}
\]

and

\[
R \frac{\partial \nu}{\partial R} = \frac{V^2}{1 + V^2}. \tag{4.2.11}
\]

The disk will only be physically realizable in terms of counter-rotating streams of particles if the mass-energy surface density \( \varepsilon \) is positive and if the dominant energy condition is satisfied; from Eq. (4.2.6) and Eq. (4.2.7), this reduces to \( \varepsilon(1 - V^2) \geq 0 \), i.e. \( |V| \leq 1 \). From Eq. (4.2.11) this implies the condition \( R \partial \nu / \partial R \leq 1/2 \) on \( z = 0 \). Not every mathematical potential on the disk will give a realistic counter-rotating disk; the potential must not be too concentrated. In practise this condition gives a limit on how compact one can make the mass distribution corresponding to a given form of disk potential of which total mass is known. From Eq. (4.2.9) the condition that \( \varepsilon \) is not negative is satisfied provided that \( \partial \nu / \partial z \) at \( z = 0_+ \) is not negative at any \( R \). This condition is automatically fulfilled for all the disks considered here because they come from positive mass density Newtonian disk solutions. Indeed, in general this condition is equivalent to the condition that the corresponding Newtonian disk have a non-negative surface density. A
sufficient but not necessary condition for that is that \( W(b) \) be non-negative. In summary, all the disks considered here will have physical sources provided \( V < 1 \); for certain types of disk this places a limit on their compactness.

The magnitude of the velocity of counter-rotating streams, as measured by the local static observers, follows from Eq. (4.2.11):

\[
V = \left[ R \frac{\partial \nu}{\partial R} \left( 1 - R \frac{\partial \nu}{\partial R} \right) \right]^{1/2}.
\] (4.2.12)

The angular velocity of counter-rotation, as measured at infinity, \( \Omega = d\varphi/dt \), is

\[
\Omega = Ve^{2\nu}/R,
\] (4.2.13)

and the specific angular momentum of a particle rotating at radius \( R \) (defined as \( h = p_\varphi/\mu_0 = g_{\varphi\varphi}d\varphi/d\tau \) where \( \mu_0 \) and \( \tau \) are the particle’s rest mass and proper time) is

\[
h = RV e^{-\nu}/(1 - V^2)^{1/2},
\] (4.2.14)

where \( V \) is given by Eq. (4.2.12). The surface mass-energy density, \( \varepsilon \), and the surface rest mass density, \( \sigma \), follow from Eqs. (4.2.8), (4.2.9) and (4.2.12) in the form

\[
\varepsilon(R) = \Sigma_0 e^{\nu - \zeta} (1 - R \partial \nu / \partial R), \quad \text{and}
\]

\[
\sigma(R) = \Sigma_0 e^{\nu - \zeta} \left[ (1 - 2R \partial \nu / \partial R) (1 - r \partial \nu / \partial R) \right]^{1/2},
\] (4.2.15)

where \( \Sigma_0(R) = (4\pi)^{-1}[\partial \nu / \partial \tau]^{0+}_0 \) is the classical density. Finally, for static spacetimes, it is of interest to know the redshift \( z \) of a photon emitted by an atom at rest and received by a static observer at infinity. This redshift factor reads

\[
1 + z = (g_{00})^{-\frac{1}{2}} = e^{-\nu}.
\] (4.2.17)

\[
\bullet \circ \bigcirc \bullet.
\]

### 4.3 Relativistic Kalnajs-Mestel disks

The classical potentials of the Kalnajs-Mestel disks are given by Eq. (4.1.3), where the weight functions are chosen as inverse powers of \( b \), i.e. , according to Eq. (4.1.5).

#### 4.3.1 Derivation

Substituting Eq. (4.1.5) into Eq. (4.1.3), yields the potential of the \( m^{th} \) order Kalnajs-Mestel disks in the form

\[
\nu_m = -\mathcal{M} \int_{b_1}^{b_2} \frac{db}{b^m \left[ R^2 + (|z| + b)^2 \right]^{1/2}}.
\] (4.3.1)

Introducing the new variables \( u \) and \( v \) by

\[
u = R^{-1} \left\{ |z| + b + \left[ R^2 + (b + |z|)^2 \right]^{\frac{1}{2}} \right\}, \quad \text{and} \quad v = 2|z|/R,
\] (4.3.2)
so that
\[
\frac{2}{R} \left( b + |z| \right) = u - \frac{1}{u},
\]  
Eq. (4.3.1) takes the form
\[
\nu_m = -\mathcal{M} \left( \frac{2}{R} \right)^m \int_{u_a}^{u_b} \frac{1}{(u-1/u-v)^m} du,
\]  
where the integration limits are determined by Eq. (4.3.2),
\[
u_a = R^{-1} \left\{ |z| + b_1 + \left[ R^2 + (b_1 + |z|)^2 \right]^{1/2} \right\},
\]  
and a similar expression for \( u_b \) involving \( b_2 \). Writing
\[
I_m = \int_{u_a}^{u_b} \frac{1}{(u-1/u-v)^m} du,
\]  
leads to the recursion relation
\[
\frac{dI_m}{dv} = m I_{m+1},
\]  
which implies
\[
I_m = \frac{1}{(m-1)!} \left( \frac{d}{dv} \right)^{m-1} I_1.
\]  
Finally, introducing \( w \) by
\[
w = (r + |z|)/R, \quad \text{so that} \quad v = w - \frac{1}{w},
\]  
yields an easy integral for \( I_1 \)
\[
I_1 = \int_{u_a}^{u_b} \frac{du}{(u-w)(u+1/w)} = \frac{w}{w^2+1} \left[ \ln \left| \frac{u-w}{u+(1/w)} \right| \right]_{u_a}^{u_b}.
\]  
In terms of \( w \) the recurrence relation Eq. (4.3.8) reads
\[
I_m = \frac{1}{(m-1)!} \left( \frac{w^2}{w^2+1} \frac{d}{dw} \right)^{m-1} I_1.
\]  
The classical potential \( \nu \) of the \( m \)-th order truncated Kalnajs-Mestel disks Eq. (4.3.4) therefore reads
\[
\nu_m = -\mathcal{M} \left( \frac{2}{R} \right)^m I_m,
\]  
where \( I_m \) is given by Eq. (4.3.8) or Eq. (4.3.11), and the generating function \( I_1 \) is given by Eq. (4.3.10), with \( u_1, u_2 \) and \( w \) given by Eq. (4.3.5) and (4.3.9) in terms of \( R \) and \( z \).

Integrating Eq. (4.1.4) by parts, leads to a classical surface density for the \( m \)-th order truncated Kalnajs-Mestel disks in the form
\[
\Sigma_m (R) = \frac{\mathcal{M}}{4\pi} \left\{ \left[ \frac{1}{b_m (R^2 + b_2^2)^{1/2}} \right]_{b_1}^{b_2} - m \left( \frac{2}{R} \right)^{m+1} I_{m+1} (w = 1) \right\}.
\]  
These expressions for \( \nu \) and \( \Sigma \) in the limit \( b_1 = a, b_2 = \infty \) are in exact agreement with those given by Evans & de Zeeuw (1992) [33] for the first four members of the Kalnajs-Mestel family.
The function $\zeta$ is then required in order to obtain the complete relativistic metrics. It may be found by integrating Eq. (4.4.6) with functions $\nu$ given by Eq. (4.3.12). Because Eq. (4.4.6) and Eq. (4.2.10) are quadratic in $\nu$, the formal solution for $\zeta$ should read

$$
\zeta = \int \int W(b_1) W(b_2) Z(b_1, b_2, r_1, r_2) \, db_1 db_2
$$

(4.3.14)

where $Z$ was found by Bicak, Lynden-Bell & Katz (1993) [11] to be

$$
Z = -\frac{1}{2r_1 r_2} \left\{ 1 - \left( \frac{r_2 - r_1}{b_2 - b_1} \right)^2 \right\} \geq 0.
$$

(4.3.15)

Here $b_1$ and $b_2$ are the distances of two points below the disk’s centre and $r_1$ and $r_2$ are the distances measured from these points to a point above the disk. Below the disk, $Z$ is given by reflection with respect to the plane $z = 0$. On the $z$-axis $r_i = b_i + |z|$ and near there $r_i = b_i + |z| + \frac{1}{2} R^2/(b_i + |z|)$. Thus, near the axis, $Z = -\frac{1}{4} R^2 (b_i + |z|)^{-2} (b_i + |z|)^{-2}$, which clearly makes $\zeta$ of order $R^2$ near the axis. This ensures regularity on and near the axis for all such solutions. Eq. (4.3.15) may be rearranged using the variables $u_i$ (cf. Eq. (4.3.2)) as

$$
\zeta = -2 \int_{u_a}^{u_b} \int_{u_a}^{u_b} \frac{W_1 W_2}{u_1 u_2 + 1} u_1 u_2 \, du_1 du_2
$$

(4.3.16)

where

$$
W_i = W \left[ \frac{1}{2} R \left( u_i - \frac{1}{u_i} \right) - |z| \right], \quad i = 1, 2.
$$

(4.3.17)

The limits of integration are given by Eq. (4.3.5). In the Kalnajs-Mestel disks, the function $W$ is just proportional to the inverse power of its argument according to Eq. (4.1.5). Hence, for the $m^{th}$ order truncated Kalnajs-Mestel disks $\zeta$ reads

$$
\zeta_m = -2 \mathcal{M}^2 \left( \frac{2}{R} \right)^{2m} \mathcal{J}_m(u_a, u_b; v_1, v_2 = v)
$$

(4.3.18)

where the function $\mathcal{J}_m$ of four variables is defined by

$$
\mathcal{J}_m(u_a, u_b; v_1, v_2) = \int_{u_a}^{u_b} \int_{u_a}^{u_b} \frac{u_1^m}{(u_2^2 - v_1 u_1 - 1)^m} (u_1 u_2 + 1)^{-2} \frac{du_1 du_2}{(u_2^2 - v_2 u_2 - 1)^m}
$$

(4.3.19)

Recall that $v_1 = v_2 = v = 2|z|/R$; so $\zeta_m$ are indeed functions of $R$ and $|z|$. Inspecting Eq. (4.3.19) leads immediately to the recursion relation

$$
\frac{\partial^2 \mathcal{J}_m}{\partial v_1 \partial v_2} = m^2 \mathcal{J}_{m+1},
$$

(4.3.20)

which implies

$$
\mathcal{J}_m = \frac{1}{[(m-2)!]^2} \left( \frac{\partial^2}{\partial v_1 \partial v_2} \right)^{m-1} \mathcal{J}_1.
$$

(4.3.21)

Hence, all $\zeta_m$’s will be determined once the generating function,

$$
\mathcal{J}_1(u_a, u_b; v_1, v_2) = \int_{u_a}^{u_b} \int_{u_a}^{u_b} \frac{u_1 u_2 du_1 du_2}{(u_1^2 - v_1 u_1 - 1)(u_2^2 - v_2 u_2 - 1)(u_1 u_2 + 1)^2},
$$

(4.3.22)

is known. Since the calculation of $\mathcal{J}_1$ is lengthy and not straightforward – indeed, it is surprising that it can be expressed in terms of elementary functions – its details are given in the Appendix. The final form of $\mathcal{J}_1$ for the general, truncated Kalnajs-Mestel disks is given in Eqs. (C.15)-(C.16).
4.3.2 Examples

- Kalnajs-Mestel $m = 0$ truncated disk

All the recursion formula above are valid for $m \geq 1$. The metric for the $m = 0$ disk must therefore be directly derived from the original expressions: Eq. (4.3.1) for $\nu$, and Eq. (4.3.18) for $\zeta$. The resulting family of disks was first derived by Bicak, Lynden-Bell & Katz (1993) [11] and called the generalised Schwarzschild disks since they are generated by the potential of a rod of length $b_2 - b_1$ with constant line density $\mathcal{M}$; the disk with $\mathcal{M} = 1/2$ corresponds to the standard Schwarzschild metric with mass $M = (b_2 - b_1)/2$. Consider here truncated $m = 0$ Kalnajs-Mestel disks which have well-behaved potentials and finite masses. The metric functions are

$$\nu_0 = N \ln \left| \frac{r_1 + |z| + b_1}{r_2 + |z| + b_2} \right|, \quad \text{and} \quad \zeta_0 = 2N^2 \ln \left| \frac{(r_1 + r_2)^2 - (b_2 - b_1)^2}{4 r_1 r_2} \right|, \quad (4.3.23a, b)$$

where

$$r_i^2 = R^2 + (|z| + b_i)^2, \quad N = M / (b_2 - b_1). \quad (4.3.24)$$

- Kalnajs-Mestel $m = 1$ disk

When $m = 1$, the potential $\nu$ is well-defined even in the un-truncated case. Let $b_1 = a$ and $b_2 \to \infty$. The final expressions for $\nu$ and $\zeta$ are simpler in terms of $r, \mu$ which are related to $R, z$ by

$$R = r \sin \theta, \quad z = r \cos \theta, \quad |\cos \theta| = \mu. \quad (4.3.25)$$

Introducing the constant $K_1$, the dimensionless radial coordinate $\bar{r}$, and the function $P$ by

$$K_1 = 4\pi a \Sigma_0, \quad \bar{r} = r/a, \quad P = [\bar{r}^2 + 2\mu\bar{r} + 1]^{\frac{1}{2}}, \quad (4.3.26)$$

where $\Sigma_0 = \mathcal{M} \pi^{-1} a^{-2}/4$ is the classical central surface density of the disks (cf. Eq. (4.3.13)), yields

$$\nu_1 = -\frac{K_1}{\bar{r}} \ln \left[ \frac{\bar{r} + \mu + P}{1 + \mu} \right], \quad \text{and} \quad (4.3.27a)$$

$$\zeta_1 = K_1 \left\{ \frac{1}{\bar{r}^2} (1 + \mu \bar{r} - P) + \ln \left[ \frac{1 + \mu \bar{r} + P}{P^2} \right] + \frac{1}{\bar{r}^2} [\mu + P (\bar{r} - \mu)] \ln \left[ \frac{\bar{r} + \mu + P}{1 + \mu} \right] \right. \right.$$  \left. \left. + \frac{\mu^2 - 1}{2\bar{r}^2} \left[ \ln \left( \frac{\bar{r} + \mu + P}{1 + \mu} \right) \right]^2 - \ln 2 \right\}. \quad (4.3.27b)$$

The potential $\nu_1$ and $\zeta_1$ were derived by direct integration and also independently by using Eq. (4.3.18) with the generating function Eq. (C.18), (C.19) and substitutions. Note that $\zeta_1 \to 0$ as $r \to \infty$ and $\zeta_1 = 0$ on the $z$-axis ($\mu = 1$). Hence, the spacetime is asymptotically flat and regular on the $z$-axis. (The condition $\zeta = 0$ for $R = 0$ guarantees that the metric Eq. (4.2.1) is regular on the axis – cf. e.g. Synge 1971).

- Kalnajs-Mestel $m = 1$ truncated disk

Although the spacetimes of the un-truncated Kalnajs-Mestel disks with $m = 1$ have reasonable properties, the total mass of each disk is infinite, as can easily be seen by integrating $W(b)$ or $\Sigma_0$ given in Eq. (4.3.13). This defect is avoided with truncated disks, as it was done
for the $m = 0$ case. Let $0 < b_1 < b_2 < \infty$ be two given constants. Starting from the general expressions Eq. (4.3.1), Eq. (4.3.18), the potential $\nu_1^*$ reads

$$\nu_1^* = -\frac{M}{r \ln (b_2/b_1)} \ln \left[ \left( \frac{b_2}{b_1} \right) \frac{r + b_1 \mu + P_1}{r + b_2 \mu + P_2} \right],$$

where $r$ and $\mu$ are given by Eq. (4.3.25), and

$$P_i = (r^2 + 2b_i r \mu + b_i^2)^{1/2},$$

and $M = \mathcal{M} \ln (b_2/b_1)$ is the total mass of a disk. The second metric function, $\zeta_1^*$, is slightly more complicated:

$$\zeta_1^* = -\left[ \frac{2M}{\ln(b_2/b_1)} \right]^2 \frac{2}{R^2J_1^*},$$

where $J_1^*$ is given in Appendix (Eq. (C.20)).

- Kalnajs-Mestel $m = 2$ Isochrone disk

In contrast to the disks with $m = 0$ and $1$, the Isochrone disks have finite mass even in the standard, un-truncated form. In terms of the same variables as Eqs. (4.3.25), and Eq. (4.3.26), the metric functions read

$$\nu_2 = -\frac{M}{a} \left[ \frac{\bar{r} + 2 \mu}{\bar{r} (1 + \mu)} - \frac{\mu}{\bar{r}^2} \ln \left( \frac{\bar{r} + \mu + P^*}{1 + \mu} \right) \right],$$

and

$$\zeta_2 = \left( \frac{M}{a} \right)^2 \frac{\mu}{\bar{r}} \left[ \frac{1}{4 \bar{r}^4} (1 - 3 \mu^2 - 4P) + \frac{1}{2 \bar{r}^3} (-7 + 9 \mu^2 + 3P) \right.
\left. \right.$$

$$+ \frac{1}{2 \bar{r}^4} (-4 + 9 \mu^2) (1 - P) + \ln \left( \frac{1 + \mu \bar{r} + P}{\bar{r}^2} \right)
\left. \right.$$

$$+ \frac{1}{2 \bar{r}^4} [\mu (-7 + 9 \mu^2) (1 - P) + (2 \bar{r}^2 - 2 \mu \bar{r} + 3 \mu^2 - 1) \bar{r} P] \ln \left( \frac{\bar{r} + \mu + P}{1 + \mu} \right)
\left. \right.$$

$$+ \frac{1}{4 \bar{r}^4} (1 - 10 \mu^2 + 9 \mu^4) \left[ \ln \left( \frac{\bar{r} + \mu + P}{1 + \mu} \right) \right]^2 - \ln 2 \right].$$

The potential $\nu_2$, after changing variables ($\bar{r}, \mu$) back to ($r, z$), as stated in Eq. (4.3.26) is equal to the expression given first by Evans & de Zeeuw (1992) [33]. Again, note that $\zeta_2 \to 0$ as $r \to \infty$ and $\zeta_2 = 0$ on the $z$-axis so that the axis is regular and the spacetime with the metric Eq. (4.2.1), in which $\nu$ and $\zeta$ are given by Eq. (4.3.31) and (4.3.32), is asymptotically flat. Direct calculations confirm that the functions $\nu_2$ and $\zeta_2$ satisfy the field equations Eq. (4.2.3)-(4.2.5).

4.4 Relativistic Kuzmin-Toomre disks

The general $n^{th}$ order classical Kuzmin-Toomre disks are described by the potential (Toomre 1963 [107], Nagai & Miyamoto 1976, Evans & de Zeeuw 1992)

$$\nu_{n+3/2} = -\frac{M}{(2n-1)!!} \sum_{k=0}^{n} \frac{(2n-k)! b^k}{2^{n-k} (n-k)!} \frac{1}{r_{b}^{k+1}} P_k (| \cos \theta |),$$

where $P_k$ is the $k^{th}$ Legendre polynomial, $| \cos \theta | = (b + |z|)/b$. The simple potential of the Kuzmin disk is obtained while putting $n = 0$ in Eq. (4.1.1).
4.4.1 Derivation

Equation (4.4.1) can be re-arranged in the form:

$$\nu_{n+\frac{3}{2}} = \sum_{k=0}^{n} \frac{(-1)^k}{k!} U_{k,n} \frac{\partial^k}{\partial b^k} \left\{ \frac{1}{\left[R^2 + (|z| + b)^2\right]^{1/2}} \right\},$$

(4.4.2)

where

$$U_{k,n} = \frac{(2n-k)!}{2^{n-k}(2n-1)!!(n-k)!} M^b_k.$$

(4.4.3)

The potential of the n-th order Kuzmin-Toomre disks at $z > 0$ therefore arises in the Kuzmin picture from the sum of the fields of elementary multipoles of the order $k$, $0 \leq k \leq n$, placed at a distance $b$ below the centre of a plane $z = 0$, the multipole moments being given by Eq. (4.4.3). Correspondingly, the weight function $W$ entering Eqs. (4.1.3) and (4.1.4) becomes

$$W(b, b') = \sum_{k=0}^{n} \frac{(-1)^k}{k!} U_{k,n} \delta^{(k)}(b - b'),$$

(4.4.4)

where $\delta^{(k)}$ denotes the k-th derivative of a $\delta$-function and the integration in Eq. (4.1.3) is over $b'$. From Eq. (4.4.3), it is easy to prove the recurrence relation

$$\nu_{n+1+3/2} = \nu_{n+3/2} - \frac{b}{2n + 1} \frac{\partial}{\partial b} \nu_{n+3/2}.$$  

(4.4.5)

The function $\zeta$ is to be extracted from Eqs. (4.2.3) and (4.2.4) to determine the complete metric outside the matter. For this purpose, the two equations can be combined to give

$$D\zeta = R(D\nu)^2,$$

(4.4.6)

where

$$D = \partial/\partial R - i \partial/\partial z$$

(4.4.7)

is an operator invariant under the translations $z \rightarrow z + b$. Introducing spherical coordinates with the origin anywhere on the z-axis and putting

$$D_b = \partial/\partial \theta - ir \partial/\partial r,$$

(4.4.8)

gives, for Eq. (4.4.6),

$$D_b \zeta = \sin \theta e^{i\theta} (D_b \nu)^2.$$

(4.4.9)

By centring the spherical coordinates at the distance $b$ below the disk, Eq. (4.4.9) takes the form

$$\left( \frac{\partial}{\partial \theta} - ir_b \frac{\partial}{\partial r_b} \right) \zeta = \sin \theta e^{i\theta} \left[ \left( \frac{\partial \nu}{\partial \theta} \right)^2 - 2ir_b \frac{\partial \nu}{\partial \theta} \frac{\partial \nu}{\partial r_b} - r_b^2 \left( \frac{\partial \nu}{\partial r_b} \right)^2 \right],$$

(4.4.10)

where $r_b^2 = R^2 + (|z| + b)^2$ and $\sin \theta = R/r_b$. Substituting $\nu$ given by Eq. (4.4.1) into Eq. (4.4.10) and integrating yields after some manipulation

$$\zeta = \frac{- \sin^2 \theta}{[(2n - 1)!!]^2} \left[ \sum_{k, \ell = 0}^{n} B_{k, \ell, n} b^{k+\ell} P_{k, \ell} (\theta) \frac{1}{r_b^{k+\ell+2}} \right],$$

(4.4.11)

where

$$B_{k, \ell, n} = \frac{(2n-k)!(2n-\ell)!}{2^{2n-\ell}(n-k)!(n-\ell)(k+\ell+2)},$$

(4.4.12)

$$P_{k, \ell} (\theta) = (k+1)(\ell+1) P_k P_{\ell} + 2(k+1) \cos \theta P_k P_{\ell}' - \sin^2 \theta P_k^2 P_{\ell}'$$

and $P_k' = (d/d|\cos \theta|)P_k$. 


4.4.2 Examples

- Kuzmin-Toomre $n = 0$ disk

For the simple Kuzmin disk, $\nu$ is given by Eq. (4.1.1) while $\zeta$ follows from Eq. (4.4.12) or from direct integration of the real component of Eq. (4.4.10):

$$\nu_{3/2} = -\frac{M}{r_b},$$  \hfill (4.4.13)

and

$$\zeta_{3/2} = -\frac{1}{2} M^2 \frac{R^2}{r_b^4}. \hfill (4.4.14)$$

The $\theta$ integration gives the same result since Eqs. (4.2.3) and (4.2.4) are consistent as a consequence of the condition $\nabla^2 \nu = 0$. The metric Eq. (4.2.1) with $\nu$ and $\zeta$ given by Eq. (4.1.1) and Eq. (4.4.14) represents the spacetime of the Kuzmin-Curzon disks.

- Kuzmin-Toomre $n = 1$ disk

The explicit form of the metric functions in the original cylindrical coordinates is the following:

$$\nu_{5/2} = -\frac{M \left( R^2 + z^2 + 3b|z| + 2b^2 \right)}{\left[ R^2 + (b + |z|)^2 \right]^{3/2}}, \hfill (4.4.15)$$

and

$$\zeta_{5/2} = -\frac{M^2 R^2}{2 \left[ R^2 + (b + |z|)^2 \right]^{5/2}} \left[ (R^2 + z^2 + 4b|z| + 3b^2)^2 - \frac{1}{2} b^2 R^2 \right]. \hfill (4.4.16)$$

Substituting these into Eq. (4.2.1) yields the explicit form of a new metric. Notice that for these two solutions $\zeta \propto R^2$ near $R = 0$ which ensures that the space is regular with no cusp at $R = 0$.

4.5 Discussion

The disk models may be compared by exhibiting the physical quantities introduced in section 2 as functions of the “circumferential” radius $\tilde{R} = R e^{-\nu(R, 0)}$. The velocity curves for two families of the truncated $m = 0$ Kalnajs-Mestel disks are shown in Fig. 4.5.1. The Schwarzschild disks with $N = 1/2$ (full lines), and the disks with $N = 1/5$ (dashed lines). The velocities (measured in units of the velocity of light) are determined by Eq. (4.2.11); $\nu$ is given by Eq. (4.3.24). Here $b_1 = a$ and $b_2 = b$. Note that with $a$ decreasing (i.e. with the upper end of the rod approaching $R = z = 0$), the velocities increase close to $R = 0$. For $N \leq 1/2$, $a$ may approach zero and the value of the maximum velocity is given by

$$V_{\text{max}} = \left[ N/ (1 - N) \right]^{1/2}, \quad N \leq \frac{1}{2}.$$  

In Fig. 4.5.1 the parameter $a$ is fixed, but the effect of increasing $b$, i.e. of increasing the length of the rod or, correspondingly, the mass $M$ (cf. Eq. (4.3.24)) at fixed $N$ is presented. All the velocity curves rapidly increase at small $R$ but with $b$ increasing they remain flat over larger and larger intervals of $R$. By sending the lower end of the rod to infinity ($b \rightarrow \infty$, $N$ fixed), one

---

2 Bicak, Lynden-Bell & Katz [11] used this terminology since Eq. (4.1.1) and (4.4.14), with $r_3$ being replaced by a standard radial coordinate, give Curzon’s vacuum solution of Einstein’s equations (Kramer et. al. 1980) [63].

3 The physical circumference of the circle $R = \text{const}$ is given by $2\pi \tilde{R}$.
Figure 4.5.1: the velocity curves for two families of the truncated $m = 0$ relativistic Kalnajs-Mestel disks: $N = 1/2$ (full lines) and $N = 1/5$ (dashed lines). The numbers on the individual curves denote the values of the dimensionless parameter $b/a$. As this parameter increases, the velocity curves become flat. When $N = 1/2$, the velocities approach the velocity of light, when $N = 1/5$, one half of this velocity.

arrives at the infinite mass, constant velocity disks of Lynden-Bell & Pineault (1978) [79,80]. Note that the curves in Fig. 4.5.1 approach the maximal velocities according to the formulae: $V_{\text{max}} = 1$ for $N = 1/2$ (full lines), while $V_{\text{max}} = 1/2$ for $N = 1/5$ (dashed lines).

The velocity curves, the curves of the specific angular momentum, the surface mass-energy density, and the surface rest mass density in the dimensionless form $V$, $h/M$, $M\varepsilon$ and $M\sigma$ are given in Fig. 4.5.2, Fig. 4.5.3 and Fig. 4.5.4 as functions of the dimensionless radius $\tilde{R}/M$. These curves were calculated starting from Eqs. (4.2.12), (4.2.14), (4.2.15) and (4.2.16), with functions $\nu$ and $\zeta$ given in Sections 3 and 4. Fig. 4.5.2 corresponds to the curves for the relativistic Kuzmin-Curzon disks with $n = 1$, described by metric functions Eq. (4.4.15) and (4.4.16). Fig. 4.5.3 corresponds to the same quantities for the truncated Kalnajs-Mestel disks with $m = 1$, starting out from Eqs. (4.3.28) and (C.21). In Fig. 4.5.4 are illustrated the properties of the (un-truncated) Isochrone disks. In all the plots the total mass $M$ of the disks is fixed but several curves are plotted for different values of the parameters $b$ (for the Kuzmin-Toomre disks), $b_1$ and $a$ (for the Kalnajs-Mestel disks). In all cases the disks become more relativistic the smaller these parameters are, i.e. the closer the distribution of mass along the negative $z$-axis is to the centre of the plane $z = 0$. In the plots of $V$ and $h$ the parameters decrease from bottom to the top of the data. For $\varepsilon$ and $\sigma$ this is the case only at small $\tilde{R}$; at larger $\tilde{R}$’s the curves begin to cross. As it is easy to see from Eqs. (4.2.6) and (4.2.7), the condition that the velocity of counter-rotating streams does not exceed the velocity of light reduces to the dominant energy condition, $-\tau_t^t > \tau_\phi^\phi$ (Hawking & Ellis (1973) [43]). The maximal relativistic velocities appear at larger $\tilde{R}/M \approx 3 - 4$ for the Kuzmin-Toomre disks with $n = 1$, i.e. at larger radii than for the Kalnajs-Mestel disks. In the highly relativistic disks, $V$ approaches the velocity of light as
Section 4.5: Discussion

Figure 4.5.2: The same quantities as in Fig. 4.5.1 for nine relativistic Kuzmin-Toomre disks of order $n = 1$ corresponding to $M/b = 0.30, 0.33, ..., 0.54$. In the highly relativistic disks, as $\tilde{R}$ decreases, $V$ approaches the velocity of light and the specific angular momenta start to increase, indicating a relativistic instability. At those radii where $V \to 1$, the surface densities, in particular the surface rest mass density, decrease rapidly with $\tilde{R}$.

$\tilde{R}$ decreases, and the specific angular momenta cease to decrease with decreasing $\tilde{R}$ but start increasing. The angular momenta of circular test-particle orbits in Schwarzschild spacetime also show such a rise for $r < 6M$ which is associated with their relativistic instability. This effect is observed in all three families of disks (Fig. 4.5.2b, Fig. 4.5.3b, Fig. 4.5.4b). For the Kuzmin-Toomre disks it is more pronounced than for the Kalnajs-Mestel disks. (Compare, for example, the Isochrone disk with $V_{\text{max}} \sim 0.8$ with the analogous Kuzmin-Toomre disk).

As the disks become more relativistic, the surface mass densities rise rapidly towards the centre. This effect is most distinct in the truncated Kalnajs-Mestel disks with $m = 1$ (Fig. 4.5.3c,d). Nevertheless, these disks have long tails of surface density falling as $\tilde{R}^{-3}$ at large $\tilde{R}$, as is derived from Eq. (4.3.13) with $m = 1$ At large $\tilde{R}$, $\varepsilon \approx \sigma$ go over to the classical density $\Sigma_0$. In fact, the most compact disks are the Kuzmin-Toomre disks with large $n$. In the limit $n \to \infty$ the classical Kuzmin-Toomre disks approach the circular Gaussian disk in which $\Sigma = \Sigma_0 \exp(-R^2/R_c^2)$ (Toomre (1963) [107]), where $R_c \sqrt{\ln 2}$ is the half-mass radius. An interesting property of the rest mass surface density may be observed in the highly relativistic disks. At those radii where the velocities of the streams approach the velocity of light, the rest mass density decreases rapidly with $\tilde{R}$, as appears in all three Fig. 4.5.2(d), Fig. 4.5.3(d) and Fig. 4.5.4(d). In extreme cases, such as in the Isochrone disk in Fig. 4.5.4(d), it even reaches a local minimum.

Finally, the redshifts Eq. (4.2.17) of these objects are largest in the centre. In the case of the nth-order Kuzmin-Toomre disks, Eq. (4.2.17) and Eq. (4.4.1) imply

$$1 + z_0 = \exp \left[ -\nu_{m+3/2} (0, 0) \right],$$
Figure 4.5.3: The same quantities as in Fig. 4.5.1 for eleven relativistic truncated Kalnajs-Mestel disks of order $m = 1$ with $b_2 = e^2b_1$, $b_1/M = 2.3, 2.1, \ldots, 0.3$. The angular momenta of the first ten disks keep decreasing with decreasing $\tilde{R}$. In the most relativistic disk the rest mass density decreases very rapidly with $\tilde{R}$ for those $\tilde{R}$ where $V$ approaches the velocity of light.

where

$$-\nu_{m+3/2}(0, 0) = \frac{M}{b} A_n, \quad A_n = \sum_{k=0}^{m} \frac{(2n-k)!}{2^{n-k} (n-k)!}. \tag{4.5.1}$$

In the first family,

$$A_0 = 1, \quad A_1 = 2, \quad A_2 = \frac{8}{3}, \quad A_3 = \frac{16}{5}.$$  

As mentioned above, with $n \to \infty$ these disks approach the circular Gaussian disk. In this case (cf. Evans & de Zeeuw 1992 [33]), one finds

$$-\nu_{\infty}(0, 0) = \frac{M}{R_c} \sqrt{\pi} = \pi \sqrt{M \Sigma_0},$$

where $M = \pi \Sigma_0$ and $\Sigma_0$ is the classical central density. However, these disks can not produce arbitrarily large central redshifts by decreasing $R_c$ because $V$ must not exceed 1. Indeed the maximum value of $R d\nu/dR$ is $0.52\sqrt{\pi} M/R_c$ which is attained at $R = 1.257 R_c$. Now by Eq. (4.2.11) this quantity can not exceed $1/2$ for $V \leq 1$. Thus the maximum central redshift, $z_c$ for a disk of this type is given by $1 + z_c = 2.58$.

The central redshifts can be arbitrarily large in the Kalnajs-Mestel disks with $m = 0$ when $b_1 \to 0$, as shown by Bicak, Lynden-Bell & Katz [11].

In the Isochrone disks (cf. Eq. (4.3.31))

$$1 + z_c = \exp\left(\frac{M}{2a}\right),$$
so the central redshift $z_c$ cannot be made arbitrarily large by putting $a \to 0$ as the dominant energy condition, where $V_{\text{max}} = 1$, occurs for $M/a = 2.92$ leading to the limit

$$1 + z_c \leq 4.30.$$  

\begin{itemize}
\item \textbullet \ \textbullet \ \textcircled{\textbullet} \ \textbullet \ \textbullet \ \textbullet
\end{itemize}

\section*{4.6 Conclusion}

Recent results in Newtonian potential theory of axisymmetric infinitesimally thin disks have been used to construct infinite sequences of new solutions of Einstein’s equations which describe static, counter-rotating disks. The disks are infinite but, except for exceptional cases such as the un-truncated Kalnajs-Mestel disks with $m = 1$, their mass is finite. Relativistic generalisations of all Kuzmin-Toomre disks and Kalnajs-Mestel disks have been constructed. Einstein’s equations can be integrated starting from the corresponding classical potentials and the resulting metrics, though in general complicated, can all be expressed explicitly in closed forms.

Although the disks become Newtonian at large distances, in their central regions interesting relativistic features arise such as velocities close to the velocity of light, large redshifts, and
peculiar behaviour of the surface rest mass density. Some of these solutions correspond to non-spherical sources which lead to arbitrarily large redshifts. The disks discussed satisfy reasonable physical requirements such as a dominant (positive) energy condition, and the spacetimes they produce are asymptotically flat.

Besides the relativistic instability exhibited in the behaviour of the specific angular momentum in the most relativistic disks, the disks described in this chapter are unstable with respect to the formation of rings since no radial pressure is available to prevent the collapse. In spite of the recent observations of classical stellar counter-rotating disks mentioned in section 1, the very existence of relativistic Counter-rotating disks needs yet to be clarified. In the following chapter, these drawbacks will be addressed by constructing massive rotating disks with partial pressure support.
In this appendix, the generating function for the potentials $\nu_m$ of the mth-order Kalnajs-Mestel disks is derived explicitly. In the main text this generating function was defined as

$$J_1(u_a, u_b; v_1, v_2) = \frac{u_1 u_2 du_1 du_2}{(u_1^2 - v_1 u_1 - 1) (u_2^2 - v_2 u_2 - 1) (u_1 u_2 + 1)^2}. \quad (C.1)$$

Defining

$$v_i = w_i - \frac{1}{w_i}, \quad (i = 1, 2), \quad (C.2)$$

$J_1$ can be cast into the form:

$$J_1 = \int_{u_a}^{u_b} dQ \frac{u_2 du_2}{(u_2 - w_2) (u_2 + \frac{1}{w_2})}, \quad (C.3)$$

where

$$Q = \int_{u_a}^{u_b} \frac{u_1 du_1}{(u_1 - w_1) \left( u_1 + \frac{1}{w_1} \right) \left( u_1 + \frac{1}{u_2} \right)}. \quad (C.4)$$

The last integral can be calculated via decomposition into partial fractions, giving

$$Q = [Q_1 + Q_2]_{u_a}^{u_b},$$

where

$$Q_1 = \frac{1}{w_1^2 + 1} \left[ \frac{w_1^2}{(w_1 + \frac{1}{w_2})} \ln |u_1 - w_1| + \frac{1}{(u_2 + \frac{1}{w_1})} \ln |u_1 + \frac{1}{w_1}| \right],$$

and

$$Q_2 = \frac{u_2}{(u_2 + \frac{1}{w_1}) (u_2 - w_1)} \ln |u_1 + \frac{1}{u_2}|. \quad (C.5)$$

After substituting for $Q$ into Eq. (C.3) and performing minor manipulations, $J_1$ becomes the sum of three integrals

$$J_1 = R_1 + R_2 + S, \quad (C.6)$$
where, given Eq. (C.2),

\[
R(w_1, w_2) = R_1 = \left[ \ln |u_1 - w_1| \int_{u_a}^{u_b} \frac{u_2 \, du_2}{(u_2^2 + 1) \left( u_2 + \frac{1}{w_1} \right) \left( u_2^2 - v_2 u_2 - 1 \right)} \right]_{u_a}^{u_b},
\]

(C.7)

and

\[
R_2 = R \left( \frac{1}{w_1}, w_2 \right),
\]

(C.8)

Thus

\[
S(v_1, v_2) = \left[ \int_{u_a}^{u_b} \frac{dQ_2}{du_2} \frac{u_2 \, du_2}{(u_2^2 - v_2 u_2 - 1)} \right]_{u_a}^{u_b},
\]

(C.9)

with \(Q_2\) given by Eq. (C.5). Differentiating the log term in \(Q_2\) and then integrating by parts in the last integral gives

\[
S(v_1, v_2) = -T + \left[ \frac{u_2^2 \ln |u_1 + \frac{1}{u_2}|}{(u_2^2 - v_2 u_2 - 1) (u_2^2 - v_2 u_2 - 1)} \right]_{u_a}^{u_b, u_b} - \left[ \int_{u_a}^{u_b} \frac{u_2 \ln |u_1 + \frac{1}{u_2}|}{u_2^2 - v_2 u_2 - 1} \frac{u_2 \, du_2}{(u_2^2 - v_2 u_2 - 1)} \right]_{u_a}^{u_b},
\]

(C.10)

where this last integral is \(S(v_2, v_1)\) by comparison with Eq. (C.9). \(T\) is

\[
T = \left[ \int_{u_a}^{u_b} \frac{u_2 \, du_2}{(u_2^2 - v_1 u_2 - 1) (u_2^2 - v_2 u_2 - 1) (u_1 u_2 + 1)} \right]_{u_a}^{u_b}.
\]

(C.11)

Note that \(T\) is symmetrical in \(v_1\) and \(v_2\). From the symmetry of \(J_1\) with respect to variables \(v_1, v_2\) it is clear that

\[
J_1(u_a, u_b; v_1, v_2) = J_1(u_a, u_b; v_2, v_1).
\]

(C.12)

Thus

\[
J_1(v_1, v_2) = \frac{1}{2} \left[ J_1(v_1, v_2) + J_1(v_2, v_1) \right],
\]

so that

\[
2J_1(v_1, v_2) = R(w_1, w_2) + R(w_2, w_1) + R \left( -\frac{1}{w_1}, w_2 \right) + R \left( -\frac{1}{w_2}, w_1 \right) + S(v_1, v_2) + S(v_2, v_1).
\]

(C.13)

From Eq. (C.10) the last two terms are simply given by

\[
S(v_1, v_2) + S(v_2, v_1) = -T + \left[ \frac{u_2^2 \ln |u_1 + \frac{1}{u_2}|}{(u_2^2 - v_1 u_2 - 1) (u_2^2 - v_2 u_2 - 1)} \right]_{u_a}^{u_b, u_b}.
\]

(C.14)

The symmetrisation avoids the last integral in Eq. (C.10) which cannot be expressed as a finite combination of elementary functions.

The function \(J_1(u_a, u_b; v_1, v_2)\) is therefore given by Eq. (C.13), where \(R_1\) and \(R_2\) are given by Eq. (C.7) and (C.8) and \(S(v_1, v_2) + S(v_2, v_1)\) by Eq. (C.14). Evaluating Eq. (C.7) and (C.11) which are readily calculated by decomposition into partial fractions, leads, after some algebra, to the final form for the primitive function \(J_1\):

\[
J_1(u_a, u_b; w_1, w_2) = \frac{1}{2} \left[ \tilde{J}(u_1, u_2; w_1, w_2) + \tilde{J}(u_1, u_2; w_2, w_1) + j(u_1, u_2; w_1, w_2) \right]_{u_a}^{u_b, u_b},
\]
where
\[ j = \frac{u_2^2 \ln \left( \frac{(u_1 u_2 + 1)^2}{u_2} \right)}{(u_2^2 - v_1 u_2 - 1)(u_2^2 - v_2 u_2 - 1)}, \quad (C.15) \]

and \( \tilde{J} \) is given by
\[
\tilde{J} (u_1, u_2; w_1, w_2) = \\
\frac{w_1^2 w_2^2 \ln |u_1 - w_1|}{(w_1^2 + 1)(w_2^2 + 1)} \left[ -\frac{(w_1^2 + 1)(w_2^2 + 1) \ln |u_2 + \frac{1}{w_2}|}{(w_1 w_2 + 1)^2 (w_1 - w_2)^2} + \frac{\ln |u_2 - w_2|}{(w_1 w_2 + 1)^2} + \frac{\ln |u_2 + \frac{1}{w_2}|}{(w_1 - w_2)^2} \right] \\
+ \frac{w_1 w_2}{(w_1^2 + 1)(w_1 - w_2)(w_1 w_2 + 1)} \left[ \frac{w_2^2 \ln |u_2 + \frac{1}{w_2}|}{u_2 - w_1} + \frac{\ln |u_1 - w_1|}{u_2 + \frac{1}{w_1}} + \frac{\ln |u_2 - w_2|}{u_1 - \frac{1}{w_1}} - \frac{\ln |u_1 - \frac{1}{w_1}|}{u_1 + \frac{1}{w_1}} \right]. \quad (C.16)
\]

This expression is the generating function for the second constituent \( \zeta \) required for the complete relativistic metric of any member of the Kalnajs-Mestel family.

Note that in terms of the \( w_i \)'s the recursion relation Eq. (4.3.21) reads
\[
\mathcal{J}_m = \frac{1}{[(m - 1)!]^2} \left[ \frac{w_1^2 w_2^2}{(w_1^2 + 1)(w_2^2 + 1)} \frac{\partial^2}{\partial w_1 \partial w_2} \right]^{m-1} \mathcal{J}_2. \quad (C.17)
\]

The explicit form of \( \mathcal{J}_1 \) for the untruncated Kalnajs-Mestel disks, follows from Eq. (C.15) and (C.16) after putting \( u_a = u \) and \( u_b \to \infty \):
\[
\mathcal{J}_1 (u_a = u, u_b = \infty; w_1, w_2) = 2w_1 w_2 \left\{ \tilde{J}_1 (u; w_1, w_2) + \tilde{J}_1 (u; w_2, w_1) \right\} + \frac{u^2}{(u - w_1)(u - w_2)(1 + u w_1)(1 + u w_2)} \ln |u + \frac{1}{u}|, \quad (C.18)
\]
\[
+ \frac{w_1 w_2 (1 + w_1 w_2)^{-2}}{(1 + u_1^2)(1 + u_2^2)} \left[ \ln |u - w_1| \ln |u - w_2| + \ln |u + \frac{1}{w_1}| \ln |u + \frac{1}{w_2}| \right].
\]

where
\[
\tilde{J}_1 (u; w_1, w_2) = \frac{w_1}{(w_2 - w_1)(1 + w_1 w_2)(1 + w_2)} \left[ \frac{\ln |u - w_1|}{1 + u w_1} + \frac{w_1 \ln |u + \frac{1}{w_1}|}{u - w_1} \right] + \frac{w_1 w_2}{(w_2 - w_1)} \ln |u + \frac{1}{w_1}| \left[ \frac{\ln |u - w_2|}{(1 + u_1^2)(1 + u_2^2)} - \frac{\ln |u - w_1|}{(1 + w_1 w_2)^2} \right]. \quad (C.19)
\]

This generating function gives \( \zeta_m \) for all un-truncated Kalnajs-Mestel disks which, together with \( \nu_m \) given by Eq. (4.3.10), then determines the metrics completely.

Another asymptotic limit is readily derived from Eq. (C.15) and (C.16) by setting \( w_1 \) and \( w_2 \) to \( w \). This yields after some algebra
\[
\mathcal{J}_1^* = \mathcal{J}_1 (u_a = u_b; w_1 = w, w_2 = w) = \mathcal{K} + \left[ \mathcal{L} + \frac{4 w^4}{(1 + w^2)^4} \ln \left| \frac{u_a + \frac{1}{w}}{u_b + \frac{1}{w}} \right| \right] \ln \left| \frac{u_b - w}{u_a - w} \right|
\]
\[
+ \left[ \mathcal{M} + \frac{2w^4}{(1+w^2)^4} \ln \left| \frac{u_a + \frac{1}{u}}{u_b} \right| \ln \left| \frac{u_a + \frac{1}{u}}{u_b} \right| + \frac{2w^2u_a^2 (1+u_aw)^{-2}}{(u_a - w)^2} \ln \left| \frac{1+u_a^2}{1+u_a u_b} \right| \right] \\
+ \frac{2w^2u^2_0 (1+u_aw)^{-2}}{(u_b - w)^2} \ln \left| \frac{1+u_b^2}{1+u_a u_b} \right| + \frac{2w^4}{(1+w^4)^4} \left[ \ln \left| \frac{u_a - w}{u_b} \right| \right]^2,
\]
where
\[
\mathcal{K} = \frac{2w^4(u_a - u_b)^2}{(1+w^2)^2(u_a - w)(u_b - w)(1+u_aw)(1+u_b w)},
\]
\[
\mathcal{L} = \frac{2w^3(u_a - u_b) \left[ -2 - w(u_a + u_b) + w^3(u_a + u_b) + 2w^4u_a u_b \right]}{(1+w^4)^3(a + u_aw)(1+u_b w)^2},
\]
\[
\mathcal{M} = \frac{2w^3(u_b - u_a) \left[ 2w^4 + w(u_a + u_b) - w^3(u_a + u_b) - 2u_a u_b \right]}{(1+w^4)^3(u_a - w)^2(u_b - w)^2}.
\]
This expression characterises \( \zeta_1 \) for the \( m = 1 \) truncated Kalnajs-Mestel disk.
5

Dynamics of rotating super-massive disks

5.1 Introduction

The general-relativistic theory of rapidly rotating objects is of great intrinsic interest and has potentially important applications in astrophysics. Following Einstein’s early work on relativistic collisionless spherical shells, Fackerel (1968) [34], Ipser (1969) [51] and Ipser & Thornes (1968) [52] have developed the general theory of relativistic collisionless spherical equilibria. The stability of such bodies has been further expanded by Katz and Horowitz (1975) [60]. Here the complete dynamics of flat disk equilibria is developed and illustrated, extending the results of the previous chapter to differentially rotating configurations with partial pressure support. A general inversion method for the corresponding distribution functions is presented, yielding a coherent model for the stellar dynamics of these disks. Differentially rotating flat disks in which centrifugal force almost balances gravity can also give rise to relatively long lived configurations with large binding energies. Such objects may therefore correspond to possible models for the latest stage of the collapse of a proto-quasar. Analytic calculations of their structure have been carried into the post Newtonian régime by Chandrasekhar. Strongly relativistic bodies have been studied numerically following the pioneering work on uniformly rotating cold disks by Bardeen & Wagoner [6]. An analytic vacuum solution to the Einstein equations, the Kerr metric [61], which is asymptotically flat and has the general properties expected of an exterior metric of a rotating object, has been known for thirty years, but attempts to fit an interior solution to its exterior metric have been unsatisfactory. A method for generating families of self-gravitating rapidly rotating disks purely by geometrical methods is presented. Known vacuum solutions of Einstein’s equations are used to produce the corresponding relativistic disks. One family includes interior solutions for the Kerr metric.

5.1.1 Background

In 1919 Weyl [111] and Levi-Civita gave a method for finding all solutions of the axially symmetric Einstein field equations after imposing the simplifying conditions that they describe a static vacuum field. In the previous chapter, complete solutions corresponding to counter rotating pressure-less axisymmetric disks were constructed by matching exterior solutions of the Weyl- Levi-Civita type on each side of the disk. This method will here be generalised, allowing
true rotation and radial pressure support for solutions describing self-gravitating disks. These solutions provide physical sources (interior solutions) for stationary axisymmetric gravitational fields.

### 5.1.2 Pressure Support via Curvature

In Newtonian theory, once a solution for the vacuum field of Laplace’s equation is known, it is straightforward to build solutions to Poisson’s equation for disks by using the method of mirror images. In general relativity, the overall picture is similar. In fact, it was pointed out by Morgan & Morgan (1969) [90] that for pressureless counter rotating disks in Weyl’s co-ordinates, the Einstein equation reduces essentially to solving Laplace’s equation (cf. previous chapter). When pressure is present, no global set of Weyl co-ordinates exists. The condition for the existence of such a set of co-ordinates is that $T_{RR} + T_{zz} = 0$. On the disk itself, this is satisfied only if the radial eigenvalue of the pressure tensor, $p_R$, vanishes ($T_{zz}$ vanishes provided the disk is thin). When $p_R$ is non-zero, vacuum co-ordinates of Weyl’s type still apply both above and below the disk, but as two separate co-ordinate patches. Consider global axisymmetric co-ordinates $(R', \phi', z')$ with $z' = 0$ on the disk itself. For $z' > 0$, Weyl’s co-ordinates $R, \phi, z$ may be used. In terms of these variables, the upper boundary of the disk $z' = 0^+$ is mapped onto a given surface $z = f(R)$. On the lower patch, by symmetry, the disk will be located on the surface $z = -f(R)$. Thus in Weyl co-ordinates, the points $z = f(R)$ on the upper surface of the disk must be identified with the points $z = -f(R)$ on the lower surface. The intrinsic curvature of the two surfaces that are identified match by symmetry. The jump in the extrinsic curvature gives the surface distribution of stress-energy on the disk. A given metric of Weyl’s form is a solution of the empty space Einstein equations, and does not give a complete specification of its sources (for example, a Schwarzschild metric of any given mass can be generated by a static spherical shell of any radius). With this method, all physical properties of the source are entirely characterised once it is specified that the source lie on the surface $z = \pm f(R)$ and the corresponding extrinsic curvature (i.e. its relative layout within the embedding vacuum spacetime) is known. The freedom of choice of $\nu$ in the Weyl metric corresponds to the freedom to choose the density of the disk as a function of radius. The supplementary degree of freedom involved in choosing the surface of section $z = f(R)$ corresponds classically to the choice of the radial pressure profile.

### 5.1.3 Rotating Disks

For rotating disks, the dragging of inertial frames induces strongly non-linear fields outside the disk which prohibit the construction of a vacuum solution by superposition of the line densities of fictitious sources. Nevertheless, in practice quite a few vacuum stationary solutions have been given in the literature, the most famous being the Kerr metric. Hoenselaers, Kinnersley & Xanthopoulos (1979) [46,47](HKX hereafter) have also given a discrete method of generating rotating solutions from known static Weyl solutions. The disks are derived by taking a cut through such a field above all singularities or sources and identifying it with a symmetrical cut through a reflection of the source field. This method applies directly to the non-linear fields such as those generated by a rotating metric of the Weyl-Papapetrou type describing stationary axisymmetric vacuum solutions. The analogy with electromagnetism is then to consider the field associated with a known azimuthal vector potential $A_{\phi}$, the analog of $g_{\mu\phi}/g_{tt}$. The jump in the tangential component of the corresponding $B$-field gives the electric current in the plane, just as the jump in the $t, \phi$ component of the extrinsic curvature gives the matter current.

The procedure is then the following: for a given surface embedded in curved space-time and a given field outside the disk, the jump in the extrinsic curvature on each side of the mirror
images of that surface is determined in order to specify the matter distribution within the disk. The formal relationship between extrinsic curvatures and surface energy tensor per unit area was introduced by Israel and is described in details by Misner Thorne Wheeler (1973) p 552.\cite{89} Plane surfaces and non-rotating vacuum fields (\textit{i.e.} the direct counterpart of the classical case) were investigated in the previous chapter and led to pressureless counter-rotating disks. Any curved surface will therefore produce disks with some radial pressure support. Any vacuum metric with $g_{t\phi} \neq 0$ outside the disk will induce rotating disks.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure5.1.1.png}
\caption{Symbolic representation of a section $z = \pm f(R)$ through the $g_{t\phi}$ field representing the upper and lower surface of the disk embedded in two Weyl-Papapetrou fields. The isocontours of $\omega(R,z)$ correspond to a measure of the amount of rotation in the disk.}
\end{figure}

In section 2, the construction scheme for rotating disks with pressure support is presented. In section 3, the properties of these disks are analysed, and a reasonable cut, $z = f(R)$, is suggested. This method is then first applied to warm counter rotating disks in section 4; the dominant energy condition given for Curzon disks by Chamorow \textit{et. al.}(1987)\cite{17}, and Lemos’ solution for the self - similar Mestel disk with pressure support are recovered. In section 5, the Kerr disk is studied and the method for producing general HKX disks is sketched.
5.2 Derivation

The jump in extrinsic curvature of a given profile is calculated and related to the stress energy tensor of the matter distribution in the corresponding self gravitating relativistic disk.

5.2.1 The Extrinsic Curvature

The line element corresponding to an axisymmetric stationary vacuum gravitational field is given by the Weyl-Papapetrou (WP) metric

\[ ds^2 = -e^{2\nu}(dt - \omega d\phi)^2 + e^{2\zeta-2\nu}(dR^2 + dz^2) + R^2e^{-2\nu}d\phi^2, \]  

(5.2.1)

where \((R, \phi, z)\) are standard cylindrical co-ordinates, and \(\nu, \zeta\) and \(\omega\) are functions of \((R, z)\) only. Note that this form is explicitly symmetric under simultaneous change of \(\phi\) and \(t\). The vacuum field induces a natural metric on a 3-space-like \(z = f(R)\) surface

\[ d\sigma^2 = g_{\alpha\beta} \partial x^\alpha \partial x^\beta dx^a dx^b \equiv \gamma_{ab} dx^a dx^b, \]

(5.2.2)

where \(\{x^a\}\) are the co-ordinates on the embedded hypersurface \(z = f(R)\), \(\{x^a\} = (t, R, \phi, z)\) are Weyl’s co-ordinates for the embedding spacetime. \(\gamma_{ab}\) stands for the embedded metric, and \(h^a_a\) is given by

\[ h^a_a = \frac{\partial x^a}{\partial x^a} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & f' \\ 0 & 0 & 1 & 0 \end{bmatrix}, \]

(5.2.3)

where hereafter a prime, \( (\cdot)'\), represents the derivative with respect to \(R\). The line element is then

\[ d\sigma^2 = -e^{2\nu}(dt - \omega d\phi)^2 + R^2e^{-2\nu}d\phi^2 + e^{2\zeta-2\nu}(1 + f'^2) dR^2. \]

(5.2.4)

Let the upward pointing normal \(N_\mu\) to the surface \(z = f(R)\) be

\[ N_\mu = \frac{\partial}{\partial x^\mu} [z - f(R)] = (0, -f', 0, 1). \]

(5.2.5)

The normalised normal is therefore

\[ n_\mu = \frac{N_\mu}{\sqrt{N_\alpha N^\alpha}} = \frac{N_\mu}{\sqrt{g^{\alpha\beta}N_\alpha N^\beta}} = \frac{N_\mu}{\sqrt{1 + f'^2}} e^{\zeta-\nu}. \]

(5.2.6)

The extrinsic curvature tensor \(K\) is given in its covariant form by

\[ K_{ab} = n_\mu \left( \partial_a h^\mu_b + \Gamma^\mu_{ab} h^\alpha_a h^\beta_b \right), \]

(5.2.7)

where \(\partial_a h^\mu_b = \partial h^\mu_b / \partial x^a\). Eq. (5.2.6) and (5.2.3) together with the Christoffels given in the Appendix lead to the computation of \(K\), the extrinsic curvature of that surface embedded in geometric units \(G = c = 1\) are used throughout this chapter.
the Weyl Papapetrou metric.

\[
K_{\mu\nu} = -\frac{e^{\nu-\zeta}}{\sqrt{1 + f^2}} \left( \frac{\partial \nu}{\partial N} \right) e^{\lambda-2\zeta}, \\
K_{R\nu} = -\frac{e^{\nu-\zeta}}{\sqrt{1 + f^2}} \left( \frac{f''}{1 + f^2} + \frac{\partial (\nu - \zeta)}{\partial N} \right) (1 + f^2), \\
K_{\phi\nu} = -\frac{e^{\nu-\zeta}}{\sqrt{1 + f^2}} \left[ \frac{\partial \nu}{\partial N} \right] \omega e^{\lambda-2\zeta}, \\
K_{\phi\phi} = -\frac{e^{\nu-\zeta}}{\sqrt{1 + f^2}} \left[ \left( \frac{f'}{R} + \frac{\partial \nu}{\partial N} \right) R^2 e^{-2\zeta} + e^{\lambda-2\zeta} \omega^2 \frac{\partial \nu - \log(\omega)}{\partial N} \right],
\]

where the notation \(\partial / \partial N \equiv (\partial / \partial z - f \partial / \partial R)\) has been used.

### 5.2.2 The Stress Energy Tensor

The stress energy tensor per unit surface \(\tau^a_b\) is the integral of the stress energy tensor carried along the normal to the surface \(z = f(R)\). In a locally Minkowskian frame co-moving with the mean flow of the disk, the corresponding orthonormal tetrad is

\[
e_i(0) = e^{\nu}(1, 0, -\omega), \\
e_i(1) = e^{\zeta-\nu} \sqrt{1 + f^2} (0, 1, 0), \\
e_i(2) = Re^{-\nu} (0, 0, 1),
\]

so that

\[
ds^2 = \eta_{(a)(b)} \left( e_i^{(a)} dx^i \right) \left( e_j^{(b)} dx^j \right),
\]

with \(\eta_{(a)(b)} = \text{Diag}(-1, 1, 1)\). In that frame,

\[
[\tau^{(a)(b)}]_\nu = \begin{bmatrix}
\varepsilon & 0 & 0 \\
0 & p_R & 0 \\
0 & 0 & p_\phi
\end{bmatrix}.
\]

After a Lorentz transformation to a more general frame in which the flow is rotating with relative velocity \(V\) in the \(\phi\) direction, this stress becomes

\[
\tau^{(a)(b)} = \frac{1}{1 - \varepsilon^2} \begin{bmatrix}
\varepsilon + p_\phi V^2 & 0 & (p_\phi + \varepsilon) V \\
0 & (1 - V^2) p_R & 0 \\
(p_\phi + \varepsilon) V & 0 & p_\phi + \varepsilon V^2
\end{bmatrix}.
\]

### 5.2.3 The Discontinuity Equations

Israel (1964) [64] has shown that Einstein’s equations integrated through a given surface of discontinuity can be re-written in terms of the jump in extrinsic curvature, namely

\[
\tau^a_b = \frac{1}{8\pi} \left[ K^a_b - K^a_a \delta^a_b \right] \equiv \mathcal{L}^a_b.
\]
where \( [ ]^+ \) stands for ( ) taken on \( z = f(R) \) minus ( ) taken on \( z = -f(R) \). \( \mathcal{L}_b^a \) is known as the Lanczos tensor.

In the tetrad frame (5.2.9), the Lanczos tensor given by Eqs. (5.2.13) and (5.2.8) reads

\[
\mathcal{L}^{(a)(b)} = \frac{e^{\nu - \zeta}}{4\pi \sqrt{1 + f'^2}} \begin{bmatrix}
\frac{f''}{1 + f'^2} + f' + \frac{\partial (2\nu - \zeta)}{\partial N} & 0 & \frac{\partial \omega}{\partial N} e^{2\nu} \\
0 & -\frac{f'}{R} & 0 \\
-\frac{\partial \omega}{\partial N} e^{2\nu} & 0 & -\frac{f''}{1 + f'^2} + \frac{\partial \zeta}{\partial N}
\end{bmatrix}.
\]

Identifying the stress energy tensor \( \tau^{(a)(b)} \) with the tetrad Lanczos tensor \( \mathcal{L}^{(a)(b)} \) according to Eq. (5.2.13), and solving for \( p_r, p_\phi, \varepsilon \) and \( V \) gives

\[
V = \frac{Re^{-2\nu}}{\partial \omega / \partial N} \left[ \left( \frac{f'}{R} + 2 \frac{\partial \nu}{\partial N} \right) - Q \right],
\]

\[
\varepsilon = \frac{e^{\nu - \zeta}}{4\pi \sqrt{1 + f'^2}} \left[ \frac{Q}{2} + \left( -\frac{f''}{1 + f'^2} + \frac{f'}{2R} + \frac{\partial (\nu - \zeta)}{\partial N} \right) \right],
\]

\[
p_\phi = \frac{e^{\nu - \zeta}}{4\pi \sqrt{1 + f'^2}} \left[ \frac{Q}{2} - \left( -\frac{f''}{1 + f'^2} + \frac{f'}{2R} + \frac{\partial (\nu - \zeta)}{\partial N} \right) \right],
\]

\[
p_r = \frac{e^{\nu - \zeta}}{4\pi \sqrt{1 + f'^2}} \left( -\frac{f'}{R} \right),
\]

where

\[
Q \equiv \sqrt{\left( \frac{f'}{R} + 2 \frac{\partial \nu}{\partial N} \right)^2 - e^{4\nu} \left( \frac{\partial \omega}{\partial N} \right)^2}.
\]

All quantities are to be evaluated along \( z = f(R) \). Eq. (5.2.15) gives the form of the most general solution to the relativistic rotating thin disk problem provided the expressions for \( \varepsilon, p_\phi, p_r \) are physically acceptable.

\[
\bullet \circ \bigcirc \bullet.
\]

### 5.3 Physical properties of the warm disks

Physical properties of interest for these disks are derived and related to the choice of profile \( z = f(R) \) compatible with their dynamical stability.

Defining the circumferential radius \( R_c \), proper radial length \( \tilde{R} \), and a synchronised proper time \( \tau_s \) by

\[
R_c = Re^{-\nu},
\]

\[
\tilde{R} = \int \sqrt{1 + f'^2} e^{\zeta - \nu} dR,
\]

\[
\tau_s = \int_{(\tau)} e^{\nu} \left( 1 - \omega \frac{d\phi}{dt} \right) dt,
\]

where the integral over \( dR \) is performed at \( z = f(R) \) and that over \( dt \) is performed along a given trajectory \( (T) \), the line element on the disk (5.2.4) reads

\[
d\sigma^2 = -d\tau_s^2 + d\tilde{R}^2 + R_c^2 d\phi^2.
\]

(5.3.2)
For circular flows ($\Phi_r = e^\nu [1 - \omega d\phi / dt] dt$), $V$ may be re-written as

$$V = R_c \frac{d\phi}{d\tau^*}.$$  \hspace{1cm} (5.3.3)

Equation (5.3.3) is inverted to yield the angular velocity of the flow as measured at infinity:

$$\Omega \equiv \frac{d\phi}{dt} = \frac{Ve^\nu / R_c}{1 + \omega Ve^\nu / R_c}.$$  \hspace{1cm} (5.3.4)

Similarly, the covariant specific angular momentum $h$ of a given particle reads in terms of these variables

$$h \equiv \frac{p_\phi}{\mu} = \gamma_{t,\phi} u^t + \gamma_{\phi,\phi} u^\phi = \frac{1}{\sqrt{1 - V^2}} \left( R_c V + \omega e^\nu \right),$$  \hspace{1cm} (5.3.5)

where $\mu$ is the rest mass of the particle. The covariant specific energy $\epsilon$ of that particle reads

$$\epsilon \equiv -\frac{p_t}{\mu} = -\gamma_{t,\phi} u^\phi - \gamma_{t,t} u^t = \frac{e^\nu}{\sqrt{1 - V^2}},$$  \hspace{1cm} (5.3.6)

which gives the central redshift, $1 + z_c = \exp(-\nu)$. The velocity of zero angular momentum observers (so called ZAMOs) follows from Eq. (5.3.5):

$$V_z = -\frac{\omega e^\nu}{R_c}.$$  \hspace{1cm} (5.3.7)

With these observers, the cuts $z = f(R)$ may be extended inside ergoregions where the dragging of inertial frames induces apparent superluminous motions as measured by locally static observers. The circumferential radius $R_z$ as measured by ZAMOs is Lorentz contracted with respect to $R_c$, becoming

$$R_z \equiv \sqrt{\bar{g}_{\phi\phi}} = R_c \left[ 1 - V_z^2 \right]^{1/2},$$  \hspace{1cm} (5.3.8)

while the velocity flow measured by ZAMOs is

$$V_{z/} = \frac{V - V_z}{1 - VV_z}.$$  \hspace{1cm} (5.3.9)

The total angular momentum of the disk $H$ follows from the asymptotic behaviour of the vacuum field $\omega$ at large radii, $\omega \to -2H/r$. Define the binding energy of the disk $\Delta E$ as the difference between the baryonic rest mass $M_0$ and the total mass-energy of the disk $M$ as measured from infinity given by $\nu \to -M/r$:

$$\Delta E = M_0 - M.$$  \hspace{1cm} (5.3.10)

$M$ is also given by the Tolmann formula, but corresponds, by construction, to the mass of the line source for the vacuum field. The baryonic rest mass can be expressed as

$$M_0 \equiv \int \Sigma \sqrt{-g} U^\mu dS_{\mu} = \int \frac{[1 - \omega \Omega]^{-1}}{\sqrt{1 - V^2}} \Sigma 2\pi R_c d\tilde{R},$$  \hspace{1cm} (5.3.11)

where the baryonic energy density $\Sigma$ can be related to the energy density $\epsilon$ through some yet unspecified equation of state or the knowledge of a distribution function. Possible equations of state are derived in the appendix and will be presented in sections 4 and 6 for concrete examples; a general method to derive distribution functions corresponding to a given flow is presented in section 5.
5.3.1 Constraints on the internal solutions

The choice of the profile $z = f(R)$ is open, provided that it leads to meaningful physical quantities. Indeed it must lead to solutions which satisfy $\varepsilon \geq 0$, $p_r \geq 0$, $p_\phi \geq 0$, and $V \leq 1$ while the dominant energy condition implies also that $\varepsilon \geq p_\phi$ and $\varepsilon \geq p_r$. These translate into the following constraints on $f$

\[
\begin{align*}
\varepsilon & \geq 0 \quad \Rightarrow \sqrt{N^2 - \mathcal{O}^2} + (N - 2Z) \geq 0; \\
0 \leq V \leq 1 & \Rightarrow N \geq \mathcal{O} \text{ and } \mathcal{O} \geq 0; \\
0 \leq p_\phi \leq \varepsilon & \Rightarrow 4Z \mathcal{N} \geq (\mathcal{O}^2 + 4Z^2) \text{ and } \mathcal{N} \geq 2Z; \\
0 \leq p_r \leq \varepsilon & \Rightarrow f' \leq 0 \text{ and } -2f'/R \leq \sqrt{N^2 - \mathcal{O}^2} + (N - 2Z); \\
\end{align*}
\]

(5.3.12)

given

\[
\mathcal{N} \equiv \frac{f'}{R} + 2\frac{\partial \nu}{\partial N}, \quad \mathcal{O} \equiv \frac{\epsilon^2}{R} \frac{\partial \omega}{\partial N}, \quad \text{and} \quad Z \equiv \left( \frac{\partial \zeta}{\partial N} - \frac{f''}{1 + f'^2} \right).
\]

(5.3.13)

The existence of a solution satisfying this set of constraints can be demonstrated as follows: in the limit of zero pressure and counter-rotation (i.e. $\mathcal{O} \to 0, N \to 2\partial \nu/\partial z$, and $z \to \partial \zeta/\partial z$), any cut $z = \text{Const} \equiv b \gg m$ satisfies Eqs. (5.3.12). By continuity, there exist solutions with proper rotation and partial pressure support. In practise, all solutions given in sections 4 and 5 satisfy the constraints (5.3.12). Note that in the limit of zero radial pressure, Eq. (5.3.12b) implies $\partial/\partial z [\omega/R + \exp 2\nu] \leq 0$.

5.3.2 Ansatz for the profile

In the following discussion, the section $z = f(R)$ is chosen so that the corresponding radial pressure gradient balances a given fraction of the gravitational field which would have occurred had there been no radial pressure within the disk. This choice ‘bootstraps’ calculations for all relevant physical quantities in terms of a single degree of freedom (i.e. this fraction $\eta$), rather than a complete functional. This gives

\[
\frac{\partial p_r}{\partial R} = -\frac{\eta}{1 - V^2} \frac{\partial \nu}{\partial R} \left[ \varepsilon + p_r + V^2 (p_\phi - p_r) \right].
\]

(5.3.14)

On the r.h.s. of Eq. (5.3.14), $p_r$ is put to zero and $V, p_\phi, \varepsilon$ are re-expressed in terms of $\zeta, \nu, \omega$ via Eqs. (5.2.15) with $f' = f'' = 0$ and $z = \text{Const}$. On the l.h.s. of Eq. (5.3.14), $p_r$ is chosen according to Eq. (5.2.15d). In practise, this ansatz for $f$ is a convenient way to investigate a parameter space which is likely to be stable with respect to ring formation as discussed in the next subsection. In principle, a cut, $f$, could be chosen so as to provide a closed bounded symmetrical surface containing all fictitious sources. Indeed, by symmetry no flux then crosses the $z = 0$ plane beyond the cut, and the energy distribution is therefore bound to the edge of that surface: this corresponds to a finite pressure supported disk. It turns out that in general this choice is not compatible with all the constraints enumerated in the previous section. More specifically, the positivity of $p_\phi$ fails for all finite disk models constructed. The ansatz described above gives for instance an upper bound on the height of that surface when assuming that all the support is provided by radial pressure. This height is in turn not compatible with positive azimuthal pressure at all radii.
5.3.3 Stability

To what extent are the equilibria studied in the previous sections likely to be stable under the action of disturbances? The basic instabilities can be categorised as follows:

- dynamical instabilities, which are intrinsic to the dynamical parameters of the disks, grow on an orbital time scale, and typically have drastic effects on the structure of the system.

- secular instabilities, which arise owing to dissipative mechanisms such as viscosity or gravitational radiation, grow on a time scale which depends on the strength of the dissipative mechanism involved, and slowly drive the system along a sequence of dynamical equilibria.

Amongst dynamical instabilities, kinematical instabilities correspond to the instability of circular orbits to small perturbations, and collective instabilities arise from the formation of growing waves triggered by the self-gravity of the disk. Rings, for instance, will be generated spontaneously in the disk if the local radial pressure is insufficient to counteract the self-gravity of small density enhancements. Even dynamically stable non-axisymmetric modes may drive the system away from its equilibrium by radiation of gravitational waves which will slowly remove angular momentum from the disk. For gaseous disks, viscosity and photon pressure will affect the equilibria. Radiative emission may disrupt or broaden the disk if the radiation pressure exceeds the Eddington limit. The energy loss by viscosity may induce a radial flow in the disk. However, the disks discussed in this chapter have anisotropic pressures inappropriate for gaseous disks and the accurate description of the latter two processes requires some prescription for the dissipative processes in the gas. The scope of this section is therefore restricted to a simple analysis of the dynamical instabilities.

Turning briefly to the corresponding Newtonian problem, Toomre (1963) [107] gave the local criterion for radial collective instability of stellar disks,

$$\sigma_R \geq \frac{3.36 G \Sigma_0}{\kappa}, \quad (5.3.15)$$

where $\Sigma_0$ is the local surface density, $\sigma_R^2$ the radial velocity dispersion, and $\kappa$ the epicyclic frequency of the unperturbed stars. This criterion is derived from the first critical growing mode of the dispersion relation for radial waves. The corresponding critical wavelength is Jeans length $\sim 2\sigma_R^2/G \Sigma_0$.

For the relativistic disks described in this chapter, spacetime is locally flat, which suggests a direct translation of Eq. (5.3.15) term by term. The constraints that stability against ring formation places on these models will then be addressed at least qualitatively via the Newtonian approach. The proper relativistic analysis is left to further investigation. The relativistic surface density generalising $\Sigma_0$ is taken to be the co-moving energy density $\varepsilon$ given by Eq. (5.2.15b). The radial velocity dispersion $\sigma_R^2$ is approximated by $p_R/\varepsilon$. The epicyclic frequency is calculated in the appendix. Putting Eqs. (D.2.4) and (D.2.5) into Eq. (5.3.15) give another constraint on $f$ for local radial stability of these disks. Note that the kinematical stability of circular orbits follows from Eq. (D.2.3) by requiring $\kappa^2$ to be positive.
5.4 Application: warm counter rotating disks

For simplicity, warm counter-rotating solutions are presented first, avoiding the non-linearities induced by the dragging of inertial frames. These solutions generalise those of chapter III, while implementing partial pressure support within the disk. Formally this is achieved by putting $\omega(R, z)$ identically to zero in Eq. (5.2.1); the metric for the axisymmetric static vacuum solutions given by Weyl is then recovered:

$$ds^2 = -e^{2\nu} dt^2 + e^{2\zeta} (-dR^2 + dz^2) + R^2 e^{-2\nu} d\phi^2,$$

where the functions $\zeta(R, z)$ and $\nu(R, z)$ are generally of the form (cf. Eq. (4.1.3), Eq. (4.3.14))

$$\nu = -\int \frac{W(b)}{\sqrt{R^2 + (|z| + b)^2}},$$

$$\zeta = \int \int W(b_1) W(b_2) Z(b_1, b_2) db_1 db_2,$$

with $Z$ given by Eq. (4.3.15). When the line density of fictitious sources is of the form $W(b) \propto b^{-m}$ (Kuzmin-Toomre), or $W(b) \propto \delta^{(m)}(b)$ (Kalnajs-Mestel), the corresponding functions $\zeta, \nu$ have explicitly been given in the previous chapter. For the metric of Eq. (5.4.1), the Lanczos tensor (5.2.14) reads

$$\begin{bmatrix}
L^{(t)(t)} \\
L^{(R)(R)} \\
L^{(\phi)(\phi)}
\end{bmatrix} = \frac{e^{\nu-\zeta}}{4\pi \sqrt{1 + f'^2}} \begin{bmatrix}
\frac{f''}{1 + f'^2} + \frac{\nu'}{f' R} + \frac{\partial (2\nu-\zeta)}{\partial N} \\
-\frac{f'}{R} \\
\frac{f''}{1 + f'^2} + \frac{\partial \zeta}{\partial N}
\end{bmatrix}.$$  

(5.4.3)

Consider again counter-rotating disks made of two equal streams of stars circulating in opposite directions around the disk centre. The stress energy tensor is then the sum of the stress energy tensor of each stream:

$$\tau^{(a)(b)} = \begin{bmatrix}
\varepsilon & 0 & 0 \\
0 & p_r & 0 \\
0 & 0 & p_\phi
\end{bmatrix} = \frac{2}{1 - V_0^2} \begin{bmatrix}
\varepsilon_0 + p_0 V_0^2 & 0 & 0 \\
0 & (1 - V_0^2) p_0 & 0 \\
0 & 0 & p_0 + \varepsilon_0 V_0^2
\end{bmatrix},$$  

(5.4.4)

the pressure $p_0$ in each stream being chosen to be isotropic in the plane. Subscript $(\cdot)_0$ represents quantities measured for one stream. Expressions for $\varepsilon, p_\phi$, and $p_r$ follow from identifying Eqs. (5.4.4) and (5.4.3) according to Eq. (5.2.13). Solving for $\varepsilon_0, V_0$, and $p_0$ in Eq. (5.4.4), given Eq. (5.4.3) and (5.2.13) gives

$$\varepsilon_0 = \frac{e^{\nu-\zeta}}{4\pi \sqrt{1 + f'^2}} \left[ \frac{f''}{1 + f'^2} + \frac{\partial (\nu-\zeta)}{\partial N} \right],$$  

(5.4.5a)

$$p_0 = \frac{e^{\nu-\zeta}}{8\pi \sqrt{1 + f'^2}} \left[ -\frac{f'}{R} \right],$$  

(5.4.5b)

$$V_0^2 = \left( \frac{\partial \zeta}{\partial N} - \frac{f''}{1 + f'^2} + \frac{f'}{R} \right) \left( \frac{f''}{1 + f'^2} + \frac{\partial (2\nu-\zeta)}{\partial N} \right)^{-1}.$$  

(5.4.5c)

The above solution provides the most general counter-rotating disk model with pressure support. Indeed, any physical static disk will be characterized entirely by its vacuum field $\nu$ and its radial
pressure profile $p_n$, which in turn defines $W(b)$ and $f(R)$ uniquely according to Eqs. (5.4.5b) and (4.3.15). The other properties of the disk are then readily derived.

The angular frequency and angular momentum of these disks follow from Eq. (5.3.5), (5.3.4) on putting $\omega$ to zero and $V$ to $V_0$. The epicyclic frequency $\kappa$ given by Eq. (D.2.4) may be recast as

$$\kappa^2 = \frac{e^{\nu} dh^2}{R^2 \partial_{\nu} \partial_{\nu}},$$

which relates closely to the classical expression $\kappa^2 = R^{-3} dh^2 / dR$.

The ansatz given by Eq. (5.3.14) for $z = f(R)$ becomes, after the substitutions $\varepsilon \rightarrow \varepsilon_0$, $V \rightarrow V_0$, $p_0 \rightarrow 0$,

$$(f'/R)' = -\eta \frac{\partial \nu}{\partial R} \left[ \frac{\partial \nu}{\partial z} - \frac{1}{2} \frac{\partial \kappa}{\partial \nu} \right],$$

$$= \eta \frac{\partial \nu}{\partial R} \frac{\partial \nu}{\partial z} \left[ R \frac{\partial \nu}{\partial R} - 1 \right],$$

where $z$ is to be evaluated at $b$. Equation (5.4.7b) follows from Eq. (5.4.7a) given the $(\xi)$ component of the Einstein equation outside the disk.

The equation of state for a relativistic isentropic 2D flow of counter-rotating identical particles is derived in the appendix and reads

$$\varepsilon_0 = 2 p_0 + \frac{\Sigma}{1 + p_0/\Sigma}.$$  (5.4.8)

Solving for $\Sigma$ gives

$$\Sigma = \frac{1}{2} \left[ \varepsilon_0 - 2 p_0 + \sqrt{\varepsilon_0 - 2 p_0 \varepsilon_0 + 2 p_0} \right],$$  (5.4.9)

which in the classical limit gives $\Sigma \rightarrow \varepsilon_0 - p_0$. The binding energy of these counter-rotating disks is computed here from Eq. (5.4.9) together with Eqs. (5.3.11) and (5.4.5). Alternatively, it could be computed via the distribution functions presented in the appendix B.

### 5.4.1 Example 1: The Self-similar disk

The symmetry of the self-similar disk allows one to reduce the partial differential equations corresponding to Einstein’s field equations to an ordinary differential equation with respect to the only free parameter $\theta = \arctan(R/z)$ (Lemos, 89 [65]). Lemos’ solution may be recovered and corrected by the method presented in section 2. Weyl’s metric in spherical polar co-ordinates $(\rho, \chi, \phi)$ defined in terms of $(R, z, \phi)$ by $\rho = \sqrt{\rho^2 + z^2}$ and $\cot(\chi) = z/R$ is

$$d\sigma^2 = -e^{2\nu} d\tau^2 + \rho^2 \left[ e^{-2\nu} \sin^2(\chi) d\phi^2 + e^{2\kappa - 2\nu} \left( \frac{d\rho^2}{\rho^2} + d\chi^2 \right) \right].$$  (5.4.10)

A self-similarity argument led Lemos to the metric

$$ds^2 = -r^{2n} e^{N} dt^2 + r^{2k} \left[ e^{2P - N} d\phi^2 + e^{2Z - N} \left( \frac{dr^2}{r^2} + d\theta^2 \right) \right],$$  (5.4.11)

where $P$, $N$, and $Z$ are given by

$$P(\theta) = \log \left[ \sin (k + n) \theta \right],$$  (5.4.12a)

$$N(\theta) = \frac{4n}{k + n} \log \left[ \cos (k + n) \theta/2 \right],$$  (5.4.12b)

$$Z(\theta) = \frac{8n^2}{(k + n)^2} \log \left[ \cos (k + n) \theta/2 \right] + 2 \log (k + n).$$  (5.4.12c)
Here the $\phi$’s in both metrics have already been identified since they should in both case vary uniformly in the range $[0, 2\pi]$. The rest of the identification between the two metrics follows from the transformation

$$
\begin{align*}
\chi &= (n + k) \theta, \quad \text{(5.4.13a)} \\
\rho &= r^{n+k}, \quad \text{(5.4.13b)} \\
\tau &= t, \quad \text{(5.4.13c)}
\end{align*}
$$

with the result that $\nu$ and $\zeta$ take the form:

$$
\begin{align*}
2\nu &= N + 2n \log r, \quad \text{(5.4.14a)} \\
2\zeta &= Z - 2 \log (n+k). \quad \text{(5.4.14b)}
\end{align*}
$$

This can be written in terms of the new variables given in Eqs. (5.4.13) as

$$
\begin{align*}
\zeta &= \frac{4n^2}{(k + n)^2} \log \left( \frac{\chi}{2} \right), \quad \text{(5.4.15a)} \\
\nu &= \frac{2n}{(k + n)^2} \log \left( \frac{\chi}{2} \right) + \frac{n}{k + n} \log (\rho). \quad \text{(5.4.15b)}
\end{align*}
$$

The procedure described earlier may now be applied. In order to preserve the self-similarity of the solution and match Lemos’ solution, $f(R)$ is chosen so that

$$
f' (R) = \text{const.} = \cot (\bar{\eta}) \quad \text{say.} \quad \text{(5.4.16)}
$$

Then

$$
\frac{\partial}{\partial N} \equiv \left( \frac{\partial}{\partial z} - f' \frac{\partial}{\partial R} \right) = -\frac{1}{\rho \sin (\bar{\eta})} \frac{\partial}{\partial \bar{\eta}}. \quad \text{(5.4.17)}
$$

Putting Eqs. (5.4.15) and Eq. (5.4.17) into Eq. (5.2.15) gives

$$
\begin{align*}
\begin{bmatrix} \varepsilon \\ p_r \\ p_\theta \end{bmatrix} &= \frac{r^{-k}}{4\pi} \left[ \cos \left( \frac{\bar{\eta}}{2} \right) \right]^{2n(k-n)\pi} \left[ \cot (\bar{\eta}) + \frac{2n}{(k+n)^2} \tan (\bar{\eta}/2) \right] \left[ \frac{2n^2}{(k+n)^2} \tan (\bar{\eta}/2) \right], \quad \text{(5.4.18)}
\end{align*}
$$

with $\bar{\eta}$ to be evaluated at $(n+k)\pi/2$. The above equations are in accordance with the solution found by Lemos by directly integrating the Einstein field equations.

### 5.4.2 Example 2: The Curzon disk

The Curzon disk is characterised by the following pair of Weyl functions:

$$
\nu = -\alpha / r, \quad \zeta = -\alpha^2 R^2 / (2 r^4), \quad \text{(5.4.19)}
$$

with $r = \sqrt{R^2 + f^2}$, given the dimensionless parameter $\alpha = M / b$ (recall that $G \equiv c \equiv 1$), all lengths being expressed in units of $b$. This disk is the simplest example of the Kuzmin-Toomre family, where $W(b) \propto \delta(b)$. It also corresponds to the building block of the expansion given in Eqs. (5.4.2). Equations (5.4.3) and (5.4.4) then imply

$$
\begin{align*}
\begin{bmatrix} \varepsilon \\ p_r \\ p_\phi \end{bmatrix} &= \frac{e^{-\frac{\alpha}{2}} (1 - \frac{2\alpha^2}{r^4})}{4\pi b \sqrt{1 + f^2}} \left[ \frac{f''}{1 + f^2} + \frac{f'}{R} + \frac{\alpha}{2} \left( 2 f (r^3 - \alpha R^2) + \left[ (R^2 - f') R f' \right] \right) \right] \quad \text{(5.4.20)}
\end{align*}
$$
The weak energy conditions $\varepsilon \geq 0$, $\varepsilon + p_r \geq 0$, and $\varepsilon + p_\phi \geq 0$ which follow are in agreement with those found by Chamorow et al. (1987) [17]. Their solution was derived by direct integration of Einstein’s static equations using the harmonic properties of the supplementary unknown needed to avoid the patch of Weyl metric above and below the disk. The cut $z = f(R)$ corresponds to the imaginary part of the complex analytic function, the existence of which follows from the harmonicity of that new function.

The ansatz (5.3.14) leads here to the cut

$$
f(R) = b + \frac{\eta}{2} \left[ \frac{b m^2}{4 (b^2 + R^2)} - \frac{2 b m^3 (4 b^2 + 7 R^2)}{105 (b^2 + R^2)^{5/2}} \right]. \tag{5.4.21}
$$

In Fig. 5.4.1, the ratio of the binding energy of these Curzon disks (which is derived from Eqs. (5.3.11), (5.4.9), (5.4.20) and (5.4.21)) over the corresponding rest mass $M_0$ is plotted with respect to the compactness parameter $b/M_0$ for different ratios $p_0/\varepsilon_0$ measured at the origin. The relative binding energy decreases in the most compact configurations because these contain too many unstable orbits ($\kappa^2 < 0$). A maximum ratio of about 5% is reached.
Figure 5.4.2: the binding energy over the rest mass $\Delta E/M_0$ for Curzon disks, plotted against the compactness parameter $\alpha = b/M$ for different ratios $p_0/\varepsilon_0$ fixed by $\eta = 0.01, 0.16, \cdots, 0.7$.

5.5 Distribution functions for rotating super-massive disks

The method described in section 2 will generally induce rotating disks with anisotropic pressures ($p_\phi \neq p_R$). These objects may therefore be described in terms of stellar dynamics. It should be then checked that there exist a stellar equilibria compatible with a given vacuum field and a given cut $z = f(R)$. Here a general procedure to derive all distribution functions corresponding to a specified stress energy tensor is presented, generalising the results of Appendix B to disks with non-zero mean rotation. The detailed description of the dynamics of the disk follows.

In Vierbein components, the stress energy tensor reads

$$T^{(\alpha\beta)} = \int \int \frac{f(\varepsilon, h) P^{(\alpha)} P^{(\beta)}}{P^{(t)}} dP^{(R)} dP^{(\phi)}, \quad (5.5.1)$$

where $f(\varepsilon, h)$ is the distribution of stars at position $R, \phi$ with momentum $P^{(R)}, P^{(\phi)}$. For a stationary disk, it is a function of the invariant of the motion $\varepsilon, h$. Now for the line element (5.2.4):

$$d\sigma^2 = - e^{2\nu} (dt - \omega d\phi)^2 + e^{2\xi - 2\nu} \left(1 + f^2\right) dR^2 + R^2 e^{-2\nu} d\phi^2, \quad (5.5.2)$$

the Vierbein momenta read

$$P^{(\phi)} = e^{\nu} R^{-1} (h - \omega \varepsilon), \quad (5.5.3a)$$
$$P^{(t)} = e^{-\nu} \varepsilon, \quad (5.5.3b)$$
$$P^{(R)} = \left[\varepsilon^2 e^{-2\nu} - \varepsilon^2 R^{-2} (h - \omega \varepsilon)^2 - 1\right]^{1/2}. \quad (5.5.3c)$$
 Calling \( \chi = h/\varepsilon \) and \( \vartheta = 1/\epsilon^2 \), Eqs. (5.5.3) becomes

\[
P^{(\phi)} = e^\nu R^{-1} \vartheta^{-1/2} (\chi - \omega), \tag{5.5.4a}
\]

\[
P^{(t)} = e^{-\nu} \vartheta^{-1/2}, \tag{5.5.4b}
\]

\[
P^{(R)} = \vartheta^{-1/2} \left[ e^{-2\nu} - e^{2\nu} R^{-2} (\chi - \omega)^2 - \vartheta \right]^{1/2}. \tag{5.5.4c}
\]

In terms of these new variables, the integral element \( dP^{(\phi)} dP^{(R)} \) becomes

\[
dP^{(\phi)} dP^{(R)} = \left| \frac{\partial P^{(\phi)} \partial P^{(R)}}{\partial \chi \partial \vartheta} \right| d\chi d\vartheta = \frac{e^{-\nu} R^{-1} \vartheta^{-1/2} d\chi d\vartheta}{2\sqrt{e^{-2\nu} - e^{2\nu} R^{-2} (\chi - \omega)^2 - \vartheta}}. \tag{5.5.5}
\]

Given Eq. (5.5.5) and Eq. (5.5.3b), Eq. (5.5.1) may be rewritten as

\[
T^{(\alpha\beta)} = \int \int P^{(\alpha)} P^{(\beta)} f(\varepsilon, h) R^{-1} \vartheta^{-3/2} d\chi d\vartheta. \tag{5.5.6}
\]

In particular,

\[
RT^{(tt)} e^{2\nu} = \int \int \frac{f(\varepsilon, h) \vartheta^{-5/2} d\chi d\vartheta}{2\sqrt{e^{-2\nu} - e^{2\nu} R^{-2} (\chi - \omega)^2 - \vartheta}}, \tag{5.5.7a}
\]

\[
R^2 T^{(t\phi)} = \int \int (\chi - \omega) \frac{f(\varepsilon, h) \vartheta^{-5/2} d\chi d\vartheta}{2\sqrt{e^{-2\nu} - e^{2\nu} R^{-2} (\chi - \omega)^2 - \vartheta}}, \tag{5.5.7b}
\]

\[
R^3 T^{(\phi\phi)} e^{-2\nu} = \int \int (\chi - \omega)^2 \frac{f(\varepsilon, h) \vartheta^{-5/2} d\chi d\vartheta}{2\sqrt{e^{-2\nu} - e^{2\nu} R^{-2} (\chi - \omega)^2 - \vartheta}}. \tag{5.5.7c}
\]

Note that given Eqs. (5.5.7), \( T^{(RR)} \) follows from the equation of radial support. Let

\[
F(\chi, R) = \frac{1}{2} \int \frac{f(\varepsilon, h) \vartheta^{-5/2} d\vartheta}{\sqrt{e^{-2\nu} - e^{2\nu} R^{-2} (\chi - \omega)^2 - \vartheta}} = \int_1^{\gamma} \frac{f^*(\chi, \vartheta) d\vartheta}{2\sqrt{\gamma - \vartheta}}, \tag{5.5.8}
\]

where \( f^*(\chi, \vartheta) = f(\varepsilon, h) \vartheta^{-5/2} \), and \( \gamma(R, \chi) = e^{-2\nu} - e^{2\nu} R^{-2} (\chi - \omega)^2 \). \( \tag{5.5.9} \)

The set of Eqs. (5.5.7) becomes

\[
RT^{(tt)} e^{2\nu} = \int F(\chi, R) d\chi, \tag{5.5.10a}
\]

\[
R^2 T^{(t\phi)} = \int F(\chi, R) (\chi - \omega) d\chi, \tag{5.5.10b}
\]

\[
R^3 T^{(\phi\phi)} e^{-2\nu} = \int F(\chi, R) (\chi - \omega)^2 d\chi. \tag{5.5.10c}
\]

Any positive function \( F(\chi, R) \) satisfying the moment constraints Eqs. (5.5.10) corresponds to a possible choice\(^2\). Note that the dragging of inertial frames requires to fix three moments of the velocity distribution to account for the given energy density, pressures and rotation law, in contrast with the Newtonian inversion described in chapter III, or the relativistic static inversion.

\(^2\) For instance, a truncated Gaussian parameterised in width, mean, and amplitude similar to that constructed in section 3 of chapter 3.
described in appendix B. Let $\tilde{F}(\chi, Y) = F(\chi, R)$, where $Y(R, \chi)$ is given by Eq. (5.5.9). Written in terms of $\tilde{F}$, the integral equation Eq. (5.5.8) is solved for $f^*$ by an Abel transform

$$f^*(\chi, \vartheta) = \frac{2}{\pi} \int_1^\vartheta \left( \frac{\partial \tilde{F}}{\partial Y} \right) \frac{dY}{\sqrt{\vartheta - Y}}.$$  \hfill (5.5.11)

The latter integration may be carried in $R$ space and yields

$$f(\varepsilon, h) = \frac{2}{\varepsilon^4 \pi} \int \left( \frac{\partial F}{\partial R} \right) \frac{dR}{\chi \sqrt{-\left( P(R) \right)^2}},$$  \hfill (5.5.12)

where $P(R)$ is given by Eq. (5.5.3c) as a function of $R, \varepsilon$, and $h$. Note the minus sign in front of $(P(R))^2$. Here $F$ is chosen to satisfy Eq. (5.5.10) which specifies the stress energy components $T^{(\alpha\beta)}$. All properties of the flow follow in turn from $f(\varepsilon, h)$. For instance, the rest mass surface density, $\Sigma$, may be evaluated from the detailed “microscopic” behaviour of the stars, leading to an estimate of the binding energy of the stellar cluster.

An equation of state for these rotating disks with planar anisotropic pressure tensors is alternatively found directly while assuming (somewhat arbitrarily) that the fluid corresponds to the superposition of two isotropic flows going in opposite directions. In other words, the anisotropy of the pressures measured in the frame co-moving with the mean flow $V$ is itself induced by the counter rotation of two isotropic streams such as those described in section 4. This prescription gives for $\Sigma$:

$$\Sigma = \frac{\varepsilon - p_\phi - p_n + \sqrt{(\varepsilon - p_\phi - p_n)(\varepsilon - p_\phi + 3p_n)}}{2\sqrt{1 - V_0^2}},$$ \hfill (5.5.13)

where $V_0 = \sqrt{(p_\phi - p_n)/(\varepsilon + p_n)}$ is the counter rotating velocity measured in the frame co-moving with $V$ which induces $p_\phi \neq p_n$. In the classical limit, $\Sigma \to (\varepsilon - p_\phi)/\sqrt{1 - V_0^2}$. The binding energy of these rotating disks is computed from Eq. (5.5.13) together with Eqs. (5.3.11) and (5.2.15).

5.6 Application: rotating disks

The problem of constructing disks with proper rotation and partial pressure support is more complicated but physically more appealing than that of counter rotating solutions. It is now illustrated on an internal solution for the Kerr metric, and a generalisation of the solutions constructed in the previous chapter to warm rotating solutions is sketched.
5.6.1 A Kerr internal solution

The functions \((\nu, \zeta, \omega)\) for the Kerr metric in Weyl-Papapetrou form expressed in terms of spheroidal co-ordinates \((R = s\sqrt{(x^2 + 1)(1 - y^2)}, \ z = sxy)\) are given by

\[
e^{2\nu} = \frac{-m^2 + s^2 x^2 + a^2 y^2}{(m + s x)^2 + a^2 y^2},
\]

\[
e^{2\zeta} = \frac{-m^2 + s^2 x^2 + a^2 y^2}{s^2 (x^2 + y^2)},
\]

and

\[
\omega = \frac{-2 a m (m + s x) (1 - y^2)}{-m^2 + s^2 x^2 + a^2 y^2},
\]

where \(s = (\pm(m^2 - a^2))^{1/2}\). Here \pm corresponds to the choice of prolate (+), or oblate (−) spheroidal co-ordinates corresponding to the cases \(a < m\) and \(a > m\) respectively. In this set of co-ordinates, the prescription given in section 2 leads to completely algebraic solutions. The normal derivative to the surface \(z = f(R)\) reads (when \(a > m\))

\[
(\partial/\partial N) = N(x, y) (\partial/\partial x) + N(y, x) (\partial/\partial y),
\]

where

\[
N(x, y) = (x^2 - 1) [y/s + x (y^2 - 1) f'/R]/(x^2 - y^2).
\]

Differentiating Eqs. (5.6.1) through (5.6.3) with respect to \(x\) and \(y\) together with Eq. (5.6.4) leads via Eqs. (5.2.15) to all physical characteristics of the Kerr disk\(^3\) in terms of the Kerr metric parameters \(m\) and \(a\), and the function \(z = f(R)\) which must be chosen so as to provide relevant pressures and energy distributions according to Eqs. (5.3.12).

On the axis \(R = 0\), \(f \equiv b\), and \(f'' = -c\), while Eqs. (5.3.12) imply \(f' = 0\) at \((x = b/s, \ y = 1)\). This in turn implies

\[
(p_\phi)_0 = (p_\phi)_0 = \frac{c}{4\pi} \sqrt{\frac{a^2 + b^2 - m^2}{a^2 + (b + m)^2}},
\]

\[
(\varepsilon)_0 = \frac{1}{4\pi} \sqrt{\frac{a^2 + b^2 - m^2}{a^2 + (b + m)^2}} \left[ \frac{2m}{(b + m)^2} \left( \frac{(b + m)^2 - a^2}{(a^2 + (b + m)^2)(a^2 + b^2 - m^2)} \right) - c \right],
\]

where \((\ )_0\) stands for \((\ )\) taken at \(R = 0\). The constraint that all physical quantities should remain positive implies

\[
b \geq \sqrt{m^2 - a^2},
\]

\[
c \leq 2m \frac{(b + m)^2 - a^2}{(a^2 + (b + m)^2)(a^2 + b^2 - m^2)}.
\]

Equation (5.6.6a) is the obvious requirement that the surface of section should not enter the horizon of the fictitious source. Similarly, ergoregions will arise when the cut \(z = f(R)\) enter the torus

\[
\left( \frac{z}{m} \right)^2 + \left( \frac{R}{m} - \frac{a}{m} + \frac{1}{2} \frac{m}{a} \right)^2 = \left( \frac{1}{2} \frac{m}{a} \right)^2.
\]

\(^3\) Bicak & Ledvinka (1993) [12] have constructed independently a cold Kerr solution when \(a < m\).
Figure 5.6.1: the azimuthal velocity (a), the radial pressure (b), the ratio of radial over azimuthal pressure and the relative fraction of Kinetic Support (d) namely $\varepsilon V^2/(\varepsilon V^2 + p_\phi + p_R)$ for the $a/m = 0.5$ Kerr disk as a function of circumferential radius $R_c$, for a class of solutions with decreasing potential compactness, $b/m = 1.1, 1.2 \cdots 1.7$, and relative radial pressure support $\eta/m = 1/5 + 2(b/m - 1)/5$.

The analysis may be extended beyond the ergoregion via the ZAMO frame. The characteristics of these frames are given by Eq. (5.3.7), (5.3.8) and (5.3.9). The central redshift

$$1 + z_c = e^{-\nu} = \frac{(m + b)^2 + a^2}{b^2 + a^2 - m^2},$$

(5.6.8)
can be made very large for such models constructed when $b \to \sqrt{m^2 - a^2}$.

On Fig. 5.6.1, Fig. 5.6.2 and Fig. 5.6.3, some characteristics of these Kerr disks with both $a > m$ and $a < m$ are illustrated. For simplicity, pressure is implemented via the cut (5.4.21). These disks present anisotropy of the planar pressure tensor ($p_\phi \neq p_R$).

The anisotropy follows from the properties of the $\omega$ vacuum field which gives rise to the circular velocity curve. Indeed, in the outer part of the disk, the specific angular momentum $h$ tends asymptotically to a constant. The corresponding centrifugal force is therefore insufficient to counter-balance the field generated by the $\nu$ function. As this construction scheme generates such equilibria, the system compensates by increasing its azimuthal pressure away from isotropy. When $a > m$, and $\eta = 0$ the height of the critical cut corresponding to the last disk with positive pressures everywhere scales like $a \left(a^2 - m^2\right)^{1/2}$ in the range $2 \leq a/m \leq 20$. Note that this limit is above that of the highly relativistic motions which only occurs when $b$ tends asymptotically to the Kerr horizon $(a^2 - m^2)^{1/2}$. The binding energy which follows Eq. (5.5.13) reaches values as high as one tenth of the rest mass for the most compact configurations.
Figure 5.6.2: as in Fig. 5.6.1 with $a/m = 0.95$, $\eta/m = 1/15 + (b - 1)/5$. Note the large relative kinetic support.

5.6.2 Other rotating solutions

The rotating disk models given in section 2 require prior knowledge of the corresponding vacuum solutions. However, while studying the symmetry group of the stationary axially symmetric Einstein Maxwell equations, Hoenselaers, Kinnersley & Xanthopoulos (1979) \cite{HKX} found a method of generating systematically complete families of stationary Weyl-Papapetrou vacuum metrics from known static Weyl solutions. Besides, the decomposition of the Weyl potential $\nu_0$ into line densities provides a direct and compact method of finding solutions to the static field. The N+1 rank zero HKX transform is defined as follows: let $\nu_0$ and $\zeta_0$ be the seed Weyl functions given by Eqs. (5.6.4). HKX define the $N \times N$ matrix $\Gamma$ parameterised by $N$ twist parameters $\alpha_k$ and $N$ poles, $a_k$, $k = 1, \ldots, N$ as

$$ (\Gamma^\pm)_{k,k'} = i \alpha_k \frac{e^{\beta_k}}{r_k} \left[ \frac{r_k - r_{k'}}{a_k - a_{k'}} \mp 1 \right], \quad (5.6.9) $$

with $r_k = r(a_k) = \sqrt{R^2 + (z - a_k)^2}$ and $\beta_k = \beta(a_k)$, $\beta$ being a function satisfying (5.6.11a) below. Defining $r = \text{Diag}(r_k)$, they introduce the auxiliary functions

$$ D^\pm = |1 + \Gamma^\pm|, \quad L_\pm = 2D^\pm \text{tr} \left[ (1 + \Gamma^\pm)^{-1} \Gamma^\pm r \right], $$

$$ \omega_\pm = D^+ \pm e^{2\nu_0} D^-, \quad M_\pm = \varpi \omega_\pm + 2 \left( L_+ \mp e^{2\nu_0} L_- \right), $$

which lead to the Weyl-Papapetrou potentials

$$ e^{2\nu} = e^{2\nu_0} \Re \left[ \frac{D^-}{D^+} \right], \quad (5.6.10a) $$
Figure 5.6.3: the azimuthal velocity (a), the azimuthal pressure (b), the relative fraction of Kinetic Support (c), namely $\varepsilon V^2/(\varepsilon V^2 + p_\phi)$ and the angular momentum (d) for the $a/m = 10$ zero radial pressure Kerr disk as a function of circumferential radius $R_c$, for a class of solutions with decreasing potential compactness $b/m = 210, 260 \cdots 510$. The first curve corresponds to a cut which induces negative azimuthal pressure (tensions).
Section 5.6 Application: rotating disks

characterising the boundary conditions at infinity. The two yet unspecified function equations suggests again that solutions should be sought in terms of the line density, by

\[ e^{2\zeta} = ke^{2\zeta_0} \Re \left[ D^{-} D^+ \right], \]
\[ \omega = 2\Im \left[ \mathcal{M}_+ (\omega_+ + \omega_-) - \mathcal{M}_- (\omega_+ + \omega_-)^* \right] \left\{ |\omega_+|^2 - |\omega_-|^2 \right\}, \]

where \( \Re \) and \( \Im \) stand respectively for the real part and the imaginary part of the argument and \( (\ )^* \) represents the complex conjugate of \( (\ ) \). \( k \) is a constant of integration which is fixed by the boundary conditions at infinity. The two yet unspecified function \( \beta \) and \( \varpi \) satisfy:

\[ \nabla \beta(a_k) = \frac{1}{r_k} ([z - a_k] \nabla + R e_\phi \times \nabla) \nu_0, \]
\[ \nabla \varpi = Re_\phi \times \nabla \nu_0, \]

where \( \nabla = (\partial/\partial R, \partial/\partial z) \) and \( e_\phi \) is the unit vector in the \( \phi \) direction. The linearity of these equations suggests again that solutions should be sought in terms of the line density, \( W(b) \), characterising \( \nu_0 \), namely

\[ \beta(a_k) = \int \frac{W(b)}{a_k} \left[ \frac{R^2 + (z - b - a_k)^2}{R^2 + (z - b)^2} \right]^{\frac{1}{2}} db, \]
\[ \varpi = \int \frac{W(b)(b - z)}{\sqrt{R^2 + (z - b)^2}} db. \]

It is therefore a matter of algebraic substitution to apply the above procedure and construct all non-linear stationary vacuum fields of the Papapetrou form from static Weyl fields. Following the prescription described in section 2, the corresponding disk in real rotation is then constructed. The requirement for the physical source to be a disk is in effect a less stringent condition on the regularity of the vacuum metric in the neighbourhood of its singularity because only the half space which does not contain these singularities is physically meaningful. A suitable Ehlers transformation on \( D_-/D^+ \) ensures asymptotic flatness at large distance from the source.

To illustrate this prescription consider the vacuum field given by Yamazaki (1981) [113]. This field arises from 2 HKX rank zero transformation on the Zipoy-Voorhess static metric given by

\[ \nu_0 = 4\delta \log \frac{(x - 1)}{(x + 1)}, \]
\[ \zeta_0 = 4\delta^2 \log \frac{(x - 1)}{(x^2 - y^2)}, \]

in prolate spheroidal co-ordinates. It corresponds to a uniform line density between \( \pm \delta \). The functions \( \beta \) and \( \varpi \) follow from Eqs. (5.6.12). Choosing poles \( a_k, \ k = 1, 2 \) at the end of the rod \( \pm \delta \), leads to

\[ \beta(a_\pm) = \frac{4\delta \log \frac{x^2 - 1}{|x \mp y|}}, \]
\[ \varpi = 2\delta y. \]

Equations (5.6.10) together with (5.6.13) and (5.6.11) characterize completely the three Weyl-Papapetrou metric functions \( \nu, \zeta, \) and \( \omega \); all physical properties of the corresponding disks follow. For instance, Fig. 5.6.4 gives the zero radial pressure velocity curve of the Yamazaki disk spanning from the Schwarzschild (\( \delta = 1/2 \)) metric when \( \alpha_1 = \alpha_2 \).
Dynamics of rotating super-massive disks Chapter 5

Figure 5.6.4: the azimuthal velocity for the $a/\kappa_Y$ zero radial pressure Yamazaki disk as a function of circumferential radius $R_c$, for a class of solutions with decreasing potential compactness $b/\kappa_Y = 3.6, 3.8, \cdots, 5$. $\kappa_Y$ is the natural unit of distance in these disks as defined by Yamazaki’s Eq. (Y-2) [113].

More generally, the extension of this work to the construction of disks with planar isotropic pressures should be possible by requiring that Eq. (5.2.15c) is identically equal to Eq. (5.2.15d) given Eq. (5.6.10).

5.7 Conclusion

The general counter rotating disk with partial pressure support has been presented. The suggested method for implementing radial pressure support also applies to the construction of stationary axisymmetric disk with rotation but requires prior knowledge of the corresponding vacuum solution. In this manner, a disk-like source for the Kerr field has been constructed. The corresponding distribution reduces to a Keplerian flow in the outer part of the disk and presents strong relativistic features in the inner regions such as azimuthal velocities close to that of light, large central redshift and ergoregions. The disk itself is likely to be stable against ring formation and the ratio of its binding energy to rest mass can be as large as $1:10$. The broad lines of how to construct all rotating disks arising from HKX-transforming the corresponding counter rotating model into a fully self-consistent model with proper rotation and partial pressure support has been sketched. It should be a simple matter to implement this method with the additional requirement that the pressure remains isotropic with a sensible polytropic index. One could then analyse the fate of a sequence of gaseous disks of increasing compactness and relate it to the formation and evolution of quasars at high redshift. Alternatively, the inversion method described in section 5 yields a consistent description of all disks in terms of stellar dynamics. In those circumstances, the inversion method for distribution functions presented in section 5 yields a complete description of the detailed dynamics of these objects. It would be worthwhile to construct distribution functions for the Kerr disk presented in section 6.
Aspects of the dynamics of relativistic disks

D.1 Equation of state for a relativistic 2D adiabatic flow

The equation of state of a planar adiabatic flow is constructed here by relating the perfect fluid stress energy tensor of the flow to the most probable distribution function which maximises the Boltzmann entropy. These properties are local characteristics of the flow. It is therefore assumed that all tensorial quantities introduced in this section are expressed in the local Vierbein frame.

Consider a infinitesimal volume element $d\vartheta$ defined in the neighbourhood of a given event, and assume that the particles entering this volume element are subject to molecular chaos. The distribution function $F$ of particles is defined so that $F(R, P) \, d\vartheta \, d^2P$ gives the number of particles in the volume $\vartheta$ centred on $R$ with 2-momentum pointing to $P$ within $d^2P$.

The most probable distribution function $F^*$ for these particles is then given by that which maximises Boltzmann entropy

$$S = -d\vartheta \int F \log F \, d^2P. \quad (D.1.1)$$

This entropy reads in terms of the distribution $F$

$$S = -d\vartheta \int F \log F \, d^2P. \quad (D.1.1)$$

The stress energy tensor corresponds to the instantaneous flux density of energy momentum through the elementary volume $d\vartheta$ (here a surface)

$$T^{\alpha\beta} = \int F^* P^\alpha P^\beta \frac{d^2P}{|E|}, \quad (D.1.2)$$

where the $1/|E|$ factor accounts for the integration over energy-momentum space to be restricted to the pseudo-sphere $P^\alpha P_\alpha = m^2$. Indeed the detailed energy-momentum conservation requires the integration to be carried along the volume element

$$\int \delta (P^\alpha P_\alpha - m^2) \, d^3P = \int \delta (P^0^2 - \mathcal{E}^2) \, d^2P_0P^0 = \frac{d^2P}{|\mathcal{E}|}, \quad (D.1.3)$$

\(1\) Synge [100] gives an extensive derivation of the corresponding equation of state for a 3D flow.
where $E = \sqrt{m^2 + P^2}$. It is assumed here that momentum space is locally flat. Similarly the numerical flux vector reads
\[
\phi^\alpha = \int F^\alpha P \frac{d^2 P}{|E|}.
\]

The conservation of energy-momentum of the volume element $d\theta$ implies that the flux of energy momentum across that volume should be conserved; this flux reads
\[
T^{\alpha\beta} n_\beta d\theta = d\theta \int F^\alpha P d^2 P = \text{Const.},
\]
where $n_\beta$ is the unit time axis vector. Keeping the population number constant provides the last constraint on the possible variations of $S$
\[
\phi^\alpha n_\alpha d\theta = d\theta \int F^\alpha d^2 P = \text{Const.}.
\]

Varying Eq. (D.1.1) subject to the constraints (D.1.5) (D.1.6), and putting $\delta S$ to zero leads to
\[
(\log F^* + 1) \delta F^* = a \delta F^* + \lambda_\alpha P^\alpha \delta F^*,
\]
where $a$ and $\lambda_\alpha$ are Lagrangian multipliers corresponding resp. to (D.1.5) (D.1.6). These multipliers are independent of $P^\alpha$ but in general will be a function of position. The distribution which extremises $S$ therefore reads
\[
F^* = C \exp (\lambda_\alpha P^\alpha).
\]

The 4 constants $C$ and $\lambda_\alpha$ are in principle fixed by Eqs. (D.1.5) and (D.1.6). The requirement for $F^*$ to be Lorentz invariant – that is independent of the choice of normal $n_\alpha$ – and self consistent, is met instead by demanding that $F^*$ obeys Eq. (D.1.2) and (D.1.4), namely
\[
C \int P^\alpha \exp (\lambda_\mu P^\mu) \frac{d^2 P}{|E|} = \phi^\alpha,
\]
\[
C \int P^\alpha P^\beta \exp (\lambda_\mu P^\mu) \frac{d^2 P}{|E|} = T^{\alpha\beta},
\]
where $\phi^\alpha$ and $T^{\alpha\beta}$ satisfy in turn the covariant constraints
\[
\frac{\partial T^{\alpha\beta}}{\partial x^\beta} = 0 \quad \text{and} \quad \frac{\partial \phi^\alpha}{\partial x^\alpha} = 0.
\]

Equations (D.1.9)-(D.1.11) provide a set of thirteen equations to constraint the thirteen functions $C, \lambda_\alpha, \phi^\alpha$, and $T^{\alpha\beta}$. Equations (D.1.9) and (D.1.10) may be rearranged as
\[
\phi^\alpha = C \frac{\partial \Phi}{\partial \lambda^\alpha}, \quad T^{\alpha\beta} = C \frac{\partial^2 \Phi}{\partial \lambda^\alpha \partial \lambda^\beta},
\]
where the auxiliary function $\Phi$ is defined as $\Phi = \int \exp (\lambda_\alpha P^\alpha) d^2 P/|E|$; $\Phi$ is best evaluated using pseudo polar coordinates corresponding to the symmetry imposed by the energy momentum conservation $P^\alpha P_\alpha - m^2 = 0$; these are
\[
P^1 = m \sinh \chi \cos \phi,
\]
\[
P^2 = m \sinh \chi \sin \phi,
\]
\[
P^0 = im \cosh \chi,
\]
where the coordinate $\chi$ is chosen so that the normal $n_\alpha$ lies along $\chi = 0$. In terms of these variables, $d^2P/|E|$ then reads
\[ d^2P/|E| = dP^4dP^2/m \cosh \chi = m \sinh \chi d\chi d\phi. \quad (D.1.14) \]
Moving to a temporary frame in which $\lambda^0$ is the time axis ($\lambda^0 = i\lambda = i(-\lambda_\alpha \lambda^\alpha)^{1/2}$, $\lambda^k = 0$ $k = 1, 2$), $\Phi$ becomes
\[ \Phi = 2\pi m \int_0^\infty \exp(-m\lambda \cosh \chi) \sinh \chi d\chi = \frac{2\pi}{\lambda} \exp(-\lambda). \quad (D.1.15) \]
From Eq. (D.1.9)-(D.1.10) and (D.1.15), it follows that
\[ T^{\alpha\beta} = \frac{2\pi C}{\lambda^3} \exp(-\lambda) \left[ \{\lambda^2 + 3\lambda + 3\} \frac{\lambda^\alpha \lambda^\beta}{\lambda} + \{\lambda + 1\} \delta^{\alpha\beta} \right], \quad (D.1.16) \]
\[ \phi^\alpha = \frac{2\pi C}{\lambda^3} \exp(-\lambda) \{\lambda + 1\} \lambda^\alpha. \quad (D.1.17) \]
On the other hand, the stress energy tensor of a perfect fluid is
\[ T^{\alpha\beta} = (\varepsilon_0 + p_0) U^\alpha U^\beta + p_0 \delta^{\alpha\beta}, \quad (D.1.18) \]
while the numerical surface density of particles measured in the rest frame of the fluid $\Sigma$ is related to $\phi^\alpha$ via
\[ \Sigma = -\phi^\alpha U_\alpha. \quad (D.1.19) \]
Eq. (D.1.16) has clearly the form of Eq. (D.1.18) when identifying
\[ U^\alpha = \lambda^\alpha/\lambda, \quad (D.1.20a) \]
\[ \varepsilon_0 + p_0 = \frac{2\pi C}{\lambda^3} \exp(-\lambda) \{\lambda^2 + 3\lambda + 3\}, \quad (D.1.20b) \]
\[ p_0 = \frac{2\pi C}{\lambda^3} \exp(-\lambda) \{\lambda + 1\}. \quad (D.1.20c) \]
Eliminating $\lambda, C$ between Eqs. (D.1.17)-(D.1.20c) yields the sought after equation of state
\[ \varepsilon_0 = 2p_0 + \frac{\Sigma}{1 + p_0/\Sigma}. \quad (D.1.21) \]
where $p_0 = \Sigma/\lambda$. This is the familiar gas law given that $1/\lambda$ is the absolute temperature. The equation of state Eq. (D.1.21) corresponds by construction to an isentropic flow. Indeed, using the conservation equations Eq. (D.1.11a) dotted with $U^\alpha$, and Eq. (D.1.11b) together with $U^\alpha U_\alpha = -1$ yields after some algebra to the identity
\[ \lambda \frac{d}{ds} \left[ \frac{1}{\lambda} \left( \lambda^2 + 3\lambda + 3 \right) \right] + 1 = -\frac{\partial U^\alpha}{\partial x^\alpha} \frac{1}{\Sigma} \frac{d\Sigma}{ds}, \quad (D.1.22) \]
where $d/ds = (dx^\alpha/dx^\alpha)(\partial/\partial x^\alpha)$ is the covariant derivative following the stream lines. Writing the r.h.s of Eq. (D.1.22) as an exact covariant derivative leads to the first integral
\[ \frac{d}{ds} \left[ \frac{1}{1 + \lambda} - 2 \log(\lambda) + \log(1 + \lambda) - \log(\Sigma) \right] = 0. \quad (D.1.23) \]
Given that $p_0 = \Sigma/\lambda$, it follows
\[ \left( 1 + \frac{p_0}{\Sigma} \right) \frac{p_0}{\Sigma} \exp\left( \frac{1}{1 + \Sigma/p_0} \right) \frac{1}{\Sigma} = \text{Const.} \quad (D.1.24) \]
which in the low temperature limit gives $\Sigma^{-2}p_0 = \text{Const}$. This corresponds to the correct adiabatic index $\gamma = (2 + 2)/2$. 

Section D.2 The relativistic epicyclic frequency
D.2 The relativistic epicyclic frequency

The equation for the radial oscillations follows from the Lagrangian

\[ \mathcal{L} = -\sqrt{e^{2\nu} \left( \dot{t} - \omega \dot{\phi} \right)^2 - e^{2\xi-2\nu} (1 + f'^2) \dot{R}^2 - e^{-2\nu} \phi^2 R^2}, \]  

(D.2.1)

where ( ) stands here for derivation with respect to the proper time \( \tau \) for the star describing its orbit. In Eq. (D.2.1), \( \phi \) and \( t \) are ignorable which leads to the invariants:

\[ \frac{\partial \mathcal{L}}{\partial \dot{\phi}} = h \quad \frac{\partial \mathcal{L}}{\partial \dot{t}} = -\epsilon. \]  

(D.2.2)

The integral of the motion for the radial motion follows from Eqs. (D.2.2) and \( U^\mu U_\mu = -1 \), namely:

\[ e^{2\xi-2\nu} (1 + f'^2) \dot{R}^2 + \frac{e^{2\nu}}{R^2} (h - \epsilon \omega)^2 - \epsilon^2 e^{-2\nu} = -1 \]

This equation, together with its total derivative with respect to proper time provides the radial equation of motion, having solved for the angular momentum \( h \) and the energy \( \epsilon \) of circular orbits as a function of radius when equating both \( \dot{R} \) and \( \ddot{R} \) to zero. the relativistic generalisation of the classical epicyclic frequency is defined here to be the frequency of the oscillator calling back linear departure from circular orbits. Hence the equation for radial perturbation reads

\[ \delta \ddot{R} + \kappa^2 \delta R = 0, \]  

(D.2.3)

which gives for \( \kappa^2 \) :

\[ \kappa^2 = e^{-2Z} (Z'' - 2Z'^2) - e^2 e^{-2L} (L'' - 2L'^2) - e^{2K} \left[ (h - \epsilon \omega)^2 (K'' + 2K'^2) - \epsilon (h - \epsilon \omega) (4K' \omega' + \omega'') + \epsilon^2 \omega'^2 \right]. \]  

(D.2.4)

where ( )' stands in this Appendix for \( d/dR \equiv \partial/\partial R + f' \partial/\partial z \) and \( L, K, \) and \( Z \) are given in terms of the potential \( \zeta \) and \( \nu \) by:

\[ L = \zeta + \log \sqrt{1 + f'^2}, \]  

(D.2.5a)

\[ Z = \zeta + \log \sqrt{1 + f'^2} - \nu, \]  

(D.2.5b)

\[ K = 2\nu - \zeta - \log \sqrt{1 + f'^2} - \log R. \]  

(D.2.5c)

Here \( \epsilon \) and \( h \) are known functions of \( R \) given by the roots of:

\[ \dot{R} = 0 \quad \Rightarrow \quad e^{-2L} \epsilon^2 = (h - \epsilon \omega)^2 e^{2N} + e^{-2Z}, \]  

(D.2.6a)

\[ \ddot{R} = 0 \quad \Rightarrow \quad e^{-2Z} Z' = e^{-2L} L' \epsilon^2 + (h - \epsilon \omega) e^{2N} [(h - \epsilon \omega) N'' - \epsilon \omega']. \]  

(D.2.6b)
D.3 Christoffel Coefficients for the WP metric

The connection coefficients of the Weyl-Papapetrou metric are given by:

\[ \Gamma'_{Rt} = \frac{\partial \nu}{\partial R} + \frac{1}{2 R^2} e^4 \nu \frac{\partial \omega}{\partial R}, \]  
(D.3.1)

\[ \Gamma'_{zt} = \frac{\partial \nu}{\partial z} + \frac{1}{2 R^2} e^4 \nu \frac{\partial \omega}{\partial z}, \]  
(D.3.2)

\[ \Gamma'_{\phi R} = \frac{\omega}{R} - 2 \frac{\partial \nu}{\partial R} - \frac{1}{2 R^2} e^4 \nu \frac{\partial \omega}{\partial R} \]  
(D.3.3)

\[ \Gamma'_{\phi z} = -2 \frac{\partial \nu}{\partial z} - \frac{1}{2 R^2} e^4 \nu \frac{\partial \omega}{\partial z}, \]  
(D.3.4)

\[ \Gamma^{R}_{tt} = e^4 \nu - 2 \zeta \frac{\partial \nu}{\partial R}, \]  
(D.3.5)

\[ \Gamma^{R}_{RR} = -\frac{\partial \nu}{\partial R} + \frac{\partial \zeta}{\partial R}, \]  
(D.3.6)

\[ \Gamma^{R}_{zR} = -\frac{\partial \nu}{\partial z} + \frac{\partial \zeta}{\partial z}, \]  
(D.3.7)

\[ \Gamma^{R}_{\phi t} = e^4 \nu - 2 \zeta \left( -\frac{\omega}{R} - 2 \frac{\partial \omega}{\partial R} \right), \]  
(D.3.8)

\[ \Gamma^{z}_{tt} = e^4 \nu - 2 \zeta \frac{\partial \nu}{\partial z}, \]  
(D.3.9)

\[ \Gamma^{z}_{RR} = \frac{\partial \nu}{\partial R} - \frac{\partial \zeta}{\partial R}, \]  
(D.3.10)

\[ \Gamma^{z}_{zR} = -\frac{\partial \nu}{\partial z} + \frac{\partial \zeta}{\partial z}, \]  
(D.3.11)

\[ \Gamma^{z}_{\phi t} = e^4 \nu - 2 \zeta \left( -\frac{\omega}{R} - 2 \frac{\partial \omega}{\partial z} \right), \]  
(D.3.12)

\[ \Gamma^{\phi}_{Rt} = \frac{1}{2 R^2} e^4 \nu \frac{\partial \omega}{\partial R}, \]  
(D.3.13)

\[ \Gamma^{\phi}_{zt} = \frac{1}{2 R^2} e^4 \nu \frac{\partial \omega}{\partial z}, \]  
(D.3.14)

\[ \Gamma^{\phi}_{\phi R} = -\frac{\partial \nu}{\partial R} - e^4 \nu \frac{\partial \omega}{\partial R} \frac{1}{2 R^2}, \]  
(D.3.15)
Epilogue

The work described in the previous four chapters presented new aspects of the dynamics of flattened self gravitating systems. It has covered the following aspects:

- Orbital instabilities in galaxies

Instabilities of stellar systems are worth studying because they may account for the structure of observed and numerical data on disks. Unstable configurations are also possible output of galaxy formation and stability analysis puts constraints on the viability of models which are supposed to account for observed galaxies. Considerable efforts have been devoted to understanding the stability of bars within galaxies, but the problem of their formation remains to a large extent a puzzle. This question was addressed here by considering the linear stability of self-gravitating stellar systems from the viewpoint of instabilities triggered by interactions between orbits. It was shown that the set of equations describing the linear instability can be exactly written in terms of the interaction of orbital streams, i.e. resonant flows of stars. For a given resonance this interaction leads to an instability of Jeans’s type. Jeans’s gravitational instability can be thought of as a process in which more and more particles are trapped in a growing potential well which may move relative to an inertial frame. In galaxies azimuthal gravitational instability works by trapping the lobes of orbital streams of stars moving azimuthally in a growing potential well which may be rotating. Such instabilities can only occur if the orbital lobes respond by moving in the direction of the torques applied on them. This co-operation occurs at the inner Lindblad resonance in the central parts of galaxies. The general picture combines the effect of all resonances compatible with the symmetry of the mode under consideration. This treatment is formally related to the potential density pair method of Kalnajs, but treats both Boltzmann and Poisson’s equations directly in action space. The basis functions on which the perturbation is expanded then correspond to known families of orbits for which the exact coupling factors describing the interaction have been computed. These correspond to the core of Poisson’s equation written in angle and action variables. Identifying the fastest growing mode and its eigen-space yields the roots of the orbital instability. The fate of the above orbital instabilities has been investigated and the type of distribution functions expected from the adiabatic re-alignment of resonant orbits under the assumption of a unique inner Lindblad resonance was derived. It was assumed that the star trajectories would evolve towards their state of maximum entropy, given the constraints fixed by conservation of circulation, overall angular momentum and energy. This analysis yields a critical temperature below which the preferred state of equilibria is that of a barred galaxy which may display negative specific heat. The amplitude and pattern of a bar is then related to the other macroscopic properties of the flow. Conversely the criteria states that barred configurations are unstable to orbital dilution for some range of energies and momenta.

- Distribution functions for flat galaxies: observational implications
Simple inversion formulae were given to construct distribution functions for flat systems whose surface density and Toomre’s Q number profile. These distributions describe stable models with realistic velocity distributions for the power law, the Isochrone and the Kuzmin disks. The purpose of these functions is to provide plausible galactic models and assess their critical stability with respect to global non-axisymmetric modes. The inversion may also be carried out for a given azimuthal velocity distribution (or a given specific energy distribution) which may either be observed or chosen accordingly. This method should be applied to real galaxies and provides a unique opportunity to confront predictions on the radial velocity distributions with the same observed quantity. The likely residual could be used to assess the departures from axisymmetry or other limiting aspects of the model. Sets of initial conditions for an N-body simulation may be extracted from the distribution function. A numerical stability analysis may then be carried out to address the intrinsic tendency for the disk to form such non-axisymmetric structures. The properties of the galaxy when evolved forward in time should provide an unprecedented link between the theory of galactic disks and the detailed observation of the kinematics of these objects which may be achieved with today’s technology.

- Relativistic stellar dynamics of flattened systems and quasar formation

Complete families of new exact analytical solutions of Einstein’s equations have been constructed in closed form for thin axisymmetric rotating disks derived from new metrics with gradient discontinuities. A general geometrical procedure is presented to implement partial radial pressure support within these disks. These solutions are relevant to astronomy for the following reasons: firstly, one of these families represents a possible internal solution for the Kerr metric which is ultimately the field generated by a rotating black hole. This internal solution had been sought for thirty years; secondly, these models correspond to dynamically (locally) stable, differentially rotating, relativistic flattened stellar clusters for which distribution functions are available. When the pressure is isotropic, these configurations describe super-massive disks corresponding to possible models for the latest stage of the collapse of a proto-quasar. The binding energy of these thin disks can reach values as high as ten percent of the rest mass, suggesting that the corresponding configuration might be sufficiently long lived to have observational implications. A natural follow-up of this work would be to study the global dynamical stability of these disks. It would also be worthwhile to address the secular stability of super-massive disks by studying the effect of viscosity and radial angular momentum transport. These would undoubtedly induce a relative broadening of the disk which should be assessed. The corresponding model could then be contrasted with that of the spherical collapse of a non-rotating super-massive star. The detailed dynamical properties of massive flattened stellar clusters also deserve further work.
Une fraction importante des disques galactiques observés présentent une structure barrée qui se caractérise par des isophotes allongés au delà du bulbe central. L’analyse photométrique de ces isophotes suggère que la barre est constituée d’étoiles âgées; la barre ne semble donc pas être le fruit d’un phénomène transitoire, et la description du système en termes de dynamique stellaire paraît appropriée pour comprendre sa formation et son évolution. De gros efforts ont été consacrés ces dernières années à l’étude de la stabilité de ces objets mais leur formation reste à ce jour un problème ouvert. Si l’est vrai que les simulations numériques sont sujettes à de fortes instabilités bi-symétriques, seulement quarante pour cent des disques observés sont constitués d’une forte barre. Pourquoi certaines galaxies forment-elles une barre et d’autres pas? J’ai tenté de répondre à cette question en menant l’étude de la stabilité d’un système stellaire autogratiant avec pour point de vue les instabilités induites par l’alignement d’orbites résonnantes. De fait, j’ai montré qu’il est possible de re-écrire les équations de la dynamique exclusivement en termes de couplage entre des flots d’étoiles décrits par une même orbite résonante. En portant mon attention sur les étoiles appartenant à un flot en résonance interne de Lindblad, j’ai montré que ce couplage conduisait à une instabilité de type instabilité de Jeans. L’instabilité de Jeans peut être interprété comme un processus au cours duquel plus en plus d’étoiles sont capturées par un potentiel d’amplitude croissante. De la même manière, ici l’instabilité azimutale procède par capture du lobe des orbites résonnantes par un potentiel tournant. Le taux de précession de ce potentiel d’amplitude croissante correspond alors au taux de précession de la barre. Ce type d’instabilité n’est possible que si le moment d’inertie adiabatique des orbites piégées est positif. Dans ce cas, l’orbite résonnante répond dynamiquement à un couple attracteur par un déplacement dans la direction de la source de ce couple. Ce mouvement coopératif a lieu à la résonance interne de Lindblad dans les parties centrales des galaxies. Le critère d’instabilité global est obtenu en considérant l’effet de toutes les résonances compatibles avec la symétrie de l’instabilité considéré. Une relation de dispersion pour la fréquence et le taux de croissance de la barre se déduit par projection, soit sur la base des fonctions propres du noyau de couplage entre les orbites dans l’approximation épicyclique, soit sur une base complète de fonctions de distribution susceptible de bien décrire la géométrie de la barre dans le régime linéaire. Dans ce dernier cas, mon analyse s’apparée à la méthode des paires densité-potentiel de Kalnajs, mais présente l’avantage de traiter le couplage directement dans l’espace des phases, ce qui permet d’identifier les orbites responsables de l’instabilité. Les caractéristiques de l’onde stationnaire de densité déduits par simple intégration sur l’espace des vitesses. J’ai par ailleurs étudié le devenir de ce type d’instabilité en construisant la fonction de distribution induite par l’alignement adiabatique d’orbites en résonance interne de Lindblad. Cette fonction de distribution est obtenue en maximisant l’entropie du système compte tenu des contraintes de conservation du nombre d’orbites, du moment angulaire et de l’énergie totale du système, et de la conservation adiabatique détaillée de la circulation de chaque étoile décrivant
son orbite. On montre alors qu’il existe une température critique en deçà de laquelle le système évolue spontanément vers un état barré avec une composante bi-symétrique naissante tournant à la fréquence prédite par l’analyse ci-dessus. Le système peut présenter dans ce cas une chaleur spécifique négative; en considérant l’ensemble des étoiles non résonnantes comme une source thermique à chaleur spécifique positive, on consit qu’il se produise l’équivalent d’une catastrophe gravothermal. La barre s’amplifie alors et son taux de précession décroit pendant que la température diminue encore, en accord avec les résultats des simulations numériques. Il est important de noter que ces résultats peuvent aussi bien rendre compte du devenir d’une barre générée spontanément par instabilité orbitale par exemple, mais aussi expliquer la disparition adiabatique d’une structure barrée qui posséderait une température d’équilibre en deçà de la température critique.

Il convient maintenant de mettre en œuvre ce type de formalisme sur des exemples concrets de galaxies réalistes, afin d’évaluer la pertinence de ce type d’instabilités pour la formation ou la destruction des barres galactiques.

L’état d’équilibre dynamique d’une galaxie est caractérisé par une fonction de distribution qui s’exprime en termes des invariants du mouvement de chaque étoile. Cette fonction présente un double intérêt sur le plan théorique et observationnel. Elle permet de contraindre les distributions en vitesse radiales et azimuthales obtenues par déconvolution des profils de raies d’absorption pour rendre compte de la nature gravitationnelle de l’équilibre. Elle constitue la première pierre d’une analyse linéaire ou numérique telle que celle décrite ci dessus pour étudier la stabilité globale des disques galactiques vis-à-vis de modes non-axissymétriques.

Dans ce domaine, j’ai trouvé une méthode d’inversion très générale qui permet de construire les fonctions de distribution correspondant à une densité surfacique donnée et un profil de température choisi. Ces fonctions de distribution conduisent à des modèles localement stables avec des distributions de vitesse réalistes. L’inversion peut aussi être menée directement à partir de données observationnelles comme la distribution en vitesse azimuthale ou la distribution en énergie spécifique. La mise en œuvre d’une telle méthode d’inversion est prometteuse pour les raisons suivantes: elle constitue une opportunité unique de confronter la distribution en vitesse radiale observée à la même quantité déduite indépendamment via la fonction de distribution, de la distribution en vitesse azimuthale. Les résidus permettent alors de contraindre le modèle et d’évaluer ses faiblesses. Avec cette méthode, il est d’autre part aisé de générer les conditions initiales d’une simulation numérique correspondant exactement d’un point de vue dynamique aux caractéristiques de la galaxie observée. Il serait intéressant de procéder à ce type d’analyse simultanément pour une galaxie S0 et pour une galaxie spirale présentant une structure non axissymétrique. Les propriétés relatives de ces deux types morphologiques et l’existence présumée de modes propres d’instabilité azimuthale dans le second cas consisteraient un lien sans précédent entre la théorie dynamique des disques galactiques et les observations détaillées de leur cinématique rendues possible par la technologie actuelle. En particulier, cette observation permettrait d’apporter des éléments de réponse pour le problème du caractère transitoire ou intrinsèque de la formation des spirales dans les galaxies.

Lors de l’effondrement d’un nuage baryonique de plus $10^6 M_\odot$, la conservation du moment angulaire conduit à son aplatissement. La contraction se poursuit jusqu’à ce qu’une description relativiste du coeur devienne nécessaire. S’il semble établi que la phase ultime de l’effondrement corresponde à un trou noir supermassif, il convient d’étudier plus en détail les étapes conduisant à la formation d’un tel objet. Pour se faire, j’ai construit des familles de solutions analytiques exactes aux équations d’Einstein correspondant à des disques supermassifs minces. Ces solutions sont établis de manière géométrique en identifiant de part et d’autre d’un plan de symétrie les points décrivant deux surfaces de sections traversant un champ du vide donné. Cette identification définit le saut des dérivées normales du champ à partir desquelles les propriétés du disque se déduisent. J’ai étudié la nature des sections conduisant à une solution physique. Pour une métrique statique, j’ai
montré que le champ du vide peut être construit par superposition linéaire de sources fictives de part et d'autre du plan de symétrie en analogie avec la démarche newtonienne. La solution correspondante peut alors être interprétée en termes de deux flots stellaires en contre rotation. Le choix d’une section à courbure extrinsèque nulle conduit à des disques dénués de pression radiale. Tout profil courbe induira au contraire un disque susceptible d’être stable vis à vis des instabilités radiales. En raison de la complexité induite par les forces gravomagnétiques (qui entrainent les repères inertiels), il n’existe pas de solution stationnaire générale pour laquelle la composante non diagonale de la métrique peut être choisie indépendamment. Néanmoins, il est possible de générer par transformation algébrique des familles complètes de métriques stationnaires à partir de solutions statiques connues. On construit de cette manière un disque avec une distribution d’énergie et de pression et une courbe de vitesse qui se déduit du choix de la métrique du vide dans laquelle le disque est plongé. À grande distance, ces disques deviennent newtoniens mais présentent dans leur région centrale des propriétés relativistes (décalage vers le rouge important, écoulement lumineux, ergoregions etc ...). J’ai transposé dans ce contexte la méthode d’inversion newtonienne présentée ci-dessus pour construire les fonctions de distribution compatibles avec un écoulement relativiste donné.

Ce type de solutions présente un intérêt à la fois astrophysique et théorique. La généralisation du formalisme de la dynamique stellaire des disques minces au cadre de la relativité d’Einstein, et la description statistique détaillée de l’écoulement pour un ensemble très général de solutions relativistes constitue un progrès théorique important, en particulier au vu du nombre très restreint de solutions physiques aux équations d’Einstein à symétrie non sphérique. À titre d’exemple, une de ces solutions représente une source physique interne pour la métrique de Kerr qui correspond ultimement au champ généré par un trou noir en rotation. Cette solution interne a fait l’objet d’une recherche continue depuis trente ans. Ces solutions sont plus généralement susceptibles de décrire des amas stellaires relativistes en rotation différentielle localement stables. Les configurations à pression isotrope correspondent éventuellement à des disques supermassifs que l’on peut associer à l’étape d’effondrement final d’un proto-quasar. J’ai calculé la fraction de l’énergie de masse qui est rayonnée lors de l’effondrement de ces objets. Celle-ci peut atteindre dix pour cent de la masse au repos du disque, ce qui suggère que ces derniers ont une durée de vie suffisante pour avoir des conséquences observationnelles.

Il convient maintenant d’étudier la stabilité dynamique globale de ces disques. Il serait aussi souhaitable d’étudier leur stabilité séculaire en ajoutant une description de la viscosité et du transport radial du moment angulaire. Ce type d’instabilité devrait induire un épaississement du disque dont il faudrait tenir compte.
Acknowledgements

It is a great pleasure to thank my supervisor, Donald Lynden Bell, for his enthusiasm and constant availability. It has been an overwhelming experience to learn so much from such an extraordinary person. I am also very grateful to Françoise Combes for her efficient supervision from abroad.

I would like to thank Robert Cannon warmly for his kind friendship and constant help on so many aspects of my life in the last three years, ranging from relativity to the repairing of hoovers. I am very grateful to Jim Collett for his precious advice, physical insight and kindness. Let me thank also Guinevere Kauffman for her enlightening feminine touch and friendship, Rodrigo Ibañez for his curious mind and generous attention, and Martin Haehnelt for his patience and continental solidarity. Thanks to Cooky for countless discussions about “the revolution” and to Ravi Sheth for his peaceful approach and concern. Also worthy of special recognition are David Earn for his positive criticism, Robin Williams for enlightening conversations, Derek Richardson for his skilful help with Unix “en version française”, Andrew Dunn for his refreshing solipsist philosophy, Simon White for his support, advice and understanding in both languages, Isobel Hook for her good mood, Erik Stengler for his sunny accent, Geraint Lewis for the Welsh touch, Angel Lopez-Saínz for Thesis, Mike Hudson and Caleb Scharf for their even tempered approach to being fourth year students of Donald, Martin Becket for all his help with fixing PCs and macs, Blane Little for late night discussions and Dave Syer for introducing me to Mathematica.

My work at the Institute of Astronomy has been greatly facilitated by the kindness and smooth efficiency of Michael Ingham and Judith Moss and by Alice Duncan’s helpful availability.

Special thanks are due to Jiri Bicak for his energetic introduction to the the relativity of counter-rotating disk, and to Martin Rees and Gerry Gilmore who have given me much valuable advice.

Je souhaite aussi remercier Christian Boily pour son soutien de francophone face aux “envahisseurs”, Ian Bonnell et Gisèle pour leur humeur joviale et leur patriotisme québécois plus vrai que nature, Jean-François Sygnet pour son flegme très parisiens, Michel Tagger pour ses conseils de carrière et son approche polémique de la théorie des ondes spirales, et Pierre Encenzaz pour ses encouragements à poursuivre dans cette voie. Un grand merci à Stéphane Colombi pour son language corosif et les journées passées à me convaincre des mérites de la recherche et à Frédérique Dizier pour les heures de téléphone pour me persuader du contraire. Merci aussi à Jonathan pour son entêtement à vouloir reconstruire le monde sur des bases plus saines après chaque repas. et à Laurent dont l’amitié et les longues conversations par courrier electronique m’ont été d’un grand secours. Je tiens a remercier Michel Tallon pour sa patience à m’enseigner quelques rudiments de physique dans la bonne humeur, et Eric Thiebaut dont l’entrain force l’admiration. Un grand merci à Renaud et Françoise pour leur gentillesse et leur hospitalité mainte fois mise à l’épreuve.

It has been a real pleasure to spend time with with Steve, Robert, Ted, Karl, Richard, Kristian, Christopher, Ariane, Jamila, Francine, Patrick, Guy, Juliet, Maryvonne, Fabienne and Cathy. I am grateful to Margaret for making England just like home, Marcus and Debby for their hospitality and kindness, Mark for his brilliant skills on the deck, the pitch, and the pub, and Greg for his enthusiasm for football and the French revolution. From all these people I have learnt a lot about friendship, kindness and sometimes even physics!

My research was generously supported by the Ecole Normale Supérieure de Lyon, the Ministère de la Recherche, the British Council and the Isaac Newton Studentship.

Let me express my gratitude to my parents: they have supported and encouraged me to follow my interests at every stage of my life.

Finally, a very special thank thank to Héloïse, Jean and Brigit for constantly reminding me what real life is all about.

Cambridge
February 1994
References

References

22. CurtisPubl Lick Obs 13, 11- (1918).
43. Hawking, S. Ellis, G. “The large scale structure of space time.” Cambridge Univ. Press.
64. Israel Nuovo-Cimento 33, 331-335 (1964).
References

77. Lynden-Bell, 1973 Topics in the dynamics of stellar systems Geneva.
95. Pichon, C. & Lynden-Bell, D “New sources of Kerr and other metrics: Rapidly rotating relativistic disks with pressure support.” submitted to Phys Rev D.


105. Toomre, A. “What amplifies the spiral?” in The structure and evloution of normal galaxies edited by S.M. Fall and D. Lynden-Bell, Cambridge University Press.


Contents