Bars, Scars & Spirals: The Response of Razor-thin Axisymmetric Stellar Discs to Perturbations

Dominic Dootson

Balliol College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity 2023

Abstract

In this thesis we study the evolution of non-axisymmetric features of Galactic discs. We do this by developing a method to solve the Boltzmann equation, making use of linear response theory and the collisionless assumption of galactic discs. This method is based on Murali (1999) and can be used to efficiently calculate the linear response of a disc to an external perturbation. Where appropriate we will compare our work to other results obtained via linear response theory, while looking to extend previous analyses.

Initially, we study swing-arm amplification without relying on the shearing sheet approximation (Julian & Toomre, 1966). By comparing the density response of the shearing sheet to a complete disc, showing reasonable agreement between the two geometries. Then we show that the amplification of the shearing sheet and full disc is comparable in size, supporting the important role of swing amplification in the growth of spiral structure.

In chapter 4 we study density waves. Initially, we review LSK theory (Lin & Shu, 1966; Kalnajs, 1965) and highlighting the intuition it teaches us. By developing a method to decompose density waves into in-going and out-going components, we finally test the qualitative predictions of LSK theory against a linear theory without the WKB assumption.

In chapter 5 we change tack and calculate the response matrix, and ultimately the modes, of a disc. We compare the spatial frequency to a simple WKB cavity approximation and show that a scarred Mestel disc supports cavity modes, with the scar acting as a reflective boundary. We then show that the cavity mode is comprised of in-going trailing waves and out-going leading waves, agreeing with the swing amplifier feedback cycle Toomre (1977, 1981); Bertin et al. (1989).

We finish by studying the response of a disc to a bar rotating at a constant pattern speed. For weak bars we find strong agreement between our linear method and N-body simulations and successfully predict the late time torque torque. For stronger bars we show that the disagreement between linear and non-linear theory is due to resonantly trapped orbits.

Bars, Scars & Spirals: The Response of Razor-thin Axisymmetric Stellar Discs to Perturbations



Dominic Dootson

Balliol College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2023

For all the childhood dreams that get lost along the way.

Acknowledgements

There are so many people, who have done so much to get me here.

John, it was such a pleasure to work with you. From our first meeting to our last, your humour was infectious and the pleasure of sharing with you what I had learnt was real. I will miss our meetings, when we would both stand, chalk in hand, hashing out the wonders of the universe (periodic cube...) on your overgrown blackboard. You gave me the support I need to become a real physicist, while providing the space for me to grow as a human too. For all this, and much more, I will always be so grateful.

James, thank you for generous support and guidance. Your door was always open to any question, no matter how silly. Your piercing insight, always delivered with quick wit and a clever turn phrase, was a pleasure to behold.

Amery, your support and constant encouragement, especially when everything seemed stagnant, got me through so many tough moments. I loved having a friend in the office who I could share my triumphs with, or moan about silly mistakes to and, most importantly of all, would walk around University parks discussing anything from architecture to philosophy.

Thank you to all the friends I made at Balliol, I am not sure how many Nigel Slater dinner or laps of University parks were shared with you, but I enjoyed every one of them: Tom, for all the walks discussing Adam Curtis and all the pub chats that got very, very deep; Pandora, for all the hours spent in Homebase and being so unashamedly you; Cécile, le meilleur professeur de Français que j'ai jamais eu; Nick, for being the kindest man I have ever met and always being at the horsebox at 4:00 for a walk; Raphael, for welcoming me to Tawney street when I needed to escape and putting up with my bad DJing.

When I arrived at Keble I had so much to learn about myself, yet with the inspiration from many people, I started to take shape: Bill, for being a comrade in arms during those years at Keble, and never failing to inspire me with your friendship. Evie, for making it to more astrophysics lectures than law lectures; Alex, for all the long chats with the best teddy bear; Svea, for making me learn how to cook; Cabbage, for teaching me to dance as if nobody's looking.

I would not be here, writing this, if it were not for some fantastic teachers who saw things that I could never see for myself: Jane, for letting me mark your

homework; Paul, for always having books I could borrow; Martin, for showing me how beautiful mathematics is; Steve, for listening when a twelve year old said he wanted to be a physicist and helping him achieve his dream.

Ali, for showing me all I have to stand up tall about.

Rory, watching you find your wings and fly has brought me so much joy and pride. You are a wonderful human being, who I am so lucky to call my twin. We have come so far together, yet I cannot wait to see where our adventures take us next.

Mum & Dad, thank you so much for all you've done in getting me here today. You are the most caring people I know and I am so lucky that I get to call you Mum & Dad. You are kind, you are clever and you are truly inspirational.

Abstract

In this thesis we study the evolution of non-axisymmetric features of Galactic discs. We do this by developing a method to solve the Boltzmann equation, making use of linear response theory and the collisionless assumption of galactic discs. This method is based on Murali (1999) and can be used to efficiently calculate the linear response of a disc to an external perturbation. Where appropriate we will compare our work to other results obtained via linear response theory, while looking to extend previous analyses.

Initially, we study swing-arm amplification without relying on the shearing sheet approximation (Julian & Toomre, 1966). By comparing the density response of the shearing sheet to a complete disc, showing reasonable agreement between the two geometries. Then we show that the amplification of the shearing sheet and full disc is comparable in size, supporting the important role of swing amplification in the growth of spiral structure.

In chapter 4 we study density waves. Initially, we review LSK theory (Lin & Shu, 1966; Kalnajs, 1965) and highlighting the intuition it teaches us. By developing a method to decompose density waves into in-going and out-going components, we finally test the qualitative predictions of LSK theory against a linear theory without the WKB assumption.

In chapter 5 we change tack and calculate the response matrix, and ultimately the modes, of a disc. We compare the spatial frequency to a simple WKB cavity approximation and show that a scarred Mestel disc supports cavity modes, with the scar acting as a reflective boundary. We then show that the cavity mode is comprised of in-going trailing waves and out-going leading waves, agreeing with the swing amplifier feedback cycle Toomre (1977, 1981); Bertin et al. (1989).

We finish by studying the response of a disc to a bar rotating at a constant pattern speed. For weak bars we find strong agreement between our linear method and N-body simulations and successfully predict the late time torque torque. For stronger bars we show that the disagreement between linear and non-linear theory is due to resonantly trapped orbits.

Contents

List of Figures **xi**

1 Introduction **1**
- 1.1 The Sky at Night . . . 1
- 1.2 Our Home – Components of the Milky Way . . . 2
- 1.3 Galactic Dynamics . . . 4
- 1.4 Bars, Scars & Spirals: An Overview of This Thesis . . . 11

2 Dynamical Modelling **15**
- 2.1 Introduction . . . 15
- 2.2 Dynamical Evolution . . . 19
- 2.3 Application to razor-thin axisymmetric discs . . . 23
- 2.4 Green Function of the Linear CBE . . . 33
- 2.5 Softening due to Potential Density Pair Expansion . . . 34
- 2.6 Conclusion . . . 36

3 Spirals **39**
- 3.1 Introduction . . . 39
- 3.2 The Shearing Sheet & JT Equation . . . 41
- 3.3 Comparison between Shearing Sheet & Linear Response Theory . . . 55
- 3.4 The Strength of Swing Amplification in Discs . . . 65
- 3.5 Conclusion . . . 72

4 Density Waves **73**
- 4.1 Introduction . . . 73
- 4.2 WKB Waves . . . 74
- 4.3 Kinematics of Waves in Cylindrical Coordinates . . . 79
- 4.4 Test Application of a Logarithmic Spiral . . . 86
- 4.5 Putting LSK to the Test . . . 90
- 4.6 Conclusion . . . 93

5 Scars **95**
5.1 Introduction 95
5.2 Modes of the Tapered Mestel Disc 97
5.3 Modes of Scarred Discs - Angular Momentum Scars 103
5.4 Conclusions 112

6 Bars **115**
6.1 Introduction 115
6.2 Bar Models 116
6.3 Linear Torque Calculation 119
6.4 Limitations Of Linear Theory 124
6.5 Conclusion 128

7 Conclusions **131**
7.1 What Just Happened? 131
7.2 What is Left to Happen? 134

Appendices

A Calculating Basis Functions via Green Function **139**

B Details of test-particle and N-body integrator **143**
B.1 Sampling the DF 143
B.2 N-Body Evolution 145

C Response Matrix Via Direct Integration **149**

Bibliography **153**

List of Figures

1.1 Orbit of a stars at ILR, CR and OLR 10

2.1 Density profile of the tapered Mestel disc. 23
2.2 The radial Kalnajs basis functions Kalnajs (1976). 25
2.3 Density response using two different basis functions. 31
2.4 The quadrupole harmonic of the kernel's Greeens function. 33
2.5 Interaction potential for $\ell = 2$ Kalnajs basis functions. 35

3.1 The swing-arm amplification mechanism (Julian & Toomre, 1966) . 40
3.2 A schematic diagram for the shearing sheet. 45
3.3 Evolution of the density response for different initial windings in the shearing sheet. 53
3.4 Amplification of the shearing sheet for different temperature discs. . 54
3.5 The axisymmetric shearing sheet and full disc density response to a perturbation 59
3.6 The non-axisymmetric shearing sheet and full disc density response to a perturbation 62
3.7 Amplification in the cold disc limit. 67
3.8 Amplification of the shearing sheet and full disc. 68
3.9 The full density response of a disc to a perturbation. 69
3.10 The effect of self-gravity in amplification for the shearing sheet and full disc. 71

4.1 LSK dispersion relation (Lin & Shu, 1966; Kalnajs, 1965). 75
4.2 Motion of an WKB wave. 77
4.3 Reconstruction of an unwinding logarithmic spiral. 87
4.4 The response to a logarithmic spiral decomposed into in-going and out-going waves. 89
4.5 Density waves decomposed into in-going and out-going waves 91

5.1 The response matrix for different mode frequencies. 100
5.2 Real space density representation for an unstable tapered disc. . . . 101
5.3 The response of coefficients in an unstable disc. 102

5.4 The modes of a tapered Mestel disc with different active fractions. . 104
5.5 Comparison of spatial frequencies for cavity modes calculated via the response matrix and WKB approximation. 106
5.6 The real space density representation of cavity modes with different scar radii. 108
5.7 The density profile of modes with different scar radius. 109
5.8 Decomposition of a cavity mode into in-going and out-going density waves. 110

6.1 Different radial bar models (Sormani et al., 2015). 117
6.2 The evolution of the torque acting on a bar to different temperature disc. 120
6.3 The evolution of coefficients due to a steadily rotating bar. 121
6.4 The torque acting on bars of different strengths. 124
6.5 Surface of section for orbits trapped at ILR and CR. 126
6.6 Phase space snapshots of an N-body simulation showing non-linear effects. 127

A.1 Gaussian basis functions 140

We fall in love for a smile, a look, a shoulder. That is enough; then in the long hours of hope or sorrow, we fabricate a person, we compose a character.

– M. Proust, *The Fugitive*

1 Introduction

Contents

1.1 The Sky at Night 1
1.2 Our Home – Components of the Milky Way 2
1.3 Galactic Dynamics 4
1.3.1 The Collisionless Approximation & CBE 4
1.3.2 N-Body Methods & their Discontents 5
1.3.3 Equilibria of Galaxies 6
1.3.4 Perturbative Approaches 8
1.3.5 Wave Mechanics in Discs 9
1.4 Bars, Scars & Spirals: An Overview of This Thesis . . 11

1.1 The Sky at Night

The sky at night is a constant source of wonder and amazement; looking upon the heavens offers a moment in which to leave the chaos within oneself behind (Nietzsche, 2005)[1]. Those outside long enough may start to question how the stars, nebulae and galaxies formed: many will only spend seconds pondering such a question, before returning to the warmth inside from a clear, dark night; some will spend much longer – maybe years wrestling with these questions; a few, very occasionally, will spend a lifetime. Those that spend time in pursuit of a theory of galaxy formation will

[1] Some would consider it a cliché to start like this, the author has no such concerns.

find many questions in need of an answer, yet two underpin the whole endeavour: where is all the 'stuff' – be it stars, gas, dust, black holes, dark matter ... – and how do they dance together in these entities astronomers call galaxies? The field concerned with answering this final question is galactic dynamics.

Individual galaxies, and the constituent parts that compose them, are not just intrinsically interesting, they also offer an environment in which to test the fundamental laws of physics: they are the habitat in which we find dark matter, a type of exotic particle which constitutes roughly a quarter of our universe and that has never been observed in the laboratory; they are the machines that turn gas into stars and planets, a crucial part of understanding our own planet and sun; their structure is imprinted with their formation history, offering clues to the processes at play in the early universe.

This thesis aims to contribute to our understanding of the processes that drive galactic dynamics. We will develop new techniques to solve the underlying equations and to study the processes that form and drive the evolution of non-axisymmetric features in discs.

1.2 Our Home – Components of the Milky Way

Our Galaxy, the Milky Way, is made up of many constituent parts. Before we discuss the methods used in the field of galactic dynamics, we will first outline the component parts of galaxies which astrophysicists choose to study.

The dark matter Halo is the largest both in extent, with an approximate viral radius of 200 Kpc (Dehnen et al., 2006), and mass, approximately one hundred times more massive than the disc (Binney & Tremaine, 2008; Posti & Helmi, 2019). Due to the dark nature of the particles constraining the size and shape of the Halo is a challenging task. However luminous objects at large radii such as globular clusters (e.g. Posti & Helmi, 2019) and stellar streams (Varghese et al., 2011, e.g.) can be used to probe its potential. Due to its large mass, the Halo plays an important role in the dynamics of the Galactic disc. In particular the first evidence of the existence of the Halo came from the flat rotation curve of stars at large radii (e.g.

Kent, 1987). The dynamical history of the Galactic disc can also probe the physical properties of dark matter (for a recent review see Mayer (2022)).

At the centre of the Halo is a supermassive black hole, Sgr A*. The sun is at a distance of 8.2 kpc from Sgr A* (The GRAVITY Collaboration, 2019), moving with a proper motion of 245 km s^{-1}(Reid & Brunthaler, 2004).

In the middle of the Halo is the Galactic disc of the Milk Way. It is composed of two discs (Binney & Vasiliev, 2022), both distinct in chemical make up and structure, superimposed on top of one another: the thin and thick disc. The thin disc plays host to the Sun and has a radial scale length of order 2.5 kpc and vertical scale height of order 300 pc (Binney & Tremaine, 2008). The thick disc, which was first observed by Yoshii (1982) and Gilmore & Reid (1983), has a much larger vertical scale height of order 1 kpc, hence the designation 'thick'. The stars of the thin disc tend to be younger with a distinct chemical composition – they have higher metallicities, but lower abundance of alpha process elements compared to iron, [α/Fe], (e.g. Fuhrmann, 1998; Bensby et al., 2007; Adibekyan et al., 2013), suggesting that stars in the thin disc formed later than in the thick disc. The thick disc is also the junior partner, with approximately 1-12% the total stellar budget (Gilmore & Reid, 1983; Siegel et al., 2002; Jurić et al., 2008) in the stellar neighbourhood, although recent work is questioning this assumption (Pouliasis et al., 2017). The evidence suggests that a nascent thin disc produced the thick disc when it was shaken by an encounter with a smaller galaxy in its youth (e.g. Samland & Gerhard, 2003).

As well as these axisymmetric features, the Milky Way also contains non-axiysmmetric features, such as a bar and spiral arms. Bars are elongated structures that rotate with a uniform pattern speed, found in a wide variety of galaxies. By aiding the motion of particles and gas to the centre of the disc, bars can play a crucial role in the evolution of galaxies. The presence of a bar in our own Galaxy was first suggested by de Vaucouleurs (1964), but its existence remained an open question for many years. However recent evidence has confirmed it presence using three different approaches: gas kinematics at the centre of the Galaxy (e.g Binney et al., 1991; Sormani et al., 2015); the infra-red luminosity at the centre of the

Galaxy (e.g Blitz & Spergel, 1991; Weiland et al., 1994); distributional asymmetries of red giants (e.g. Nakada et al., 1991; Weinberg, 1992). From N-body simulations it has long been known that formation of a bar is inevitable in discs that are unstable to non-symmetric modes (e.g. Hohl, 1971; Athanassoula et al., 2013), but the details are still to be completely understood.

Spiral arms are another important non-axisymmetric feature of our Galaxy. Grand design spirals are arguably some of the most impressive structures in galactic physics and can be found in many galaxies (Lintott et al., 2011). Although easy to identify in other galaxies, they are much harder to observe in our own and must be observed indirectly: distance measurements of OB stars and other photometric star counts (e.g. Morgan et al., 1953; Binney et al., 1997; Benjamin et al., 2005); distance measurements of MASERS (e.g. Reid et al., 2014, 2019); measurements of H1 emission (e.g Van de Hulst et al., 1954; Levine et al., 2006); dynamics of red giants (e.g. Eilers et al., 2020); and observation of interstellar dust (e.g. Kh et al., 2018). Despite their ubiquitous occurrence in disc galaxies, the physics that drives the formation of these structures is little understood, with even Gaia as yet failing to give conclusive proof about the mechanism (e.g. Sellwood et al., 2019).

1.3 Galactic Dynamics

1.3.1 The Collisionless Approximation & CBE

Galaxies are dynamically young objects: in orbital years the Sun is roughly twenty years old – only just out of its teenage years. On these short timescales two body-interactions are not important, rather it is the evolution in the mean field, $\Phi(\mathbf{x}, t)$, that drives the evolution of galaxies. The mean field, in which the particles[2] distinction between stars and dark matter that constitute our Galaxy evolve, is sourced via the density distribution of particles whose motion we wish to study. As such, galactic dynamics is concerned with the evolution of the galaxy's distribution function (DF), $F(\mathbf{x}, \mathbf{v}, t)$, which is a phase space mass density, which can either

[2]In this thesis we will ignore gas and not make a

represent a single or multiple particle species. Once the DF has been obtained, the the potential can be calculated via Poisson's equation

$$\nabla^2\Phi(\mathbf{x},t) = 4\pi G \int F(\mathbf{x},\mathbf{v},t)d^3\mathbf{v}. \tag{1.1}$$

If we need the DF to calculate the potential, we need an equation that governs the evolution of the DF in the mean field. On such short timescales, which correspond to $\sim 10^8$ years for our Galaxy, it is sensible to assume that stars are conserved. This is *not* the case: stars are constantly being born and die due to their own stellar processes and very occasionally destroyed via collisions with other objects. Indeed, in this thesis we will assume conservation of stars. From the continuity equation we arrive at the collisionless Boltzmann equation (CBE)

$$\frac{dF(\mathbf{x},\mathbf{v},t)}{dt} = 0. \tag{1.2}$$

Of course, the dynamics of galaxies over secular timescales is an active and interesting area of research (e.g. Heyvaerts (2010), Chavanis (2012), Fouvry et al. (2015b)).

1.3.2 N-Body Methods & their Discontents

Given that we now have an equation that governs the evolution of our galaxy under the collisionless approximation, we now must solve it. The most powerful way of doing so is N-body simulations. The development of 'off the shelf' N-body general purpose codes e.g. (e.g. Springel, 2005) that are optimised for large simulations have allowed researchers to study the evolution of systems in ever more detail. Or alternative bespoke N-body codes, which are designed with a particular application in mind, such as those used by Sellwood in Sellwood (2010, 2012); Sellwood & Carlberg (2014) can be used to shed light on specific processes. At its heart, collisionless N-body simulations are a Monte-Carlo approach to solving the CBE via the method of characteristics. There are two general approaches to N-body codes: 'honest' N-body codes, where each particle in the simulation represents an individual object, such as a star or black hole, e.g. NBODY6 (Wang et al., 2015); or collisionless N-body codes, where each 'particle' represents samples of

the characteristic curves for the CBE. Even as computers become more powerful, allowing for an ever increasing number of particles, there is still one fundamental limitation with both of these N-body techniques: Poisson noise. Indeed, due to the finite N, there is a potentially perilous problem: due to the finite number of particles, Poisson statistics will generate fluctuations in the distribution that will not be removed by averaging over many realisations (e.g. Sellwood, 2012).

N-body simulations showed that galaxies undergo an initial period of violent relaxation (Lynden-Bell, 1967) and phase mixing at the beginning of their formation, leading to a quasi-stationary state within a few dynamical times. This quasi-stationary state evolves on a secular timescale, much longer that the dynamical timescale of galaxies. We could consider this state to be a constant equilibrium, around which perturbations evolve in its mean field. The method of perturbation particles (Leeuwin et al., 1993), is a monte-carlo method designed to exploit this observation. In these simulations the leading order potential is due to the equilibrium model. To this potential a first order correction, that is sourced by density fluctuations, is added, making the effects of Poisson noise a second order effect. Even with this improvement the simulation would still be subject to the same, albeit smaller, errors. These errors occur due to finite N effects in N-body codes, to fix this problem could we just solve the CBE directly?

The direct numerical integration of the CBE is a numerically difficult task due to the ever finer structure generated by phase mixing and chaotic mixing (Mangeney et al., 2002; Colombi et al., 2015) and the high dimensionality of the problem. Instead of solving the CBE directly, could we use perturbation theory to study the evolution of a perturbation to some well understood equilibrium? If we are to do this, what exactly is a 'well understood equilibrium?'.

1.3.3 Equilibria of Galaxies

There is a large zoo of equilibrium models of self-gravitating systems, each capturing different structure or phenomena that are the focus of different areas of study (e.g. Mestel, 1963; King, 1966; Navarro et al., 1996).

We want our equilibrium model of our galaxy to be stationary. The easiest way to ensure this is via Jean's theorem (Jeans, 1915), which states, "Any steady-state solution of the collisionless Boltzmann equation depends on the phase-space coordinates only through integrals of motion in the given potential, and any function of the integrals yields a steady-state solution of the collisionless Boltzmann equation" (pp 283 Binney & Tremaine, 2008). As such, a steady-state equilibrium DF can be constructed by ensuring that it is a function of integrals of motion, a function that is only conserved along a particles orbit.

If we choose our integrals of motion such that their conjugate variable is 2π-periodic and that they form a canonical coordinate system the integrals are referred to as 'actions', $\mathbf{J}$, while their conjugate positions are 'angles' $\boldsymbol{\theta}$ (Arnold, 1987). By Hamilton's equation, the evolution of the angles is trivial, $\boldsymbol{\theta}(t) \propto t$, vastly simplifying the CBE.

There has been much work constructing self-consistent models, where the density profile sources a potential (e.g. §4 Binney & Tremaine, 2008). A recent example comes from one use of Vasiliev (2019)'s software, AGAMA, that combines different components into one self-consistent model, allowing the study of complex multi-component systems. This modelling approach, along with the unparalleled view of our galaxy give by the Gaia Galactic survey (Gaia Collaboration, 2018), has led to the most complete, albeit still axisymmetric, models of our galaxy ever constructed (Binney & Vasiliev, 2022).

Another key feature of many equilibrium models is that they must be dynamically stable, such that any infinitesimally small perturbation will damp away and leave the system unchanged. The question 'is a model unstable?', can be studied using N-body methods, but the discretization of the DF can obfuscate the answer, e.g. too low N could lead to a stable disc becoming Toomre unstable (Toomre, 1964). A more prudent way to answer this important question is to use perturbative methods.

Although most of the models of interest are stable, there are many physical phenomena that are due to instabilities, for example: secular diffusion can form grooves and ridges in the DF which lead to spiral arms forming instabilities (Sellwood,

2010, 2012; Fouvry et al., 2015b; De Rijcke et al., 2019a) and ultimately bar formation (Hohl, 1971; Athanassoula et al., 2013). It is also worth highlighting that linear instabilities will be brought to heel by non-linear phenomena: the linear growth of the spiral in (Sellwood & Carlberg, 2014) is curtailed by non-linear phenomena (Hamilton, 2023).

With the development of more accurate equilibria models, how do perturbations to these models evolve? To answer this we can use perturbative methods.

1.3.4 Perturbative Approaches

There are two important issues to address when studying the linear response of a disc: what coordinate system to express the CBE in and how, given the coordinate system, to solve Poisson's equation for the potential that the perturbation sources. It makes most sense to solve the CBE in real space coordinates $(\mathbf{x}, \mathbf{v})$, as this is the natural coordinate system for Poisson's equation. However, as galaxies are inhomogenous, in this coordinate system the motion of particles follow complex orbits. To simplify this problem, we could express the orbits in a more natural coordinate system, such as action angle variables, however this does drastically complicate solving Poisson's equation. These are important and technical considerations, which we will return to in section 2.1.

Another important decision, the implications of which we will discuss in detail in section 2.1, is whether to solve the CBE via Laplace transform or in the time domain. Laplace transforming will result in a dispersion relation for the equilibrium model of the disc, perfect for studying the stability of a disc. However, inverting the Laplace transform back to the time domain is difficult and so it can be hard to obtain the time dependent evolution of a disc. If we are interested in the time dependent response of the disc to a perturbation, then why not solve the linearised CBE in the explicit time domain? In chapter 2 we will develop a method that solves the linearised CBE in the time domain, that we will then use to study wave mechanics in disc during the rest of this thesis.

Before developing a novel method to study the linear dynamics of discs, we first review some key pieces in our current understanding of disc dynamics.

1.3.5 Wave Mechanics in Discs

Density Waves

It was first noted by Lin & Shu (1966), that density perturbations to a disc propagate in a similar way to waves across the surface of a pond. The 'density waves' are composed of periodic compressions and rarefactions of the equilibrium model's surface density. Those waves that wrap in the opposite (same) direction to the rotation of the disc are leading (trailing). One way to obtain a dispersion relationship for these density waves is to assume that the waves are tightly wound, i.e. that their radial wavelength is much less than their radius. This leads to the Lin-Shu-Kalnajs (LSK) dispersion relationship (Lin & Shu (1966), Kalnajs (1965)) which is often used to gain intuition about the motion of density waves. In particular, LSK predicts that an initially tightly wound leading wave, will unwind. As it does so it will transition from leading to trailing, during which it will violate the tight winding approximation. Once trailing LSK predicts that it will continue to wind up, getting tighter and tighter. It also predicts that the ILR and OLR are barriers, through which density waves cannot pass (Binney & Tremaine, 2008). In chapter 4, we will study the motion of density waves, and compare linear theory to that of LSK.

Epicycle Approximation and Resonances

In dynamically cold discs the radial motion of stars are small. For stars that are on nearly circular orbits in a disc, the effective potential in which they orbit can be approximated by a quadratic term. Under this assumption, the angular velocity, $\Omega(R_{\rm g})$, of the star's guiding centre is constant, while the star undergoes harmonic motion about the guiding centre, $R_{\rm g}$, with the epicycle frequency, $\kappa(R_{\rm g})$. In more explicit language, this means that the radius and angular velocity of the star, which is given via conservation of angular momentum, are

$$\begin{aligned} R(t) &= R_{\rm g} + A\cos(\kappa t), \\ v_\phi(t)R(t) &= L_z = \text{constant}, \end{aligned} \tag{1.3}$$

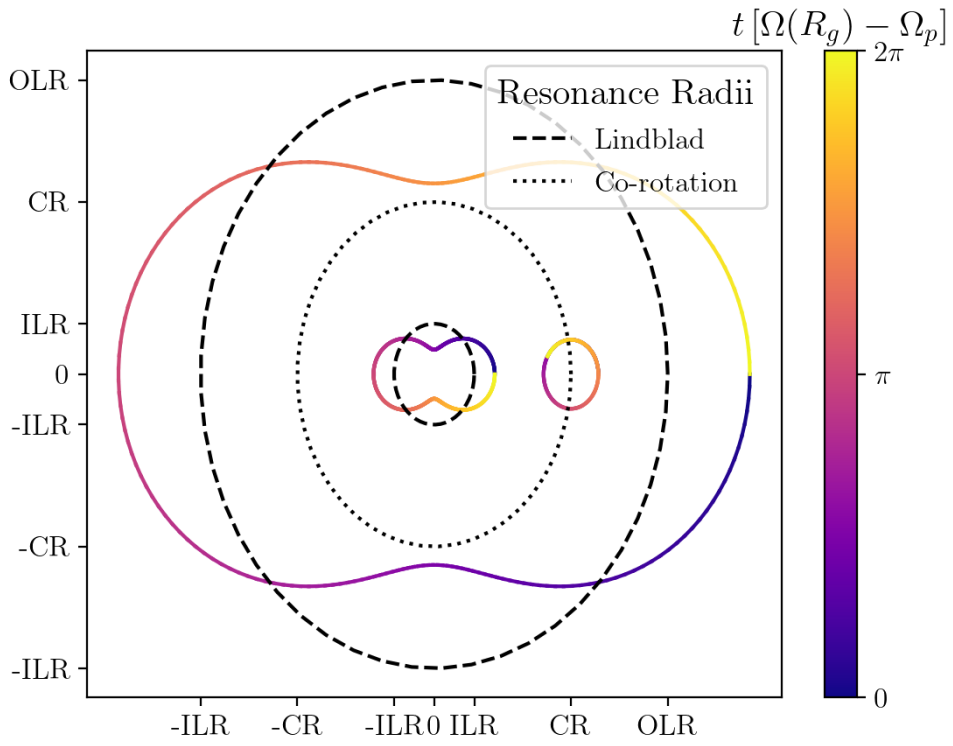


Figure 1.1: The orbit of stars in the rotating frame of an $\ell = 2$ perturbation, at ILR (inner circle), CR (middle circle) and OLR (outer circle). The colour of each orbit is proportional to the time, normalised by the time period of the orbit in the co-rotating frame. The star at CR just moves radially, while those at ILR and OLR circulate in this frame. The star at the ILR orbits faster than the perturbation and hence moves in an anti-clockwise direction, while the star at OLR orbits slower and moves in an clockwise direction. Any non resonant orbits would not be closed in this frame, and would, after enough time, visit every point on an annulus centred at $R_{\rm g}$.

where A sets the size of the epicycle motion, L_z is the conserved angular momentum of the star and we have ignored initial conditions.

If there is another frequency set by a perturbation to the disc, such as the pattern speed of a bar, the orbit can resonate with the perturbation. The most common of these resonances are:

- *Co-rotation resonance (CR)*: the star's guiding centre frequency is the same as the pattern speed of the perturbation, $\Omega_{\rm p}$, of the perturbation, i.e. $\Omega(R_{\rm g}) = \Omega_{\rm p}$.
- *Lindblad* resonance: the star's epicyclic frequency coincides with the frequency at which it experiences peaks in strength of the perturbation, i.e.

$\ell\,[\Omega(R_{\rm g}) - \Omega_{\rm p}] = \pm\kappa(R_{\rm g})$, where ℓ is the Fourier harmonic of the perturbation. The + corresponds to the inner Lindblad resonance (ILR), while the − corresponds to the outer Lindblad resonace (OLR)

In Figure 1.1, we plot the orbit of stars in the rotating frame of an $\ell = 2$ perturbation, at ILR (inner circle), CR (middle circle) and OLR (outer circle). The colour of each orbit is proportional to the time, normalised by the time period of the orbit in the co-rotating frame. A star at CR just orbits a point, while those at ILR and OLR circulate in this frame. The star at the ILR orbits faster than the perturbation and hence moves in an anti-clockwise direction, while the star at OLR orbits slower and moves in an clockwise direction. However both the stars orbiting at ILR and OLR have their apocentre every half orbit in this frame, consistent with the ILR and OLR for a quadrupole perturbation. Any non resonant orbits would not be closed in this frame, and would, after enough time, visit every point on an annulus centred at $R_{\rm g}$ with width $2A$.

Shearing Sheet

In order to study the motion of a density wave during its transition from leading to trailing, Julian & Toomre (1966) studied a local patch of a differential rotating disc – the so-called 'shearing sheet' (c.f. Figure 3.2). By studying the time dependent response of the shearing sheet to a small perturbation Julian & Toomre (1966) showed that as a wave passed from leading to trailing it can undergo large amplification. This process, swing amplification, is an important part in our understanding of the formation of spiral arms. In chapter 3, we will study this process without relying on the shearing sheet approximation.

1.4 Bars, Scars & Spirals: An Overview of This Thesis

In this thesis we will study a variety of phenomena that all play a part in the formation of spiral structure in galaxies. In order to do so we will develop a linear

response machinery that will allow us to study the response of a razor-thin disc to perturbations in a self-consistent manner.

In the next chapter we outline our novel linear response method, where the disc's density response is recast as a Volterra integral equation. To simplify the equations of motion we choose to solve the CBE in action-angle coordinates. This, however, introduces a different problem: the real space Laplacian in these coordinates is non-local, making Poisson's equation much harder to solve. To resolve this issue we introduce potential-density pairs (Kalnajs, 1976). We then depart from previous work and apply our general linear response calculation to a two dimensional razor-thin disc, rather than a spherical halo (e.g. Murali, 1999; Rozier et al., 2022). We then discuss how the kernel of the integral equation – and in particular the action-angle representation of the potential-density pairs – is calculated. To illustrate how these equations are used we calculate the density response to a perturbation using two different potential density pairs. We finish with a discussion about the Green function for the CBE and the softening that is inherent to the potential density pair method.

In chapter 3 the linear response machinery is applied to the study of swing amplification (Julian & Toomre, 1966; Binney, 2020). Since the publication of Julian & Toomre (1966) swing amplification has been a crucial piece of the puzzle in understanding the formation of large spiral structure, yet the assumptions on which this seminal work is based – the so-called shearing sheet approximation – are not always physically justifiable. Due to the general nature of our linear response machinery, the same problem can be studied without the need for such a limiting assumption. In order to ultimately compare amplification in a disc to that in the shearing sheet, we will start by reviewing the shearing sheet. We will then look to the regimes in which these assumptions fail and use the linear response machinery to calculate the power of swing amplification in a full disc.

In chapter 4 we test the LSK dispersion relationship (Lin & Shu, 1966; Kalnajs, 1965), which rests on a questionable WKB approximation, that is a key part of our understanding of density waves. In order to make comparisons to the LSK

predictions we must be able to decompose density waves into in-going and out-going parts. Although conceptually simple (we need only project the response onto the eigenfunctions of the wave equation in cylindrical coordinates) it is a computationally hard task as the radial eigenfunctions diverge at the origin. This complication is overcome by instead projecting onto the far field limit of these basis functions. We then compare the corresponding in-going and out-going waves packets to that predicted by WKB theory (Toomre, 1969; Mark, 1974, 1977).

After three chapters of resisting calculating modes, we finally succumb to the pressure and in chapter 5 do just that. We start by showing the relationship between the time dependent and frequency dependent response. There are, however, subtle limitations to this method that we discuss in detail. To validate our approach we calculate the modes of the unstable Mestel disc (Mestel, 1963) and compare to previous work (e.g. Evans & Read, 1998). We confirm the presence of 'cavity' modes in scarred discs by comparing the eigenmodes to a naive WKB approximation (Mark, 1977). The findings of De Rijcke et al. (2019a) are corroborated, while showing the importance of accurate density models.

In our final chapter, before we conclude we calculate the torque that a disc imparts on the bar that it hosts. We start by comparing the response for a weak bar, which is in the linear regime, and the linear response calculated using the methods of chapter 2. We show good agreement between the two approaches and calculate the steady-state torque that the bar imparts on the disc using the approach of Lynden-Bell & Kalnajs (1972). We then show the limitations of linear theory as we increase the strength of the bar, introducing non-linear effects. Using N-body simulations, we show that this is due to resonant trapping.

In the final chapter we conclude and highlight possible areas for improvement. We will discuss any possible future avenues down which this work, and in particular the linear response machinery, might be taken.

What follows ... is further an attempt to follow out an idea consistently, out of curiosity of where it might lead.

– S. Freud, *Beyond the Pleasure Principle*

2 Dynamical Modelling

Contents

2.1 Introduction 15
2.2 Dynamical Evolution 19
2.3 Application to razor-thin axisymmetric discs 23
2.3.1 Potential-Density Pairs in 2D 24
2.3.2 Angle–action variables for 2d axisymmetric discs 28
2.3.3 Computing the evolution Kernel 29
2.3.4 Comparison of the effects of basis choice 32
2.4 Green Function of the Linear CBE 33
2.5 Softening due to Potential Density Pair Expansion . . 34
2.6 Conclusion 36

2.1 Introduction

An important part of our understanding of self-gravitating systems comes from the study of the Boltzmann equation. For systems such as stellar discs, on dynamical times shorter than the relaxation time from two body encounters, the collisonal term in the Boltzmann equation is negligible and can be dropped, resulting in the collisonless Boltzmann equation (CBE).

There are many different approaches to solve the CBE, the most common being N-body methods. However, as discussed in chapter 1, this approach has

many limitations.

Techniques for studying the solution of the linearised CBE are less well developed than their N-body counterparts. In particular, one foundational complication is due to the inhomogeneous nature of galaxies, resulting in stars moving along complex orbits, with associated unpleasant equations of motion. There are two possible approaches to dealing with these complex equations of motion: to introduce simplifying assumptions that result in 'simple' equations of motions or to solve the CBE in action-angle coordinates, $(\boldsymbol{\theta}, \mathbf{J})$ (Arnold, 1987), a coordinate system in which stars move in straight line.

The most common of these assumptions is the 'epicycle' approximation (Binney & Tremaine, 2008). If a disc is dynamically cold, then its response can be studied by making use of this approximation (e.g. Fouvry et al., 2015a). Another approach to obtain simplified equations of motion is to simplify the geometry of the problem, such as in the case of the periodic cube Magorrian (2021) or the shearing sheet (Julian & Toomre, 1966; Binney, 2020). The potential sourced by the disc can then be found by solving Poisson's equation using standard techniques in real space coordinates, e.g. by Fourier transform. Which can be used to solve the CBE.

Despite how hard one might try, simplified equations of motion might not be possible in real space coordinates (without making physically unjustifiable assumptions). Although a common issue to all methods, this is more problematic in analytic methods as orbits cannot always be numerically integrated. A common approach is to introduce action-angle variables (Arnold, 1987), however this will introduce a complication further down the line – what is the mapping between action-angle variables and real space coordinates? There is, however, another more pressing question with working in action-angle variables: the CBE is closed under Poisson's equation, as the stars' potential ultimately governs their evolution, which is most naturally solved in real space coordinates; in action-angle variables the Laplacian operator becomes non-local, making Poisson's equation much harder to solve.

In order to overcome this complication Kalnajs (1976) introduced potential-density pairs: a set of basis functions such that a given density element, $\rho^p(\mathbf{x})$, sourced a corresponding potential element, $\psi^p(\mathbf{x})$.

Now that the linearised CBE is in a coordinate system in which the equations of motion can be easily solved and that Poisson's equation can be solved in such a system, we now must solve the linearised CBE. Depending on the phenomena we want to study, we are left with one final choice: to Laplace transform leading to a modal analysis or not to Laplace transform, i.e. staying in the time domain? Doing so will result in a dispersion relation whereas not doing so will allow the initial value problem to be more easily studied.

Modal Analysis

A standard approach to solving the linearised CBE is to Laplace transform in time with the resulting modal equation acting as a dispersion equation for the disc. This equation can be solved to easily ascertain the stability of a self-gravitating system, in fact Kalnajs (1976) introduced potential-density pairs so that the stability of the Mestel disc could be studied (Kalnajs, 1977). Due to the importance of this question in the hunt to understand galaxy models and the strengths of this method over N-body in assessing the stability of a system, lots of work has been conducted using these techniques (Weinberg, 1991; Vauterin & Dejonghe, 1996; Pichon & Cannon, 1997; Evans & Read, 1998; Jalali & Hunter, 2005; Khoperskov et al., 2007; Fridman & Polyachenko, 2012; Polyachenko & Just, 2015; De Rijcke et al., 2019a).

From the Laplace approach one can in principle calculate the response of the disc in the explicit time domain, doing this, however, is a complicated task: first, all modes – both stable and unstable – must be found. As calculating numerically stable modes is computationally difficult, this is a hard task in the common case where the disc is stable. Second, when inverting the Laplace transform the residues of the poles, and hence each mode's contribution to the response, must be found. If the response is required in the time domain it is more prudent to solve the initial value problem.

Direct Time Evolution

Instead of Laplace transforming the linearised CBE, the differential equation can be recast as an integral equation. This is also the most natural way to study how a perturbation, such as an orbiting galaxy, with an explicit, possibly complex time dependence perturbs a gravitating system. There have been lots of different studies using these methods, however many rely on different assumptions: when studying swing-arm amplification the shearing sheet approximation is used (Julian & Toomre, 1966; Binney, 2020); in order to analytically calculate the time integral an explicit form (which can be integrated analytically) is assumed for the external perturbation (e.g. Banik & van den Bosch, 2021; Al Kazwini et al., 2022), ruling out the study of how a disc's response might affect a perturbation, e.g. an orbiting satellite whose wake affects its orbit; or to simply ignore the self-consistent response all together (e.g. Colpi & Pallavicini, 1998; Monari et al., 2016).

Murali (1999) was the first approach that studied the response of a spherical halo while making use of action-angle variables, potential-density pairs and not limiting the time evolution of the perturbation. The approach was left relatively untouched until Rozier et al. (2022) used it to study the response the Milky Way's Halo to the Large Magellanic Cloud.

Taking inspiration from Murali (1999), this chapter will develop a new method to study the response of a 2D razor-thin disc to an external perturbation in a self-consistent manner. This method will then be harnessed to help elucidate the physical phenomena that drive our Galaxy's dynamics, with particular application to galactic spiral structure, in the following chapters. In the following section we will recast the linearised CBE as an integral Volterra equation that describes the evolution of a self-gravitating system to an external perturbation, making use of potential density pairs (Kalnajs, 1976). In section 2.3 we will specialise our analysis to a razor-thin axisymmetric disc and construct the building blocks need to solve the Volterra equation. We will finish the chapter by calculating the Green function for the linear CBE for a disc and studying the softening inherent to the truncated potential density pair expansion.

2.2 Dynamical Evolution

Linearization of the CBE

Consider a galaxy that is in equilibrium with its stars orbiting in a steady potential $\Phi(\mathbf{x})$ and having phase space probability density function $F(\mathbf{x}, \mathbf{v})$. This equilibrium is perturbed by an externally imposed potential, $\epsilon\psi^{\mathrm{e}}(\mathbf{x}, t)$. In response, the stars' DF changes from $F(\mathbf{x}, \mathbf{v})$ to $F(\mathbf{x}, \mathbf{v}) + \epsilon f(\mathbf{x}, \mathbf{v}, t)$, with the change ϵf in the DF adding $\epsilon\psi^f(\mathbf{x}, t)$ to the gravitational potential of the system. The subsequent evolution of the galaxy is governed by the CBE,

$$\frac{\partial}{\partial t}(F + \epsilon f) + [F + \epsilon f, H] = 0, \tag{2.1}$$

with Hamiltonian $H = H_0 + \epsilon\psi$, in which we assume that $H_0 \equiv \frac{1}{2}\mathbf{v}^2 + \Phi$ and the perturbation potential

$$\psi = \psi^{\mathrm{e}} + \psi^f, \tag{2.2}$$

is the sum of the externally imposed stimulus ψ^{e} plus the galaxy's response ψ^f. The latter is related to f through Poisson's equation,

$$\psi^f(\mathbf{x}, t) = \int d^3\mathbf{x}' G(\mathbf{x}, \mathbf{x}')\rho^f(\mathbf{x}', t), \tag{2.3}$$

where $G(\mathbf{x}, \mathbf{x}')$ is the Green function for Poisson's equation and the density response, ρ^f, is related to the perturbed DF by $\rho^f = \int d\mathbf{v} f$.

We assume that our background Hamiltonian, H_0, is integrable, allowing us to construct angle-action variables $(\boldsymbol{\theta}, \mathbf{J})$ (Arnold, 1987). Invoking Jeans' theorem we may assume that $F = F(\mathbf{J})$ and $H_0 = H(\mathbf{J})$ only (Binney & Tremaine, 2008). Therefore expanding the Poisson brackets in Equation 2.1 gives

$$\frac{\partial f(\boldsymbol{\theta}, \mathbf{J}, t)}{\partial t} + \epsilon\left(\boldsymbol{\Omega}(\mathbf{J}) \cdot \frac{\partial f}{\partial \boldsymbol{\theta}} - \frac{\partial F(\mathbf{J})}{\partial \mathbf{J}} \cdot \frac{\partial \psi}{\partial \boldsymbol{\theta}}\right) + \epsilon^2 [f, \psi] = 0, \tag{2.4}$$

where we have defined $\boldsymbol{\Omega}(\mathbf{J}) \equiv \frac{\partial H_0}{\partial \mathbf{J}}$. The terms in the first bracket of Equation 2.4, of order ϵ, constitute the linear response of the CBE, while those in the second bracket, of order ϵ^2, constitute the non-linear terms. Solving the CBE including

non-linear terms analytically is difficult. However this is more naturally done by N-body simulations.

As the angles are by construction periodic, it is natural then to express the response DF f as a Fourier series in the angles,

$$f(\boldsymbol{\theta}, \mathbf{J}) = \sum_{\mathbf{m}} f_{\mathbf{m}}(\mathbf{J}) \mathrm{e}^{\mathrm{i}\mathbf{m}\cdot\boldsymbol{\theta}}, \tag{2.5}$$

with analogous expansions for ψ^{e} and ψ^{f}.

Substituting these expansions into (2.1) and dropping non-linear terms, those of $\mathcal{O}(\epsilon^2)$, the linearised evolution equation for the response DF becomes

$$\frac{\partial f_{\mathbf{m}}(\mathbf{J}, t)}{\partial t} + \mathrm{i}\mathbf{m} \cdot \left[\boldsymbol{\Omega}(\mathbf{J}) f_{\mathbf{m}}(\mathbf{J}, t) - \frac{\partial F(\mathbf{J})}{\partial \mathbf{J}} \psi_{\mathbf{m}}(\mathbf{J}, t) \right] = 0, \tag{2.6}$$

One way to proceed from here would be to Laplace transform in time in order to remove the derivative in the first term, resulting in a simple algebraic equation for the (transformed) response, $\tilde{f}_{\mathbf{m}}(\mathbf{J}, \omega)$. As noted in the introduction, this is how most work attempting to solve the linearised CBE proceeds. Instead we follow the approach taken in Murali (1999) and note that the formal solution to (2.6), using an integrating factor, is simply

$$f_{\mathbf{m}}(\mathbf{J}, t) = \mathrm{i}\mathbf{m} \cdot \frac{\partial F}{\partial \mathbf{J}} \int_{-\infty}^{t} dt' \, \mathrm{e}^{-\mathrm{i}\mathbf{m}\cdot\boldsymbol{\Omega}(\mathbf{J})(t-t')} \psi_{\mathbf{m}}(\mathbf{J}, t'). \tag{2.7}$$

Potential-Density Pairs

A complication is that the $\psi_{\mathbf{m}} = \psi^{\mathrm{e}}_{\mathbf{m}} + \psi^{f}_{\mathbf{m}}$ factor in the integrand depends on the $f_{\mathbf{m}}$ through Poisson's equation (2.3). Following Kalnajs (1976), we deal with this by introducing a complete set of potential–density pairs $(\psi^{(p)}(\mathbf{x}), \rho^{(p)}(\mathbf{x}))$, the elements of each pair p being related by Poisson's equation

$$\psi^{(p)} = \int d^3\mathbf{x}' G(\mathbf{x}, \mathbf{x}') \rho^{(p)}(\mathbf{x}'). \tag{2.8}$$

For any choice of such basis there is a matrix

$$\mathcal{E}^{pq} = -\int \mathrm{d}\mathbf{x} \left[\psi^{(p)}(\mathbf{x}) \right]^{*} \rho^{(q)}(\mathbf{x}), \tag{2.9}$$

that describes how the elements of basis pairs p and q are related. We expand the external potential and the induced response

$$\psi^{\mathrm{e}}(\mathbf{x},t) = A_p(t)\psi^{(p)}(\mathbf{x}), \tag{2.10a}$$

$$\psi^{f}(\mathbf{x},t) = B_p(t)\psi^{(p)}(\mathbf{x}), \tag{2.10b}$$

with the coefficients $A_p(t)$ and $B_p(t)$ respectively. We take the convention that repeated indices are summed over, unless otherwise stated. In principle the number of terms in this sum is infinite. However in order to make computational progress these expansions will be limited to a finite number of terms.

Given $\psi^{\mathrm{e}}(\mathbf{x})$, the coefficients A_p that minimise the mean-square error

$$\int \left| \psi^{\mathrm{e}} - \sum_p A_p \psi^{(p)} \right|^2 \mathrm{d}\mathbf{x}, \tag{2.11}$$

can be found by solving the system of of equations

$$\mathcal{P}^{pq} A_q = \int \mathrm{d}\mathbf{x} \left(\psi^{(p)}\right)^\star \psi^{\mathrm{e}}, \tag{2.12}$$

in which the matrix, $\mathcal{P}^{pq}$, is given by

$$\mathcal{P}^{pq} = \int d\mathbf{x} \left(\psi^{(p)}\right)^\star \psi^{(q)}. \tag{2.13}$$

Due to the linear nature of Poisson's equation, the density that sources this potential is given by

$$\rho^{\mathrm{e}}(\mathbf{x},t) = A_p(t)\rho^{(p)}(\mathbf{x}). \tag{2.14}$$

By the same argument, the potential that is sourced by the response of the disc is

$$\rho^{f}(\mathbf{x},t) = B_p(t)\rho^{(p)}(\mathbf{x}). \tag{2.15}$$

The expansion coordinates, $A_p(t)$ and $B_p(t)$, completely describe the evolution of our stimulus and response potentials, respectively.

Volterra Equation

We can now combine our potential density pair expansion, with Equation 2.7. To get the evolution of the density response,

$$\mathcal{E}^{pq} B_q(t) = \int_{-\infty}^{t} \mathrm{d}t' \, \mathcal{K}^{pq}(t-t') \left[A_q(t') + B_q(t') \right]. \tag{2.16}$$

This is a linear Volterra equation, where the kernel, $\mathcal{K}^{pq}$, is defined as

$$\begin{aligned} &\mathcal{K}^{pq}(t-t') = -\mathrm{i}(2\pi)^d \times \\ &\int \mathrm{d}^d\mathbf{J} \sum_{\mathbf{m}} \mathbf{m} \cdot \frac{\partial F(\mathbf{J})}{\partial \mathbf{J}} \mathrm{e}^{-\mathrm{i}\mathbf{m}\cdot\boldsymbol{\Omega}(\mathbf{J})(t-t')} \left(\hat{\psi}^{(p)}_{\mathbf{m}}(\mathbf{J}) \right)^* \hat{\psi}^{(q)}_{\mathbf{m}}(\mathbf{J}), \end{aligned} \tag{2.17}$$

and we have been temporarily paralysed by indecision over the number of dimensions d for which we will write this. It is straightforward to solve Equation (2.16) numerically for the unknown response coefficients $B_q(t)$ once the kernel (2.17) and the matrix (2.9) have been constructed. By using the potential-density pairs, we have recast our problem to one of matrix arithmetic. Although the lower limit of this integral goes to negative infinity, we will assume that our disc is unperturbed until $t = 0$, mean that we have the initial condition $B_q(t = 0) = 0$.

There is however, no free lunch. By expanding in potential-density pairs, we have embedded Laplace's equation into our formalism. Although this has allowed us to trivially include self-gravity, it has come at some cost - from Equation 2.16 we can only solve for the real space response rather than the full phase space response. In principle, however, this problem can be rectified because once the potential evolution is determined, Equation 2.7 can be solved to obtain the full phase space evolution.

The kernel given in in Equation 2.17 has an intimate relationship to the frequency dependent response matrix, $\hat{\mathcal{M}}^{pq}(\omega)$ (e.g. Binney & Tremaine, 2008), as it represents how a disk will respond to an external perturbation and its own response. In chapter 5 we will use the response matrix to calculate the Landau modes of a disc. In fact, by assuming that $A_p = \hat{A}_p \mathrm{e}^{-\mathrm{i}\omega t}$ where $\hat{A}_p$ is constant and similarly for B_p, we can obtain

$$\mathcal{E}^{pq} \hat{B}_q(\omega) = \hat{\mathcal{M}}^{pq}(\omega) \left[\hat{A}_q(\omega) + \hat{B}_q(\omega) \right], \tag{2.18}$$

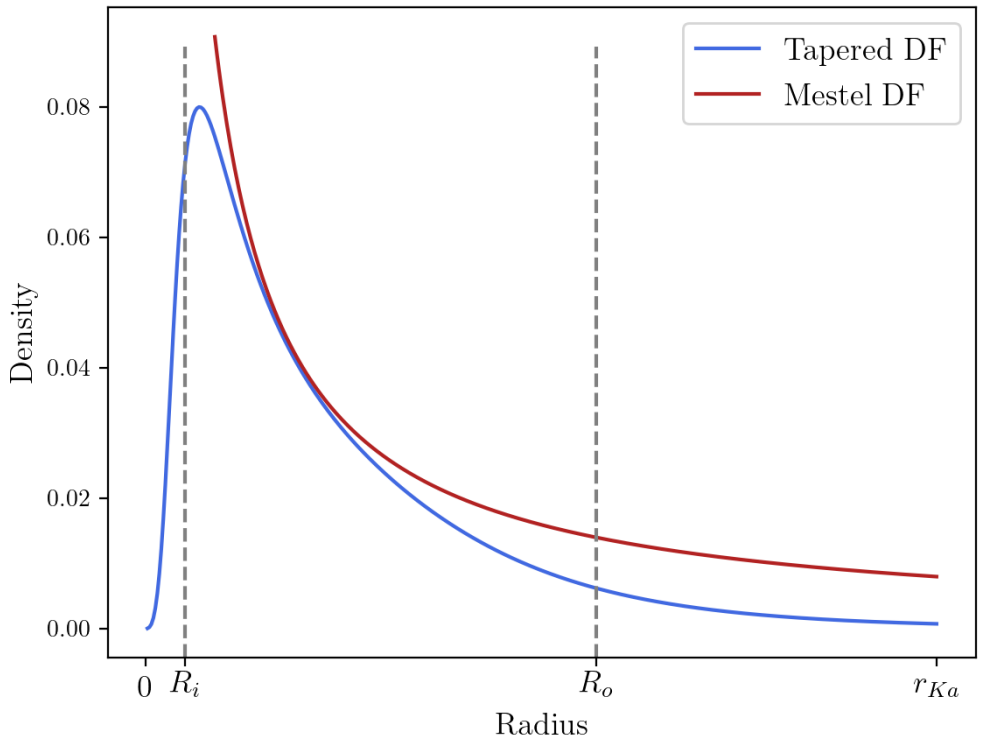


Figure 2.1: The density distribution for the tapered and untapered Mestel disk. We take $R_i = 1$ and $R_o = 11.4$.

from Equation 2.16 where the response matrix is

$$\hat{\mathcal{M}}^{pq}(\omega) = -i(2\pi)^d \times \int \mathrm{d}^d\mathbf{J} \sum_{\mathbf{m}} \mathbf{m} \cdot \frac{\partial F(\mathbf{J})}{\partial \mathbf{J}} \frac{\left(\hat{\psi}^{(p)}_{\mathbf{m}}(\mathbf{J})\right)^* \hat{\psi}^{(q)}_{\mathbf{m}}(\mathbf{J})}{\omega - \mathbf{m} \cdot \boldsymbol{\Omega}(\mathbf{J})}. \tag{2.19}$$

Although this reduces Equation 2.16 to a simple matrix equation it introduces a vanishing denominator that causes much of the difficulties in the modal analysis.

2.3 Application to razor-thin axisymmetric discs

In rest of this we will restrict ourselves to in-plane perturbations of the razor-thin Mestel disc (Mestel, 1963). In equilibrium the disc's potential is $\Phi(R) = v_\mathrm{c}^2 \log(R/R_0)$ where v_c is the constant circular orbit speed which sets the mass scale of the disk. The DF that sources this potential is

$$F_m(E, L) = \mathcal{N} L^q e^{-E/\sigma_R^2}, \tag{2.20}$$

where σ_R^2 is proportional to the disc's temperature, $q = v_c^2/\sigma_R^2 - 1$ and $\mathcal{N}$ is the normalisation constant

$$\mathcal{N} = \frac{v_c^2}{2^{1+q/2}\pi^{3/2}G\sigma_R^2\Gamma(1/2+q/2)R_0^{q+1}}. \tag{2.21}$$

This DF generates a density profile $\rho(R) = v_c^2/(2\pi GR)$. The scale free nature of this makes it a convenient model for analytic work, however less so for numerical work. In order to deal with the singularity at the centre of the disk and the infinite extent of the Mestel disk we follow the approach taken in Sellwood & Carlberg (2014) and Fouvry et al. (2015b) and introduce inner and outer tapers

$$T_{\rm inner}(L) = \frac{L^{\nu_t}}{(R_i v_c)^{\nu_t} + L^{\nu_t}}, \tag{2.22}$$

$$T_{\rm outer}(L) = \left[1 + \left(\frac{L}{R_o v_c}\right)^{\mu_t}\right]^{-1}. \tag{2.23}$$

The two parameters ν_t and μ_t set the sharpness of the inner and outer taper respectively with $R_{\rm i}$ and $R_{\rm o}$ setting the position of the tapers. We also assume that only a fraction $0 < \xi \leq 1$ of the disk is active. This gives the total DF

$$F(E, L) = \xi T_{\rm inner}(L) T_{\rm outer}(L) F_m(E, L). \tag{2.24}$$

We take $v_c = G = R_0 = 1$. For the distribution function we take $\nu_t = 4$, $\mu_t = 5$, $R_i = 1$, $R_o = 11.4$ and $\xi = 1/2$. We take $\sigma_r = 0.35$, unless otherwise specified.

Where appropriate we express our time units in terms of dynamical times at the inner taper, in our units we take $T_{\rm dyn} = 2\pi R_{\rm i}/v_{\rm c} = 2\pi$.

2.3.1 Potential-Density Pairs in 2D

Taking advantage of the axisymmetry of the background DF we choose to decompose our basis functions as

$$\psi^{(p)}(r, \phi) = e^{i\ell\phi}\mathcal{U}_n^\ell(R), \qquad \rho^{(p)}(R, \phi) = e^{i\ell\phi}\mathcal{D}_n^\ell(R), \tag{2.25}$$

in which the label $p = (n, \ell)$. We assume that the radial functions, $\mathcal{U}_n^\ell(R)$ and $\mathcal{D}_n^\ell(R)$, are real. If they are chosen well then we will be able to capture the

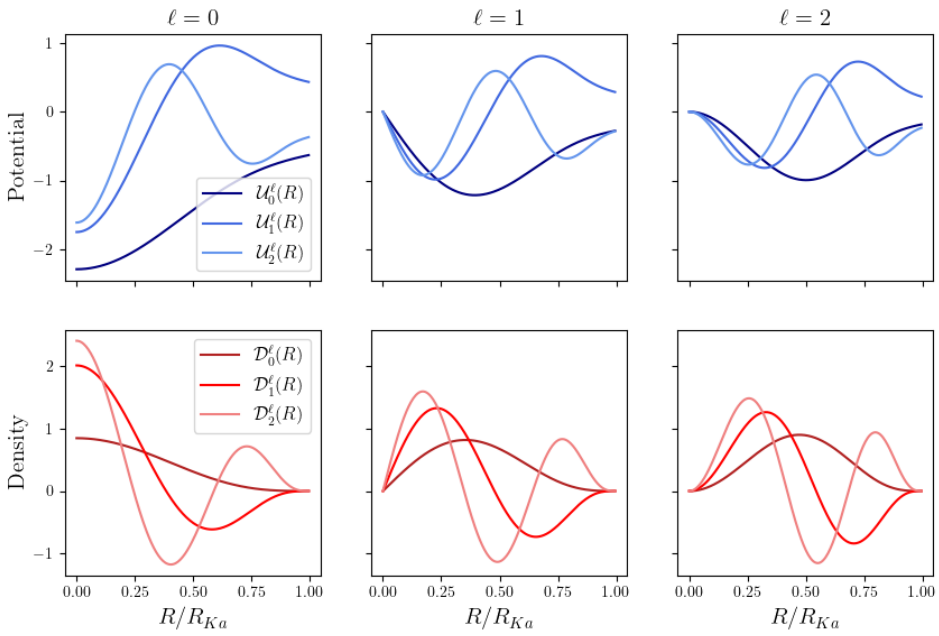


Figure 2.2: The radial part of the Kalnajs basis functions. The top row shows the potential functions, $\mathcal{U}_n^\ell(R)$, and the bottom row the density functions, $\mathcal{D}_n^\ell(R)$. Each column plots the first three n modes for a given ℓ. The radial scale length is given by R_{ka} which we take $R_{ka} = 15$ for the whole of this work.

relevant physics of the perturbation of interest using only a small number of coefficients, which will minimise the number of elements we need to include in the kernel $\mathcal{K}^{pq}$ (equation 2.16). In what follows we will outline two different sets basis functions: Kalnajs and Gaussian.

Kalnajs Basis Functions

The first is the well-known Kalnajs (1976) basis, which are constructed to satisfy the biorthogonality relation

$$-2\pi \int_0^{R_{\rm Ka}} \mathcal{D}_{n_1}^l(R)\mathcal{U}_{n_2}^l(R)\, R\mathrm{d}R = \delta_{n_1,n_2}. \tag{2.26}$$

Then the matrix $\mathcal{E}^{pq}$ that appears on the right-hand side of (2.9) is simply $\mathcal{E}^{pq} = \delta^{pq}$. The Kalnajs basis have two parameters: $k_{Ka} \in \mathbb{N}$, which sets the highest polynomial power and a scale radius $R_{Ka} \in \mathbb{R}^+$, the radius beyond which the density functions, $\mathcal{D}_n^\ell(R)$, are zero. We will keep $k_{Ka} = 4$ and $R_{Ka} = 15$ through out this thesis.

In Equation 2.27 – Equation 2.29 $\ell \to |\ell|$. In the following we express $\tilde{R}$ as a dimensionless quantity $\tilde{R} = R/R_{Ka}$. A limited number of the basis functions are plotted in Figure 2.2. The potential, $\mathcal{U}_n^\ell(\tilde{R})$, and density, $\mathcal{D}_n^\ell(\tilde{R})$, radial functions are given by

$$\begin{aligned}
\mathcal{U}_n^\ell(\tilde{R}) &= -\sqrt{\frac{G}{R_{Ka}}}\mathcal{P}(k_{Ka},\ell,n)\tilde{R}^\ell \sum_{i=0}^{k_{\rm Ka}}\sum_{j=0}^{n}\alpha_{Ka}(k_{Ka},\ell,n,i,j)\tilde{R}^{2i+2j},\\
\mathcal{D}_n^\ell(\tilde{R}) &= \frac{(-1)^n}{\sqrt{GR_{Ka}^3}}\mathcal{S}(k_{Ka},\ell,n)(1-\tilde{R}^2)^{k_{Ka}-1/2}\tilde{R}^\ell \sum_{j=0}^{n}\beta_{Ka}(k_{Ka},\ell,n,j)(1-\tilde{R}^2)^j.
\end{aligned} \tag{2.27}$$

In which the constants $\mathcal{P}$, $\mathcal{S}$, α and β can be expressed in terms of the Pochammer symbol, $[a]_i$, as

$$\begin{aligned}
\mathcal{P}(k,\ell,n) &= \left\{\frac{\Gamma(2k+\ell+2n+1/2)\Gamma(2k+\ell+n+1/2)\Gamma(\ell+n+1/2)}{\Gamma(\ell+n+1/2)\Gamma^2(\ell+1)\Gamma(n+1)}\right\}^{1/2},\\
\alpha_{Ka}(k,\ell,n,i,j) &= \left\{\frac{[-k]_i[\ell+1/2]_i[2k+\ell+n+1/2]_j}{[\ell+1]_i[1]_i[\ell+i+1]_j[\ell+1/2]_j[1]_j}\right\}[i+\ell+1/2]_j[-n]_j.
\end{aligned} \tag{2.28}$$

The normalisation for the density are,

$$\begin{aligned}
\mathcal{S}(k,\ell,n) &= \frac{\Gamma(k+1)}{\pi\Gamma(2k+1)\Gamma(k+1/2)}\\
&\quad\times\left[\frac{\Gamma(2k+\ell+2n+1/2)\Gamma(2k+n+1)}{\Gamma(\ell+n+1/2)\Gamma(n+1)\Gamma^{1/2}(2k+\ell+n+1/2)}\right]^{1/2},\\
\beta_{Ka}(k,\ell,n,j) &= \frac{[2k+\ell+n+1/2]_j[k+1]_j[-n]_j}{[2k+1]_j[k+1/2]_j[1]_j},
\end{aligned} \tag{2.29}$$

where the Pochhammer symbol is defined as

$$[a]_i = \begin{cases} 1, & \text{if } i=0,\\ a(a+1)...(a+n-1), & \text{if } i>0. \end{cases} \tag{2.30}$$

Gaussian Basis Functions

Our second way of representing the density and potential is inspired by De Rijcke et al. (2019a). The density basis functions are given by

$$\mathcal{D}_n^\ell(R) = \frac{1}{G\sqrt{2\pi\sigma_n^2}}\exp\left[-\frac{(R-R_n)^2}{2\sigma_n^2}\right], \tag{2.31}$$

in which R_n sets the centre of the $n^{\rm th}$ basis function and σ_n its width. The corresponding potential function is the convolution of the density with the Green function of Poisson's equation. We discuss this calculation in detail in Appendix A. Strictly speaking, as these functions do not span the full set of responses to arbitrarily fine resolution – which is chosen a priori by the width of the functions – they do not constitute as basis. Nevertheless, we refer to the functions as 'basis' elements.

Unlike the Kalnajs basis, these offer the possibility of tailoring the details of the basis to optimise the sampling both of the imposed perturbation and its response. We do not investigate that possibility here, but instead adopt a simpler approach in line with De Rijcke et al. (2019a). To ensure good sampling of the potential response at small radii, where the density is largest, we space R_n logarithmically between $0.50R_0$ and $15R_0$. The width, σ_n, is picked such that $R_n + \sigma_n = R_{n+1}$, making each basis function unresolved. Unlike the Kalnajs basis, these basis functions are not bi-orthonormal, but it is straightforward to use (2.9) to calculate $\mathcal{E}^{pq}$.

Given either of these bases – or indeed any other – one must choose the order at which to curtail the expansion, $N_{\rm max}$. The tapered Mestel disk's stability is highly sensitive to the inner taper (Evans & Read, 1998); the basis functions must therefore be able to resolve fine structure at small radii. One must also ensure that the perturbation can be sufficiently well resolved at larger radii. For the Kalnajs basis functions, for which the resolution increases linearly with $N_{\rm max}$, ensuring good resolution of the inner taper, will ensure a good resolution of any external perturbation. However for the Gaussian basis functions, for which resolution logarithmically decreases with radius one has to pay more attention. By comparing the true perturbation with its representation in the basis functions one can ascertain if the number of basis functions needs to be increased. We discuss how limiting the basis expansion leads to a blurring of the density response, akin to softening in N-body simulations (§ 4.7.1 Binney & Tremaine, 2008), in section 2.5.

2.3.2 Angle–action variables for 2d axisymmetric discs

After choosing a basis $\psi^{(p)}(\mathbf{x})$ the next step is to expand it as a Fourier series in angle–action coordinates,

$$\hat{\psi}^{(p)}_{\mathbf{m}}(\mathbf{J}) = \frac{1}{(2\pi)^2}\int \mathrm{d}\theta_1\mathrm{d}\theta_2 \mathrm{e}^{-\mathrm{i}m_1\theta_1}\mathrm{e}^{-\mathrm{i}m_2\theta_2}\psi^{(p)}(R,\phi), \tag{2.32}$$

in which $\mathbf{m} = (m_1, m_2)$ and R and ϕ vary as the angles θ_1, θ_2 advance along the orbit having actions $\mathbf{J}$. By definition (e.g. Arnold, 1987) the actions are

$$\begin{aligned} J_1 &= J_r = \frac{1}{\pi}\int_{R_-}^{R_+} dr\sqrt{2(E-\Phi(R)) - L^2/R^2}, \\ J_2 &= J_\phi = L, \end{aligned} \tag{2.33}$$

where (E, L) are the orbit's specific energy and angular momentum and (R_+, R_-) are its apo- and peri-centre radii, respectively, which satisfy $2(E-\Phi(R)) - \frac{L^2}{R^2} = 0$. Each orbit has a pair of associated frequencies, $\Omega_1 = \Omega_R$ and $\Omega_2 = \Omega_\phi$, which are given by (e.g. Binney & Tremaine, 2008; Fouvry et al., 2015b)

$$\begin{aligned} \frac{2\pi}{\Omega_1} &= 2\int_{R_-}^{R_+} \frac{dR}{\sqrt{2(E-\psi(R)) - L^2/R^2}}, \\ \frac{\Omega_2}{\Omega_1} &= \frac{L}{\pi}\int_{R_-}^{R_+} \frac{dR}{R^2\sqrt{2(E-\psi(R)) - L^2/R^2}}. \end{aligned} \tag{2.34}$$

Given $\Phi(R)$ we can specify an orbit uniquely using any of (R_-, R_+), (E, L) or (J_r, J_ϕ). It is straightforward to map from one set of coordinates to the other, except for the map $(J_r, J_\phi) \to (E, L)$ which requires the inversion of a non-linear integral map. In practice we transform the integral over $\mathbf{J}$ in equation (2.17) to one over (R_-, R_+), due to the trivial mapping to other coordinates and because the physical extent of the disk is more naturally expressed in (R_-, R_+) than in (E, L). Directly from the definition of apocentre and pericentre, we have that

$$E = \frac{R_+^2\Phi_+ - R_-^2\Phi_-}{R_+^2 - R_-^2}, \qquad L = \sqrt{\frac{2(\Phi_+ - \Phi_-)}{R_-^{-2} - R_+^{-2}}}, \tag{2.35}$$

where $\Phi_\pm \equiv \Phi(R_\pm)$.

The angles associated with the actions Equation 2.33 are (Lynden-Bell & Kalnajs, 1972)

$$\theta_1(R,\phi) = \Omega_1\int_{C_1[R]} dR' \frac{1}{\sqrt{2(E-\Phi_0(R')) - L^2/R;'^2}}, \tag{2.36a}$$

$$\theta_2(R,\phi) = \phi + \int_{C_1[R]} dR' \frac{\Omega_2 - L/R'^2}{\sqrt{2(E-\Phi_0(R')) - L^2/R'^2}}, \tag{2.36b}$$

in which the limits, C_1, goes from pericentre R_- to the radius of interest, R. Tremaine & Weinberg (1984) showed that $\hat{\psi}^{(n,\ell)}_{\mathbf{m}}$ vanish unless $\ell = m_2$, i.e.

$$\hat{\psi}^{(n,\ell)}_{\mathbf{m}}(\mathbf{J}) = \delta^{\ell}_{m_2} \mathcal{W}^{m_1}_{\ell m_2 n}(\mathbf{J}), \tag{2.37}$$

thanks to the symmetry assumed in our decomposition of the basis functions (Equation 2.25). It is important to note that the Kronecker delta here means that different angular harmonics decouple, a direct consequence of the decomposition in Equation 2.25. If the radial function is real then $\mathcal{W}$ becomes,

$$\mathcal{W}^{m_1}_{\ell m_2 n}(\mathbf{J}) = \frac{1}{\pi} \int_{R_-}^{R_+} \mathrm{d}R \frac{\mathrm{d}\theta_1}{\mathrm{d}R} \mathcal{U}^{\ell}_n \cos\left[m_1\theta_1 + m_2(\theta_2 - \phi)\right]. \tag{2.38}$$

In transforming the θ_1 integral into one over R we have introduced a Jacobian

$$\frac{\mathrm{d}\theta_1}{\mathrm{d}R} = \frac{\Omega_1}{\sqrt{2(E-\Phi(R)) - L^2/r^2)}}. \tag{2.39}$$

The cosine in (2.38) is an even function, which means that $\mathcal{W}^{m_1}_{\ell m_2 n} = \mathcal{W}^{-m_1}_{\ell -m_2 n}$, reducing the number of calculations needed. We do the calculation of $\psi^{n\ell}_{\mathbf{m}}$ for a grid of points in (R_-, R_+), where each R_- and R_+ is spaced uniformly with spacing ΔR_{grid}. The specific parameters required depend on the potential-density pairs used.

2.3.3 Computing the evolution Kernel

Once we have the basis functions we must calculate the kernel (Equation 2.17 for each time step from $t - t' = 0$ to $t - t' = t_{\mathrm{end}}$. As our basis functions are on a grid of points in (R_-, R_+) space we need to transform our integral measure and derivatives in (2.17) to reflect the change of coordinates. By the chain rule the derivative with respect to the background DF that appears in (2.17) is

$$\mathbf{m} \cdot \frac{\partial F}{\partial \mathbf{J}} = m_1\Omega_1 \left(\frac{\partial F}{\partial E}\right)_L + m_2 \left[\Omega_2 \left(\frac{\partial F}{\partial E}\right)_L + \left(\frac{\partial F}{\partial L}\right)_E\right]. \tag{2.40}$$

We split the Jacobian associated with the change in integration variables from $(J_1, J_2) \to (R_-, R_+)$ into two factors. The first is the Jacobian from $(J_1, J_2) \to$

(E, L), which is given by

$$\frac{\partial(J_1, J_2)}{\partial(E, L)} = \frac{1}{\Omega_1}. \tag{2.41}$$

The second is the Jacobian from $(E, L) \rightarrow (R_-, R_+)$, which is best calculated numerically from (2.35). The overall Jacobian is then the product of the two.

Due to the time translation symmetry of the kernel we only calculate it at a set of points $t - t_0 \in [0, \Delta t, 2\Delta t, ..., t_{end}]$ where we use a uniform time step, Δt. We can further reduce the number of calculations needed by exploiting the structure of equation (2.37). Due to the symmetries in the basis functions, the kernel satisfies the properties

$$\mathcal{K}^{n_q \ell_q}_{n_p \ell_q}(t - t') = \left[\mathcal{K}^{n_q \ell_{-q}}_{n_p \ell_{-q}}(t - t')\right]^*, \tag{2.42}$$

$$\mathcal{K}^{n_q \ell_q}_{n_p \ell_p} = \mathcal{K}^{n_p \ell_p}_{n_q \ell_q}. \tag{2.43}$$

We can use these properties to reduce the time required to produce each kernel by approximately a factor of four.

Once we have the kernel we can then use standard techniques to solve the integral equation (2.16). We use the trapezium rule to approximate the integral on the right-hand side by a sum. Rearranging this sum gives an explicit expression for the response coefficients, $B_p(t_n)$, in terms of their earlier values:

$$\begin{aligned} B_s(t_n) = \Delta t \left[\mathcal{E} - \frac{1}{2}\mathcal{K}(t_n)\right]^{-1}_{sp} \times \Bigg[\frac{1}{2}\mathcal{K}^{pq}(t_n)\left(A_q(0) + B_q(0)\right) \\ + \sum_{i=1}^{n-1} \mathcal{K}^{pq}(t_n - t_i)\left[A_q(t_i) + B_q(t_i)\right]\Bigg]. \end{aligned} \tag{2.44}$$

Here we have assumed that the external perturbation, $A_p(t)$, is turned on only for times $t \geq t_0 = 0$. The terms involving $A_q(t_i)$ on the right-hand side describe how the external perturbation drives a response in the disk. The $B_q(t_i)$ in the inner square bracket represents the self gravity of the response.

In some situations we will want to tease out the role of self gravity. To do this, we will compare the full response calculated using the full expression in Equation 2.44, to the test particle response, where the self-gravity is not included.

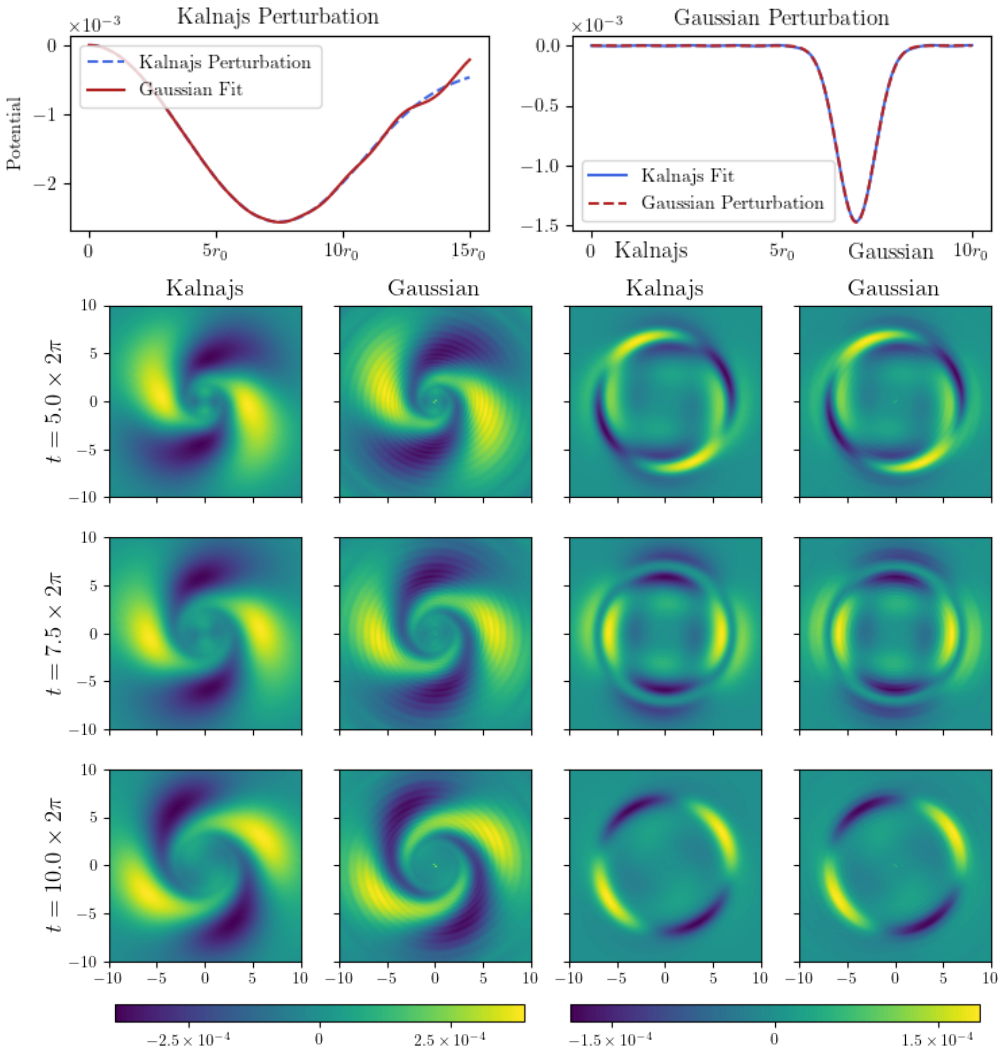


Figure 2.3: Comparing the effect of the choice of basis functions on the computed density response. The top row shows two different $\ell = 2$ potential perturbations applied at $t = 0$ together with their representations in the Kalnajs and Gaussian bases. The rows underneath plot the subsequent density evolution computed using the Kalnajs (left) and Gaussian (right) bases. There is strong agreement between the results calculated using different basis functions. Nevertheless, imprints from the basis functions can be seen in the response - in particular the rings from the Gaussian basis functions.

To calculate the test particle response we use Equation 2.16 while dropping the response term in the integral, explicitly

$$\mathcal{E}^{pq} B_q(t) = \int_{-\infty}^{t} \mathrm{d}t' \, \mathcal{K}^{pq}(t - t') A_q(t'), \tag{2.45}$$

which can be solved once again by numerically integrating the LHS via the trapezium rule, akin to Equation 2.44.

2.3.4 Comparison of the effects of basis choice

In practice we must restrict the number of basis functions used to represent the disturbing potential and the subsequent response of the galaxy. Therefore our model does not in general see the 'true' perturbing potential, nor can it predict all details of the response: instead, both are projected onto our finite basis. To test for systematic bases introduced by this projection we can compare predictions using different basis.

When calculating kernels with the Kalnajs and Gaussian basis functions we take the following parameters for the rest of this chapter. The number of terms in our basis expansion $N_{max} = 48$, which is taken as a balance between computational speed and numerical convergence of the density response. When calculating the action-angle representation of the potential-density pairs, $\mathcal{W}^{m1}_{\ell m 2 n}(\mathbf{J})$ (Equation 2.38, we do so on a uniform grid in (R_-, R_+) coordinates, with step size $\Delta R_{grid} = 0.06$ and $R_- \leq R_+ \leq R_{\rm Ka}$. To obtain the kernel from these we sum over m_1. Therefore we calculate the action-angle representation for of the basis functions for $-7 \leq m_1 \leq 7$. We check for numerical convergence in all of these parameters and find them to best balance computational speed and accuracy.

Figure 2.3 shows the density response when the same perturbation is represented in two different basis functions. The top row shows two different impulsive potentials $\psi^{\rm e}(\mathbf{x})$ that were applied to the disc at $t = 0$, i.e. so that the time dependence of $A_p(t) \propto \delta(t)$, both with $\ell = 2$. The left-hand side shows a large-scale perturbation, which due to the structure of the Kalnajs basis functions is zero beyond $R = R_{\rm Ka} = 15$ and the right-hand side shows a more localised perturbation. The rows underneath each plot the time evolution of the corresponding density response, calculated using either the Kalnajs basis (left) or the Gaussian basis (right).

For the large-scale perturbation (left-hand panels), the Gaussian basis functions fail to recover the structure of the global perturbation at larger radius. This is because the Gaussian basis functions do not have compact support. Therefore when fitting the Gaussian functions, they try to fit $\rho(R) = 0$ for $R > 15R_0$. Due to the low density at larger radius, the effect of this on the subsequent evolution is minimal, as illustrated by the strong agreement between the Kalnajs and Gaussian

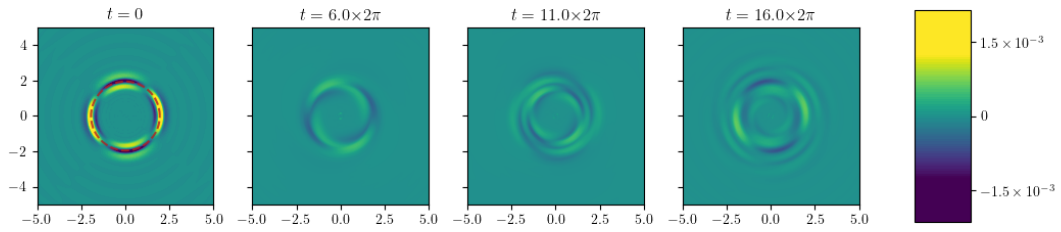


Figure 2.4: The quadrupole harmonic of the Green function at times given by the figures title. In order to produce the fine structure required to represent a point perturbation, we use the Gaussian basis functions. We perturb the disk at a radius of $R_p = 2$, shown by the red circle in the $t = 0$ plot.

density response. One can also see the imprint of the Gaussian basis functions, with the artificial rings imprinted on the density response.

As the Gaussian basis functions are logarithmically spaced, they are able to reproduce finer structure at smaller radii compared to the Kalnajs which are 'linearly spaced'. The right hand panels of Figure 2.3 illustrate this clearly, where the Gaussian basis function can represent smaller scales close to the centre. The resolution of the Kalnajs basis function can be improved. However due to its linear (rather than logarithmic) spacing one would need more basis functions than for the Gaussian case.

2.4 Green Function of the Linear CBE

From Equation 2.16 it is a straightforward exercise to calculate the Green function of the density response, $\mathcal{G}_\rho$. By definition the Green function is

$$\rho^f(\mathbf{x}, t) = \int d\mathbf{x}' dt' \mathcal{G}_\rho(\mathbf{x}, \mathbf{x}'; t - t')\rho^e(\mathbf{x}', t'). \tag{2.46}$$

Note that the Green function is the same as the response function (c.f. § 5.1.1 Binney & Tremaine, 2008). The Green function is the fundamental object for solving initial value problems; from it all quantities of interest can be calculated, allowing the evolution of the system to be completely determined. To extract $\mathcal{G}_\rho$ from Equation 2.46 one can perturb the disk at a point in space and time, i.e. $\rho^e(\mathbf{x}', t') = \delta(\mathbf{x}' - \mathbf{x}_p)\delta(t' - t_p)$ and then use the formalism developed in section 2.2 to evaluate the density evolution. In order to calculate the density evolution using

Equation 2.16, which is the left-hand side of Equation 2.46, we must express this Dirac delta external perturbation in our basis functions. However for a finite basis we can only approximate a delta function, i.e.

$$\delta(\mathbf{x}' - \mathbf{x}_p)\delta(t' - t_p) \approx \delta(t - t'_p) \sum_{q=0}^{N_{\max}} A_q(\mathbf{x}_p)\rho^q(\mathbf{x}'), \tag{2.47}$$

where the left-hand side equals the right-hand side in the limit $N_{\max} \to \infty$ and we have introduced an explicit summation to indicate the dependence on $N_{\max}$. In this expression, we have made the perturbation's expansion coefficients an explicit function of $\mathbf{x}_p$. With this approximation, we get

$$\rho^f_{\mathbf{x}_p t_p}(\mathbf{x}, t) = \int d\mathbf{x}' \mathcal{G}_\rho(\mathbf{x}, \mathbf{x}'; t - t_p) \sum_{q=0}^{N_{\max}} A_q(\mathbf{x}_p)\rho^q(\mathbf{x}'). \tag{2.48}$$

If we were to take the limit $N_{\max} \to \infty$ we could integrate out the convolution to obtain

$$\mathcal{G}_\rho(\mathbf{x}, \mathbf{x}_p; t - t_p) = \rho^f_{\mathbf{x}_p t_p}(\mathbf{x}, t). \tag{2.49}$$

But as our basis expansion must be curtailed with a finite number of terms, we want to choose a basis that can better approximate a delta function, with a small number of terms. To do this we use the Gaussian basis functions and plot the quadrupole harmonic of the Green function in Figure 2.4.

2.5 Softening due to Potential Density Pair Expansion

Despite not explicitly including softening in the calculation of our basis functions, we have implicitly done so when we have curtailed our basis expansion to a finite number of terms. In fact, we can consider 'two types' of softening: angular and radial softening.

- *Angular Softening*: By decomposing the response into a finite number of individual angular harmonics, rather than summing over all ℓ, we soften the angular interactions. However as angular harmonics decouple (c.f. Equation 2.37), the ignored higher harmonics will not affect the included lower harmonics in linear response theory.

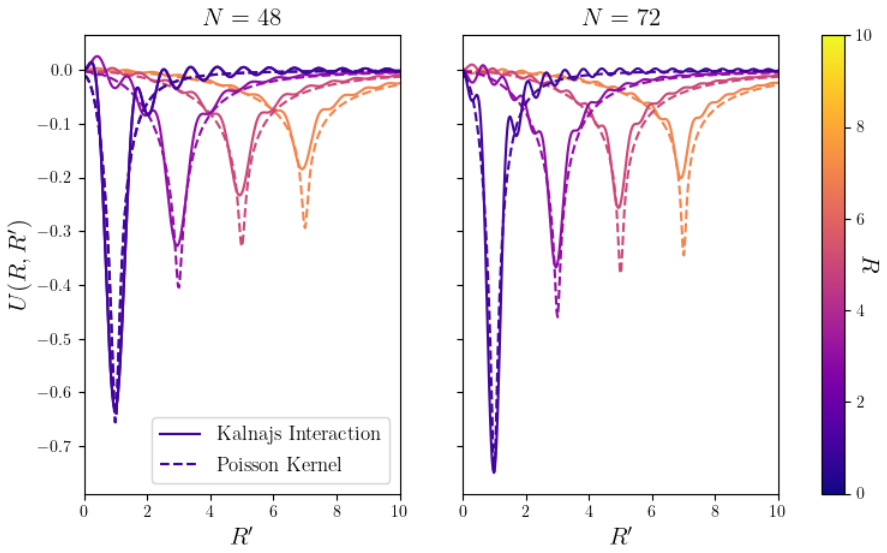


Figure 2.5: The quadrupole interaction potential of the Kalnajs basis functions and softened Plummer potential as a function of R' (x-axis) and R (colour) . *Left column:*, the interaction potential for $N_{\max} = 48$ basis functions (solid) and softening Plummer potential with $\epsilon = 0.10$. *Right column:*, the interaction potential for $N_{\max} = 72$ basis functions (solid) and softened Plummer potential with $\epsilon = 0.08$.

- *Radial Softening*: The smoothing can also happen radially, due to our finite sum of radial basis functions. Unlike for the angular case, where each harmonic evolves independently of the others, each radial harmonic will drive evolution in the other radial harmonics, a process that is captured in the off diagonal terms of the evolution kernel. Naively for the Kalnajs basis functions we would expect a softening length $\sim R_{\rm Ka}/N_{\max}$.

We will now make this intuition mathematically explicit. The particular integral of Poisson's equation is the convolution of the Green function with the density distribution, i.e.

$$\Phi(\mathbf{R}) = \int d^3\mathbf{R}' U(\mathbf{R}, \mathbf{R}')\rho(\mathbf{R}'), \tag{2.50}$$

where $U(\mathbf{R}, \mathbf{R}')$ is the Green function (or the 'interaction potential'). If an unsoftened interaction is required then the interaction potential is simply the Green function of Poisson's equation, $-G/|\mathbf{R} - \mathbf{R}'|$. A common modification in

N-body simulations, however, is to use a softened Plummer potential,

$$U(\mathbf{R}, \mathbf{R}') = -\frac{G}{\sqrt{|\mathbf{R} - \mathbf{R}'|^2 + \epsilon^2}}, \tag{2.51}$$

where ϵ is the softening length of the interaction potential (Binney & Tremaine, 2008). Applying the definition of the interaction potential to a set of potential density pairs gives $\psi^p(\mathbf{R}) = \int d^3\mathbf{R}' U(\mathbf{R}, \mathbf{R}')\rho^p(\mathbf{R}')$. Along with the ortho-normality relationship, this yields

$$U(\mathbf{R}, \mathbf{R}') = -\sum_{p,q=1}^{p,q=N_{\max}} \mathcal{E}_{pq}\psi^p(\mathbf{R})\psi^q(\mathbf{R}'), \tag{2.52}$$

where we have made the summation explicit to highlight its dependence on $N_{\max}$. As potential density pairs solve Poisson's equation by construction, any deviation of the interaction potential calculated by Equation 2.52 from $-1/|\mathbf{R} - \mathbf{R}'|$ can be described as softening due to our finite potential density pair expansion.

In Figure 2.5 we plot the interaction potential of the $\ell = 2$ Kalnajs basis functions (solid line) and a fitted quadrupole Plummer softened interaction potential (dashed line) for $N_{\max} = 48$ (left column) and $N_{\max} = 72$ (right column). As the two interaction potentials have the same angular softening (they are both the quadrupole harmonic) we need only consider the interaction potentials as a function of $R = |\mathbf{R}|$ and $R' = |\mathbf{R}'|$. We find that a relatively small softening length ($\epsilon = 0.10, 0.08$ for $N_{\max} = 48, 72$ respectively) gives a reasonable fit to the Kalnajs interaction potential. This is smaller than the naive calculation of $R_{\mathrm{Ka}}/N_{\max} = 0.30, 0.21$ with $R_{\mathrm{Ka}} = 15$. This discrepancy might be due to the limitations of Plummer softening (e.g. Dehnen, 2001; De Rijcke et al., 2019a).

2.6 Conclusion

In this chapter we have developed a method to study the self-consistent, linear response of a 2D axisymmetric disc to an external perturbation inspired by Murali (1999). In order to deal with the inhomogeneous nature of galaxies, we transformed the linearised CBE into action-angle variables. The difficulties of solving Laplace's

equation in this coordinate system was alleviated by the introduction of potential density pairs (Kalnajs, 1976). With these two building blocks, we recast the linear CBE as a matrix integral equation which could be solved to study the linear response of a disc to an external perturbation (Murali, 1999).

We then compared the response to a perturbation for two different basis functions: Kalnajs and Gaussian. Although both produced the same density response, we found that the Kalnajs basis functions were much easier to work with due to their compact support and orthogonal relationship.

We will now use the techniques set out in this chapter to study the processes that drive galaxy evolution.

6.52. ... We feel that even when all possible scientific questions have been answered, the question of life remains untouched.

— L. Wittgenstein, *Tractatus Logico-Philosophicus*

3 Spirals

Contents

3.1 Introduction 39
3.2 The Shearing Sheet & JT Equation 41
3.2.1 Derivation of the Julian & Toomre Equation 42
3.2.2 Amplification by the Shearing Sheet 54
3.3 Comparison between Shearing Sheet & Linear Response Theory 55
3.3.1 Axisymmetric Comparison 57
3.3.2 Comparison For Non-Axisymmetric Perturbations . . . 62
3.4 The Strength of Swing Amplification in Discs 65
3.4.1 Comparison in the Cold Disc Limit 66
3.4.2 Comparison of Density Response 70
3.4.3 The Role of Self-Gravity 71
3.5 Conclusion 72

3.1 Introduction

There are many different ways to tackle a physics problem: some cry out for large computational hammers, while others only need the deft use of piercing insight and precise mathematical tweezers. When Julian & Toomre (1966) first studied the shearing sheet they took the latter approach. In doing so they explained swing amplification – a key mechanism in the growth of spiral structure in a disc.

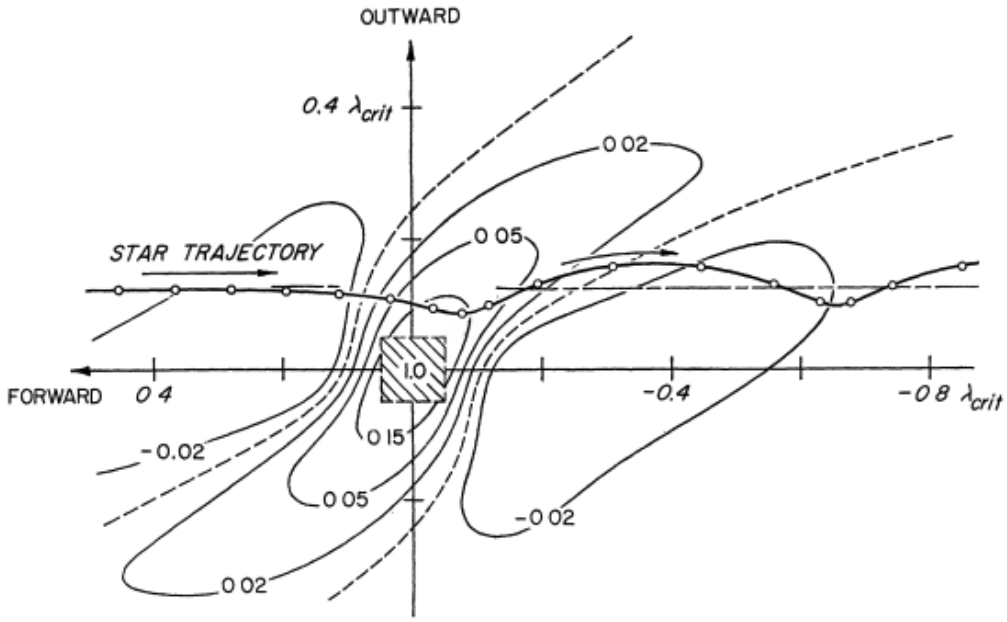


Figure 3.1: The diagram from Julian & Toomre (1966) (figure 12) that explains the mechanism behind swing-arm amplification in the shearing sheet. It shows the density response of the shearing sheet in the co-rotating frame of an imposed mass.

The physical insight of Julian & Toomre (1966) is clearly explained in their figure 12 (reproduced in Figure 3.1). The shearing sheet is a local approximation to a patch co-rotating with a imposed mass at its centre. In this frame orbits which are further outwards from the centre of rotation, and hence would rotate more slowly in a disc, travel backwards from left to right. As a star, whose trajectory is represented by the black line in Figure 3.1, approaches the origin, it is attracted to the perturbing mass from, e.g. a satellite embedded in the disc. Once closer to the mass, the star seems to linger in its vicinity as the star's speed relative to the mass decreases (the dots, which are placed at equal time steps, gather together). This results in the mass becoming 'dressed' in a cloud of stars, increasing its effective gravitational mass and amplifying the effect of the perturbation. Julian & Toomre (1966) used the analysis of the shearing sheet to propose a mechanism for the production of spiral structure: due to the shear of the disc, this dressing gets elongated out into a spiral structure

With this insight, Julian & Toomre (1966) derived an equation which governs the linear response of the shearing sheet to an imposed external perturbation. Solving the so-called JT equation led to two important conclusions: first, it is the 'swinging'

of a spiral arm from leading to trailing that leads to large amplification (hence the name 'swing-amplification'); second, self-gravity is crucial in this process, amplifying the response by orders of magnitudes. Both of these conclusions are discussed in subsection 3.2.2. Given the importance of this seminal work, a rather simple question remains: how well does the shearing sheet work as an adequate model of a full disc?

One complementary approach to study the shearing sheet would be to use the large computational hammer that N-body simulations provide. Surprisingly, this is difficult due to the nature of the Hamiltonian: the shearing sheet Hamiltonian contains a Coriolis term, due to the transformation into the co-rotating frame of the shearing sheet, which requires a complex numerical integration scheme to deal with (Rein & Tremaine, 2011; Quillen et al., 2018).

A more common approach is to study the formation of spiral arms in a whole disc using N-body simulations (e.g. Sellwood & Carlberg, 1984). However, there is a fundamental problem with N-body methods in testing the direct validity of the shearing sheet: N-body simulations are non-linear and so, for a fair comparison to Julian & Toomre (1966), a linear theory should be used. This chapter will test the two key conclusions of Julian & Toomre (1966) in the linear regime without relying on the shearing sheet approximation.

In the following section we will outline the derivation of the JT equation (Julian & Toomre, 1966) paying close attention to Binney (2020). In section 3.3 we will compare the density response of the shearing sheet and full disc to an external perturbation. Finally in section 3.4 we will put Julian & Toomre (1966) to the test in a full disc.

3.2 The Shearing Sheet & JT Equation

In this section we will outline the derivation of the equation governing the evolution of the shearing sheet - the so-called JT equation and the associated JT kernel. Although originally derived in Julian & Toomre (1966), due to its pedagogical approach the key steps of the derivation highlighted here will closely follow Binney (2020). Understanding the approach taken to obtain the JT equation, and its

geometrical assumptions, will be crucial when comparing the shearing sheet's evolution to that of our linear response calculations in section 3.3.

3.2.1 Derivation of the Julian & Toomre Equation

Solving the Linearised CBE

As with chapter 2, when studying the linear response of the shearing sheet we will start with the linearised CBE, It can be expressed in two equivalent ways

$$\frac{\partial f}{\partial t} + [f, H_0] = [\psi, F_0], \tag{3.1a}$$

$$\frac{df}{dt} = [\psi, F_0], \tag{3.1b}$$

where we have assumed that we can expand around a background DF, such that $F = F_0 + \epsilon f$, which sources a background Hamiltonian, $H = H_0 + \epsilon\psi$. We have two ways to proceed, depending on which equation we choose to solve:

1. *Eulerian*, Equation 3.1a: the flow in phase space where the observer stays at a specific phase space coordinate. To solve the initial value problem one would expand the Poisson bracket on the LHS of Equation 3.1a, which corresponds to the motion in the mean field, and solve the resulting differential equation. We chose this method in chapter 2 when working in action angle variables for two reasons: first, by Jeans theorem H_0 and F_0 are functions of action only, meaning that the Poisson brackets could be reduced to a single term each, allowing the differential equation to be solved simply with an integrating factor; second, as actions are integrals of motion, evolution in the mean field is trivial, reducing the first Poisson bracket on the LHS to $\boldsymbol{\Omega}(\mathbf{J}) \cdot \partial f / \partial \boldsymbol{\theta}$. This, however, introduces a new problem: how to solve Poisson's equation in action angle coordinates? To deal with this we introduce potential density pairs and the tedious transformation of them into action angle variables.

2. *Lagrangian*, Equation 3.1b: the flow in phase space where the observer follows a phase space element. To solve the initial value problem we integrate both sides along orbits in phase space. This is the approach we will take when

deriving the JT equation for two reasons: first, the orbits in the shearing sheet are easily integrable in real space coordinates; second, as we will stay in real space coordinates, as opposed to action angle variables, we will be able to solve Poisson's equation by Fourier transform, a method that is much simpler than introducing potential density pairs.

Of course these two methods are equivalent, however this does not mean that one approach would not be more work than the other. In what follows we will follow the Lagrangian approach, i.e. we will integrate

$$f(\mathbf{x},\mathbf{p},t) = \int_{t_\mathrm{i}}^{t} \frac{\partial \psi(\mathbf{x}[t'],t')}{\partial \mathbf{x}} \cdot \frac{\partial F_0(\mathbf{p})}{\partial \mathbf{p}} dt' \tag{3.2}$$

along unperturbed orbits, which we have made explicit functions of time by labeling $\mathbf{x} = \mathbf{x}[t']$ in the integral. We assume that the disc is unperturbed for $t' < t_\mathrm{i}$. Taking inspiration from chapter 2, we will decompose the perturbing potential into two parts: ψ^e, the contribution to the perturbing potential from an externally imposed field and ψ^f, the potential due to the self consistent response. As the self-consistent response is driven by the density perturbation, to close Equation 3.2 we integrate over $\mathbf{p}$ to obtain

$$\begin{aligned} \Sigma^f(\mathbf{x},t) &= \int f(\mathbf{x},\mathbf{p},t) d^2\mathbf{p} \\ &= \int \left[\int_{t_\mathrm{i}}^{t} \frac{\partial}{\partial \mathbf{x}} \left(\psi^f(\mathbf{x}[t'],t') + \psi^\mathrm{e}(\mathbf{x}[t'],t') \right) \cdot \frac{\partial F_0(\mathbf{p})}{\partial \mathbf{p}} dt' \right] d^2\mathbf{p}. \end{aligned} \tag{3.3}$$

where Σ^f is the density response of the sheet which is driven by a self-consistent potential, ψ^f, and an external potential ψ^e.

We now turn to actually deriving the JT equation. Although we will take a slightly different approach to the derivation of Equation 2.16, as highlighted above, the approach will follow a similar structure:

1. We will obtain the background Hamiltonian for the shearing sheet from a local analysis of a differentially rotating disc. Once we have got this, we will then solve the equations of motion for particles in the unperturbed background Hamiltonian. This will be important when solving Equation 3.2, as orbits

are to be integrated along unperturbed orbits. In chapter 2 we were agnostic about our geometry, beyond assuming that orbits in our disc are integrable, so that action-angle variables can be constructed (c.f. Arnold, 1987) and that the frequencies are not degenerate. For the derivation of the JT equation we will make specific assumptions about the nature of the Hamiltonian.

2. We Fourier transform in order to solve Poisson's equation, linking the density perturbation, Σ^f, to the potential perturbation that it sources, ψ^f. This is akin to the introducing of potential density pairs in section 2.2: in the sheet's case the density basis functions will be $\psi^k = \exp(\mathrm{i}\mathbf{x} \cdot \mathbf{k})$, with a continuous index, $\mathbf{k}$.

3. We will obtain a suitable background DF for the shearing sheet. In section 2.3 we took an 'off the shelf' DF for our disc: the Mestel disc. For the JT equation we will take a different background DF, namely a Gaussian in both velocity coordinates. We will discuss how the Mestel DF differs to our choice for the sheet's background DF.

4. Finally we will pull all of these parts together in Equation 3.3, and calculate the integral over $\mathbf{p}$ to obtain the JT equation and the JT kernel. In chapter 2 we did this by integrating over $\mathbf{x}$ and $\mathbf{v}$, exploiting the bi-orthogonality relationship for the potential density pairs, and using the canonical nature of the action angle coordinate system (c.f. the end of section 2.2). For the sheet, as we stay in real space coordinates, we can play the same trick by just integrating over $\mathbf{p}$ (c.f. Equation 3.3) as we have implicitly integrated over $\mathbf{x}$ by changing to Fourier space.

Orbits in the Shearing Sheet

At its heart the shearing sheet is a local approximation to a differentially rotating disc. In Figure 3.2 we schematically represent the coordinate system of the disc and its relation to the shearing sheet. In their derivation Julian & Toomre (1966) considered the evolution in the vicinity of a co-rotating patch, centred at a radius

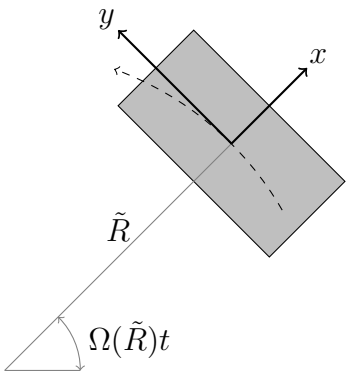


Figure 3.2: A schematic diagram showing the relationship between the global coordinate system for a disc and the local coordinate system used for the sheet. The sheet is centred on a point with radius $\tilde{R}$ that rotates with angular frequency $\Omega(\tilde{R})$. We expand in Cartesian coordinates around this point, with x pointing away from the centre of the disc and y in the direction of rotation. As $\Omega(R)$ is in general a function of R, there will be a differential shear across the patch.

$\tilde{R}$ and rotating with angular frequency $\Omega(\tilde{R})$. Locally to this point we can use a Cartesian coordinate system with y pointing in the direction of rotation and x pointing towards the anti-centre of the disc (c.f. Figure 3.2). In these coordinates the local approximation is simply $x, y \ll \tilde{R}$.

In the co-rotating frame the Lagrangian of the system is

$$\mathcal{L} = \frac{1}{2}\left[\dot{x}^2 + \left(\tilde{R}+x\right)^2\left(\Omega\left(\tilde{R}\right) + \frac{\dot{y}}{\tilde{R}+x}\right)^2\right] - \Phi_0(\tilde{R}+x). \tag{3.4}$$

For conciseness, we will abbreviate $\Omega = \Omega\left(\tilde{R}\right)$ when working with the shearing sheet. From this Lagrangian we can read off the generalised momenta

$$\begin{aligned} p_x &= \dot{x} \\ p_y &= \frac{1}{\tilde{R}}\left(\tilde{R}+x\right)^2\left[\Omega + \frac{\dot{y}}{\tilde{R}+x}\right] \simeq \tilde{R}\Omega + 2\Omega x + \dot{y} + \mathcal{O}\left(\left(\frac{x}{\tilde{R}}\right)^2\right), \end{aligned} \tag{3.5}$$

where we have ignored terms that are second order in $x/\tilde{R} \ll 1$ in the approximate expression for p_y. From these coordinates the Hamiltonian governing the shearing sheet can be constructed as

$$H = \frac{1}{2}\left[p_x^2 + \frac{p_y^2}{(1+x/\tilde{R})^2}\right] - \Omega\tilde{R}p_y + \Phi_0(\tilde{R}+x). \tag{3.6}$$

As the Hamiltonian is independent of both t and y, H and p_y are constants of motion. At this point it is prudent to introduce

$$\Delta_y \equiv p_y - \tilde{R}\Omega = 2\Omega x + \dot{y}, \tag{3.7}$$

which corresponds to the azimuthal velocity in the corotating frame. As such, it is first order in small quantities x, y. The potential term in the Hamiltonian can also be expanded as

$$\Phi_0\left(\tilde{R}+x\right) \simeq \Phi_0\left(\tilde{R}\right) + \tilde{R}\Omega^2 x + \frac{1}{2}\Omega(\Omega - A)x^2, \tag{3.8}$$

where A is Oort's first constant and is related to the epicycle frequency κ via $\kappa^2 \equiv 4\Omega(\Omega - A)$. The second term on the right side corresponds to the shear across the sheet and the third to the epicycle oscillations that the stars will undergo. This allows us to write the Hamiltonian for the shearing sheet as

$$H \simeq \frac{1}{2}\left[p_x^2 + p_y^2 - \tilde{R}^2\Omega^2\right] + \Phi_0\left(\tilde{R}\right) - x\Omega\Delta_y + \frac{1}{2}\kappa^2 x^2. \tag{3.9}$$

From this expression we see that the Hamiltonian can be split into two parts,

$$H = \left[\frac{p_x^2}{2} + \frac{\kappa^2}{2}\left(x - \frac{2\Omega\Delta_y}{\kappa^2}\right)^2 - \frac{\Omega^2\Delta_y^2}{\kappa^4}\right] - \frac{\Omega^2\Delta_y^2}{\kappa^4} + \frac{p_y^2}{2} + \Phi_0(\tilde{R}), \tag{3.10}$$

the first of which, in the square brackets, is just the Hamiltonian for a harmonic oscillator in the x direction whose centre depends on Δ_y, while the rest of the Hamiltonian is constant. The centre of oscillations is

$$\bar{x} \equiv \frac{2\Omega\Delta_y}{\kappa^2}. \tag{3.11}$$

It is worth highlighting that a key difference between the shearing sheet and any reasonable disc model will be the potentials: in the shearing sheet the effective potential is quadratic, whereas for a disc this is only the case for orbits with small radial motion, which will undergo epicyclic motion. We will study the difference in evolution between the Mestel disc and shearing sheet in the cold limit in subsection 3.4.1.

With the Hamiltonian of the shearing sheet we can now calculate the orbits of an individual particles. Due to the dependence of H on x and p_x, x oscillates harmonically about $\bar{x}$ with frequency κ, therefore

$$\begin{aligned} x &= \bar{x} + X\cos(\kappa t), \\ p_x &= -\kappa X \sin(\kappa t), \end{aligned} \tag{3.12}$$

where X is the amplitude of the oscillation and any phase factor has been ignored as it will disappear when we orbit average. As the angle $\theta_x = \kappa t$ grows lineally with time it is an angle with action,

$$J_x = \frac{1}{2\pi} \oint p_x dx = \frac{1}{2}\kappa X^2. \tag{3.13}$$

If we wanted to, we could make use of this to derive the JT equation. However, due to the shear term in the Hamiltonian, this would be a difficult task as it would need to be expressed in action angle variables. Now that we have p_x we can express the azimuthal speed in the sheet's co-rotating frame, v_ϕ, as

$$v_\phi = 2Ax + \dot{y} = 2B(x - \bar{x}) = 2BX\cos(\kappa t) \tag{3.14}$$

where we have used Oort's second constant, $B = -\kappa^2/4\Omega < 0$. We can use this expression and our expression for $x(t)$ (Equation 3.12) to obtain

$$y(t) = y_0 + \Delta_y \frac{A}{B} t - 2\frac{\Omega}{\kappa} X \sin(\kappa t) \tag{3.15}$$

where y_0 is an integration constant.

We can use v_ϕ to eliminate x from our expression for the Hamiltonian, which allows us to split our Hamiltonian into two parts

$$H = \frac{1}{2}\left[p_x^2 + \frac{\Omega}{-2B} v_\phi^2\right] + \left[\frac{\Delta_y^2}{2}\frac{A}{A-\Omega} - \frac{R^2}{2}\Omega^2 + \Phi_0(R)\right] \tag{3.16}$$

where the first bracket we will label $H_x(p_x, v_\phi)$ and the second bracket H_y which depends solely on the constant of motion Δ_y.

$$H_x \equiv \frac{1}{2}\left[p_x^2 + \frac{\Omega}{-B} v_\phi^2\right] = \frac{1}{2}\kappa^2 X^2, \tag{3.17}$$

which again only relies on constants of motion.

Solving Poisson's Equation for the Shearing Sheet

In the previous section we have worked hard to calculate orbits in x and y coordinates. The reason for this is simple: we can easily solve Poisson's equation

via Fourier techniques. In particular if we Fourier expand our density response and perturbation in the plane as

$$\begin{aligned}\Sigma^f(\mathbf{x},t) &= \tilde{\Sigma}^f(t)\mathrm{e}^{\mathrm{i}\mathbf{k}(t)\cdot\mathbf{x}},\\ \Sigma^{\mathrm{e}}(\mathbf{x},t) &= \tilde{\Sigma}^{\mathrm{e}}(t)\mathrm{e}^{\mathrm{i}\mathbf{k}(t)\cdot\mathbf{x}},\end{aligned} \tag{3.18}$$

where $\mathbf{k}$ is our Fourier variable, it will source a potential (e.g. Binney & Tremaine, 2008)

$$\psi^f(\mathbf{x},t) = -\frac{2\pi G}{|\mathbf{k}|}\tilde{\Sigma}^f(t)\mathrm{e}^{\mathrm{i}\mathbf{k}(t)\cdot\mathbf{x}}, \tag{3.19}$$

and similarly for ψ^{e}. Note that in the shearing sheet we are in a Cartesian coordinate system and as $\mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}}$ are eigenfunctions of the Laplacian in this coordinate system, different $\mathbf{k}$ will decouple.

We have chosen to make $\mathbf{k}$ a function of time because even if there was no self-gravity or perturbation present, a density profile will naturally shear of its own accord. We could have kept $\mathbf{k}$ independent of time, which would have made the evolution of $\tilde{\Sigma}_1$ more complicated. Instead we pick a coordinate system which naturally deals with the shear. The upshot of this decision, especially when calculating the real space density response, will be discussed in subsection 3.3.2. So in order to calculate the potential term, we first must find how $\mathbf{k}$ evolves for a density profilenaturally shearing density profile.

Consider a collection of particles with the same 'guiding centre', $\bar{x}$, with with different X. Compared to the particle on a circular orbit, particles with $x < 0$ will overtake the circular orbit, while those with $x > 0$ will lag behind (c.f. Equation 3.14). This means that only the circular orbits will experience no shear. Therefore the phase must be constant for a particle moving along a circular orbit, i.e. $\mathbf{k}(t)\cdot\mathbf{x}$ is constant. If we equate the phase at an initial time, t_{i}, and some time later, t, we obtain

$$\mathbf{k}(t)\cdot\mathbf{x} = k_x(t_{\mathrm{i}})x(t_{\mathrm{i}}) + k_y y(t_{\mathrm{i}}) = k_x(t)x(t_{\mathrm{i}}) + k_y y(t). \tag{3.20}$$

For a circular orbit, we know that $\dot{y} = -2Ax$ as there is no radial motion ($X = 0$ in Equation 3.14, so we obtain

$$k_x(t) = k_x(t_i) + 2k_y A(t - t_{\mathrm{i}}), \tag{3.21}$$

We can take one further step and choose to shift our origin of time, so that in the old time coordinate the origin in the new coordinate system occurs at

$$t_{\rm c} = -\frac{k_x(t_{\rm i})}{2k_y A}. \tag{3.22}$$

What we have done is changed our time, so that at times $t < 0$ the spiral is leading and at times $t > 0$ the spiral is trailing. For a density profile, which is initially spread in $\mathbf{k}$, each frequency will have a different origin of time. As each $\mathbf{k}$ decouples, it does not introduce any complications that are more than bookkeeping. From Equation 3.21 we can immediately write

$$k_x(t) = 2Ak_y t \implies |k| = k_y\sqrt{1 + 4A^2t^2}. \tag{3.23}$$

From our intuition about circular orbits, we have calculated $k_x(t)$. What we now need, however, is $\mathbf{k}(t) \cdot \mathbf{x}$ for non circular orbits. As we already have the position of a particle travelling in the shearing sheet, i.e. Equations 3.12 and 3.15, we can write

$$\mathbf{k}(t) \cdot \mathbf{x}(t) = k_y y_0 + \psi(t), \tag{3.24}$$

where we define the phase

$$\psi(t) = 2k_y X \left[At\cos(\kappa t) - \frac{\omega}{\kappa}\sin(\kappa t)\right], \tag{3.25}$$

which, as a sanity test, is zero along a circular orbit. It follows that for two points on an orbit, t and t', the phases are related by

$$\mathbf{k}(t') \cdot \mathbf{x}(t') = \mathbf{k}(t) \cdot \mathbf{x}(t) + \psi(t) - \psi(t'). \tag{3.26}$$

In the integral of Equation 3.2 we need to calculate the derivative of the perturbing potential with respect to $\mathbf{x}$. From Equation 3.19 we have

$$\begin{aligned}\frac{\partial\psi(\mathbf{x}[t'], t')}{\partial\mathbf{x}} &= -i\mathbf{k}(t')\frac{2\pi G}{|\mathbf{k}(t')|}\tilde{\Sigma}(t')\mathrm{e}^{i\mathbf{k}(t')\cdot\mathbf{x}(t')}, \\ &= -i\mathbf{k}(t')\frac{2\pi G}{|\mathbf{k}(t')|}\tilde{\Sigma}(t')\mathrm{e}^{i[\mathbf{k}(t)\cdot\mathbf{x}(t)+\psi(t)-\psi(t')]}.\end{aligned} \tag{3.27}$$

Apart from $\tilde{\Sigma}_1(t')$, which is ultimately what we will solve Equation 3.2 to obtain, we have now calculated all of this is a function of time along an orbit.

Shearing Sheet DF

We now need to calculate the second term in the integrand in Equation 3.2, to do this we first need to find a suitable DF for the shearing sheet. Due to Jeans' theorem, any equilibrium DF is solely a function of integrals of motion. One can then posit that a suitable choice for the shearing sheet's equilibrium DF is simply,

$$F_0 = \mathcal{F}\mathrm{e}^{-H_x/\sigma^2}, \tag{3.28}$$

where we have two free parameters: $\mathcal{F}$, the normalisation that will set the density of the shearing sheet, and σ, that will set the velocity dispersion. The derivatives of this with respect to $\mathbf{p}$, which are needed for Equation 3.2, are

$$\begin{aligned} \frac{\partial F_0}{\partial p_x} &= -\frac{p_x}{\sigma^2}F_0, \\ \frac{\partial F_0}{\partial p_y} &= 2\Omega\frac{(x-\bar{x})}{\sigma^2}F_0, \end{aligned} \tag{3.29}$$

where $v_\phi = p_y$ and we have used Equation 3.14.

This DF generates a uniform surface density that is related to $\mathcal{F}$ by,

$$\Sigma_0 = \int dp_x dv_\phi f_0 = \frac{4\pi\sigma^2 B\mathcal{F}}{\kappa}. \tag{3.30}$$

As we will want to compare our shearing sheet calculation to those for the Mestel disc we can fix $\mathcal{F}$ by forcing the condition that

$$\Sigma\left(\tilde{R}\right) = \Sigma_0, \tag{3.31}$$

where $\Sigma\left(\tilde{R}\right)$ is the surface density of the Mestel Disc at $\tilde{R}$, i.e.

$$\Sigma_0 = \frac{\xi v_c^2}{2\pi\tilde{R}}. \tag{3.32}$$

We can also make the radial velocity distribution the same for the shearing sheet and the Mestel disc. The Mestel disc's DF radial velocity, v_R, dependence is Gaussian with dispersion σ_R. As the transformation into the shearing sheet's coordinate frame is a rotation from the Mestel disc's inertial frame, the radial velocity in the sheet, p_x, is the exactly equivalent to v_R. As the shearing sheet's DF, Equation 3.28,

is Gaussian in p_x to make equivalent to the Mestel disc we choose $\sigma = \sigma_R$, so that the radial velocity dispersion in both systems in exactly the same.

Unfortunately, we cannot make an exact comparison between the Mestel disc and the shearing sheet's DF: the Mestel disc has a gamma distribution in angular velocity, $v_\phi^{\rm d}$, whereas the shearing sheet's DF is assumed Gaussian in $v_\phi^{\rm s}$ (we use the superscript to make explicit which geometry we are referring to, d for the disc and s for the sheet). Although there is no way to match each of the higher moments of these DFs, we can, however, ensure that the ratio of the second moments are the same for the disc and the sheet, i.e. we want the ratio of the dispersion to satisfy

$$\frac{\sigma_\phi^{\rm d}}{\sigma_R} = \frac{\sigma_\phi^{\rm s}}{\sigma_x}, \tag{3.33}$$

where $\sigma_\phi^{\rm d}$ ($\sigma_\phi^{\rm s}$) is the dispersion in the disc's angular velocity (sheet's angular velocity). For the Mestel disc $\kappa = \sqrt{2}\Omega$. Therefore as $\kappa^2 = -4\Omega B$ we must set $\Omega/-B = \frac{2}{\sigma^2}$ for the sheet to have the same shear rate as the Mestel disc. Integrating over the DFs to calculate the relevant moments, we obtain

$$1 + q - 2\left[\frac{\Gamma\left(\frac{q}{2}+1\right)}{\Gamma\left(\frac{q+1}{2}\right)}\right]^2 = \frac{1}{2}, \tag{3.34}$$

we have used $q = v_c/\sigma_R - 1$. The first two terms on the LHS of Equation 3.34, $1 + q$, are due to the expectation of $(v_\phi^{\rm d})^2$, whereas the final term on the LHS is the expectation of $v_\phi^{\rm d}$, i.e. the mean, squared. By numerically plotting the first two terms in this expression, we see that

$$\lim_{q\to\infty}\left[1 + q - 2\left[\frac{\Gamma\left(\frac{q}{2}+1\right)}{\Gamma\left(\frac{q+1}{2}\right)}\right]^2\right] = \frac{1}{2} \tag{3.35}$$

so that this equation is solved as $q \to \infty$, or as $\sigma_R \to 0$. Obviously, for a finite active fraction, ξ, we cannot take $\sigma_R \to 0$ as the disc would be linearly unstable via the Toomre instability criteria Toomre (1964) Despite this and due to the weak dependence of the LHS on q, a cold yet stable disc will approximately satisfy Equation 3.33: for example a disc with $\sigma_R = 0.25$ ($Q = 1.3$) the relative difference in ratios is 1.3%, while for a disc with $\sigma_R = 0.28$ ($Q = 1.5$) the relative difference is

6.0%, where Q is the Toomre Q parameter Equation 3.41. In subsection 3.4.1 we will test the agreement between the sheet and the Mestel disc in the limit that $\sigma \to 0$.

The Linearised CBE

We are now in a position to evaluate the integrals in Equation 3.2. From the previous sections where we have calculated the two derivatives, Equation 3.27 and 3.29, required to solve Equation 3.2. Putting this together explicitly gives

$$f(\mathbf{x},\mathbf{p},t) = 2\pi \mathrm{i} G\mathcal{F}\frac{\mathrm{e}^{-H_x/\sigma^2}}{\sigma^2}\mathrm{e}^{\mathrm{i}\mathbf{k}(t)\cdot\mathbf{x}(t)}\int_{t_\mathrm{i}}^{t} dt' \left[\left(\tilde{\Sigma}^f(t') + \tilde{\Sigma}^\mathrm{e}(t')\right)\mathrm{e}^{\mathrm{i}[\psi(t')-\psi(t)]}\frac{2At'\kappa X\sin(\kappa t') + 2\Omega X\cos(\kappa t')}{\sqrt{1+4A^2t'^2}}\right] \tag{3.36}$$

where we have explicitly integrated over unperturbed orbits. We can now close this expression by integrating both sides over $\mathbf{p}$. Although this requires some tedious algebra, it requires no new physics (Binney, 2020). Once this is done we arrive at the JT equation - the equation at the heart of Julian & Toomre (1966)'s study of the shearing sheet -

$$\tilde{\Sigma}_1(t) = \int_{t_\mathrm{i}}^{t} dt' K_\mathrm{JT}(t,t')\left[\tilde{\Sigma}_\mathrm{e}(t') + \Sigma_1(t')\right] \tag{3.37}$$

where both the kernel and the lower bound of the integral depend on $\mathbf{k}$. As with Equation 2.16, we sweep the complications involved with solving this linear Volterra equation into the JT kernel

$$K_\mathrm{JT}(\mathbf{k},t,t') = 4\mathbf{c}(t,t')\cdot\mathbf{b}(t,t')\frac{\exp\left(-0.572Q^2|\mathbf{b}(t,t')|^2\right)}{\sqrt{1+4A^2t'^2}}, \tag{3.38}$$

where $\mathbf{b}$ and $\mathbf{c}$ are vectors with magnitudes $|\mathbf{b}|$ and $|\mathbf{c}|$ respectively. Using the abbreviations $C \equiv \cos(\kappa t)$, $C' \equiv \cos(\kappa t')$ and similarly for S and S' with sine, the coefficents of the vectors in JT kernel are

$$\begin{aligned} b_x &= \frac{k_y}{k_\mathrm{crit}}\left[A(t'S' - tS') + (\Omega/\kappa)(C' - C)\right], \\ b_y &= \frac{k_y}{k_\mathrm{crit}}\left[A(t'C' - tC') - (\Omega/\kappa)(S' - S)\right], \\ c_x &= -At'C' + (\Omega/\kappa)S', \\ c_y &= At'S' + (\Omega/\kappa)C'. \end{aligned} \tag{3.39}$$

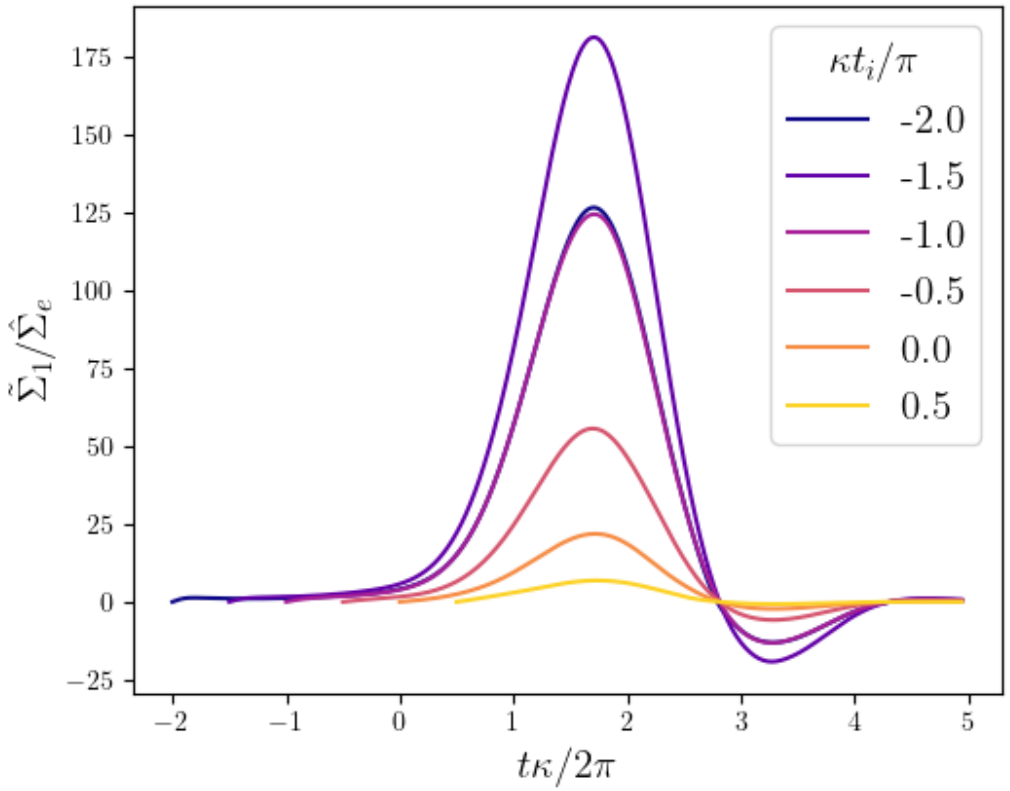


Figure 3.3: The evolution of the density response to an impulse perturbation. As we encode the initial k_x using t_i, more negative t_i corresponds to an initially more tightly wound leading spiral. By taking this choice of t, all waves pass from leading to trailing at $t = 0$ and a maximum density response at the same t. Note that the maximum density response is for $\kappa t_\mathrm{i} = -3\pi/2$.

In these expressions we introduce two new parameters that depend on constants of motion and parameters in the DF, namely the critical wavenumber

$$k_\mathrm{crit} = \frac{\kappa^2}{2\pi G\Sigma_0}, \tag{3.40}$$

and the Toomre Q parameter (Toomre, 1964)

$$Q \equiv \frac{\kappa\sigma}{3.36G\Sigma_0}. \tag{3.41}$$

Note that as $Q^2|\mathbf{b}(t,t')|^2 \propto \sigma^2$ in the JT kernel and that it is independent of Σ_0.

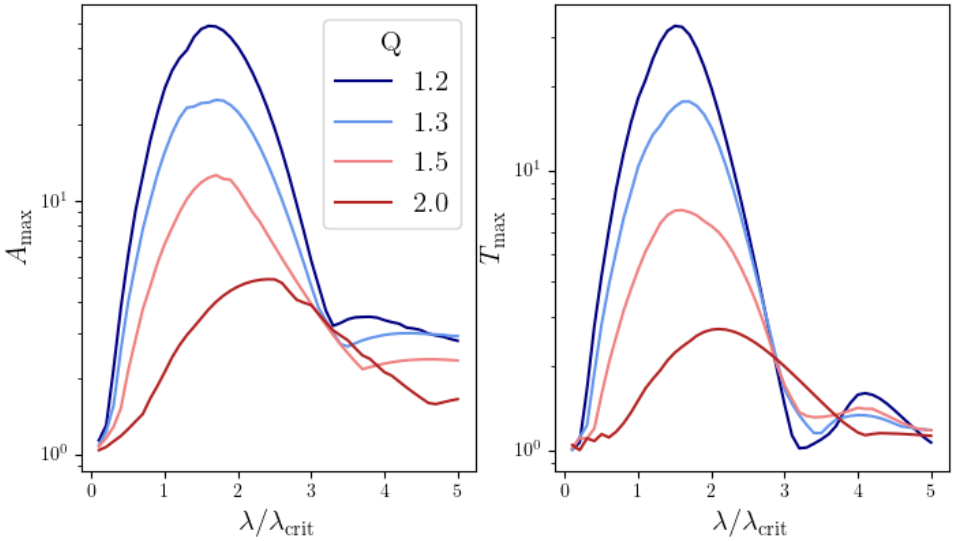


Figure 3.4: The figure shows two different metrics that quantify the strength of amplification due to swing arm amplification as a function of $\lambda/\lambda_{\rm crit} = k_{\rm crit}/k$ for different Q. *Left Panel:* Shows the evolution of $A_{\max}$, which is the ratio of the maximum response with self gravity to without it, which highlights the importance of self-gravity - especially in cold discs. *Right Panel*: Shows the evolution of $T_{\max}$, which is the ratio of the maximum response of the disc when stimulate by a leading wave to trailing wave (i.e. k_x, and hence $t_{\rm i}$, only differ by a sign). This highlights the effectiveness of swing arm amplification in amplifying leading waves as they unwind to trailing waves. Although $T_{\max}$ is not the same quantity as that plotted on the y-axis of Fig. 7 of Toomre (1981), it is designed to measure the same effect (c.f. Binney, 2020).

3.2.2 Amplification by the Shearing Sheet

The study the evolution of the shearing sheet we initially solve the JT equation with a perturbation of the form

$$\tilde{\Sigma}_{\rm e}(t) = \frac{\hat{\Sigma}_{\rm e}}{\kappa}\delta(t - t_{\rm i}). \tag{3.42}$$

which reduces the JT equation for chosen $\mathbf{k}$ to

$$\tilde{\Sigma}_1(t, t_{\rm i}) = \hat{\Sigma}_{\rm e} K_{\rm JT}(\mathbf{k}, t, t_{\rm i}) + \int_{t_{\rm i}}^{t} dt' K_{\rm JT}(\mathbf{k}, t, t')\tilde{\Sigma}_1(t, t_{\rm i}), \tag{3.43}$$

where we have been explict in the response's dependence on $t_{\rm i}$, by writing $\tilde{\Sigma}_1(t, t_{\rm i}) = \tilde{\Sigma}_1(t)$. In Figure 3.3 we plot the evolution of $\tilde{\Sigma}_1$ for different initial windings, which is encoded in Equation 3.43 via $t_{\rm i}$. Due to initial times' dependence on k_x the waves all pass from leading to trailing at $t = 0$.

To quantify the strength of the amplification we will define two different metrics: $A_{\max}$ which will characterise the importance of self gravity and $T_{\max}$ which will characterise the amplification as a wave passes from leading to trailing.

To measure the importance of self gravity, we will use the maximum ratio between the self-consistent and test particle density, i.e.

$$A_{\max} = \max_{t_\mathrm{i}} \left[\frac{\max_t \tilde{\Sigma}_1^{\mathrm{self}}(t, t_\mathrm{i})}{\max_t \tilde{\Sigma}_1^{\mathrm{test}}(t, t_\mathrm{i})} \right], \tag{3.44}$$

where $\tilde{\Sigma}^{\mathrm{self}}$ is the self consistent response, given by Equation 3.43 and $\tilde{\Sigma}^{\mathrm{test}}$ is the test particle response, given by the same equation, but with the integral term ignored. As the JT equation is linear, $A_{\max}$ is independent of $\hat{\Sigma}_\mathrm{e}$. As such we set it to unity in our calculations. To study the amplifying effects of passing from leading to trailing we use the metric

$$T_{\max} = \max_{t_\mathrm{i}} \left[\frac{\max_t \tilde{\Sigma}_1(t, t_\mathrm{i})}{\max_t \tilde{\Sigma}_1(t, -t_\mathrm{i})} \right], \tag{3.45}$$

where we assume that $t_\mathrm{i} < 0$, such that that numerator corresponds to an initially leading wave. Maximising with respect to t_i means that we pick the largest density response, i.e. using the lines shown in Figure 3.3 we would pick $\kappa t_\mathrm{i} \approx -3\pi/2$.

In Figure 3.4 we plot the evolution of these two metrics for different wavelengths. The left-hand panel shows the importance of self-gravity and gives similar results to Fig. 6 of Binney (2020). The right-hand panel highlights the effectiveness of swing arm amplification in amplifying leading waves as they unwind to trailing waves. Although $T_{\max}$ is not the same quantity as that plotted on the y-axis of Fig. 7 of Toomre (1981), it is designed to measure the same effect (c.f. Binney, 2020) and similarly shows the importance of swing amplification.

3.3 Comparison between Shearing Sheet & Linear Response Theory

Our ultimate aim is to reproduce the amplification measurements of Toomre (1981) and Binney (2020) by using our linear response calculations that do not rely on

the shearing sheet approximation. Before we do that we will compare the shearing sheet's response to a disc's response to an external perturbation.

To achieve this aim we will first compare the shearing sheet in the axisymmetric limit, which corresponds to sending $k_y \to 0$ for the shearing sheet and to the axisymmetric harmonic of the disc $\ell = 0$. To calculate the disc's response using our linear response machinery, we will then simply just use the $\ell = 0$ kernel. This will allow us to test the agreement of the two approaches, without having to worry about the complications introduced by the $k_y \leftrightarrow \ell$ correspondence and that k_x is a function of time. Once we have done that, we will then compare the non-axisymmetric responses, where $k_y \neq 0$ and $\ell \neq 0$ for the sheet and disc respectively. Not only will we have to include the previously highlighted complications, but we will also have to modify the JT equation to account for the multiple k_x in our perturbation.

Comparison between Mestel and Sheet set up

Before we proceed with comparisons between the shearing sheet and the Mestel disc, we first highlight the differences between the two systems, and the steps we take to ensure, where possible, a fair comparison. The geometrical differences required a different potential between the two systems: for the sheet the effective potential is quadratic for all orbits regardless of their p_x, whereas for the Mestel disc this is only the case for orbits with small radial motion. As a result we expect better agreement for cool discs, where the potential in which particles move in is better approximated by a quadratic.

Due to the nature of the two potentials, the frequencies in the disc will be a function of radius, whereas for the sheet they are constant. Despite this, we can at least ensure that the relationship between the two frequencies are the same for the two systems, namely $\kappa = \sqrt{2}\Omega$. In order to make the fairest, although certainly not perfect comparison, we will take $\kappa = \sqrt{2}\Omega(\tilde{R})$ in our sheet calculations. As with the frequencies, the density for the Mestel disc is a function of radius, whereas for the shearing sheet it is constant. For the fairest possible comparison, we ensure

that the radius at which we place the sheet, $\tilde{R}$, has the same density as that given by the Mestel disc, i.e. we pick the density of the shearing sheet, Σ_0, as

$$\Sigma_0 = \frac{\xi v_c^2}{2\pi \tilde{R}}. \tag{3.46}$$

3.3.1 Axisymmetric Comparison

JT Kernel in the Axisymmetric Limit

We will now compare the axisymmetric limit of the shearing sheet to the axisymmetric response of the full disc. Making the link between a non-axisymmetric disc and the shearing sheet is complicated by the question of how to compare k_y for the sheet to ℓ for the disc. One can temporarily simplify this problem, among others, by comparing the axisymmetric response of the disc, for which $\ell = 0$, to the axisymmetric response of the sheet, for which $k_y \to 0$. To make this comparison, the asymmetric formulation of the JT kernel must be obtained.

The limit $k_y \to 0$ must be done judiciously, as $k_x = 2Atk_y$ must remain finite. This necessitates us taking $k_y \to 0$ while taking $|t| \to \infty$, such that $k_x t$ remains finite. Making use of

$$\begin{aligned} \cos\left[\kappa\left(t-t'\right)\right] &= \cos(\kappa t)\cos(\kappa t') + \sin(\kappa t)\sin(\kappa t') \\ \sin\left[\kappa\left(t-t'\right)\right] &= \sin(\kappa t)\cos(\kappa t') + \cos(\kappa t)\sin(\kappa t') \end{aligned} \tag{3.47}$$

we can manipulate the trigonometric terms and obtain the axisymmetric form of the JT kernel (Binney, 2020),

$$\mathcal{K}_{\rm JT}(k_x, t-t') = \frac{|k_x|}{k_{\rm crit}} \sin\left[\kappa(t-t')\right] e^{(\cos([\kappa(t-t')]-1)\chi}, \tag{3.48}$$

where we have used used $\chi = (k_x \sigma_R/\kappa)^2$.

This allows the JT equation to be written in a form where the similarity to Equation 2.16 is self evident,

$$\tilde{\Sigma}_1(k_x, t) = \int_{-\infty}^{t} \kappa dt' \, \mathcal{K}_{\rm JT}(k_x, t-t') \left[\tilde{\Sigma}_e(k_x, t') + \tilde{\Sigma}_1(k_x, t')\right]. \tag{3.49}$$

We also highlight that in this limit k_x is constant, unlike for the non-asymmetric case.

The LSK Relationship from the JT Equation

As a slight digression we can use this expression to obtain the Lin-Shu-Kalnajs (LSK) relation for the disc (Equation 4.1) (Lin & Shu, 1966; Kalnajs, 1965). In order to obtain a dispersion relationship we will follow the same approach that led us to an expression for the response matrix in chapter 2, namely we assume that $\Sigma_1(k_x,t) = \hat{\Sigma}_1(k_x)\exp(-\mathrm{i}\omega t)$ and that there is no external perturbation. This gives

$$1 = \int_0^\infty \mathcal{K}_{\mathrm{JT}}(k_x,\tau)\mathrm{e}^{-\mathrm{i}\omega\tau}\kappa d\tau, \tag{3.50}$$

where we have taken a complex conjugate as the JT kernel in the axisymmetric limit is real. Using

$$\mathrm{e}^{z\cos(\theta)} = \sum_{n=-\infty}^{n=\infty} I_n(z)\cos(n\theta), \tag{3.51}$$

where I_n are the modified Bessel functions of the first kind, in Equation 3.48, we obtain

$$1 = \frac{|k|\mathrm{e}^{-\chi}}{k_{\mathrm{crit}}}\sum_{n=-\infty}^{n=\infty} I_n(\chi)\frac{(n+1)\kappa^2}{(n+1)^2\kappa^2-\omega^2}. \tag{3.52}$$

where we have used $|k_x| = |k|$ as in the axisymmetric limit $k_y = 0$. After some, not too difficult, algebra we obtain

$$1 = \frac{|k|\mathrm{e}^{-\chi}}{2k_{\mathrm{crit}}}\sum_{n=-\infty}^{n=\infty}\left[I_{n-1}(\chi)-I_{n+1}(\chi)\right]\frac{n}{n^2-s^2}. \tag{3.53}$$

where we have defined the normalised frequency, $s = \omega/\kappa$. Finally using the recurrence relationship for the modified Bessel functions, $I_{n-1}(\chi) - I_{n+1}(\chi) = 2nI_n(\chi)/\chi$, we obtain the LSK dispersion relationship (c.f. Equation 4.1)

$$0 = 1 - \frac{|k|}{k_{\mathrm{crit}}}\mathrm{e}^{-\chi}\sum_{n\geq 1}\frac{I_n(\chi)}{\chi}\frac{1}{1-(s/n)^2}. \tag{3.54}$$

Density Response of the Sheet and Disc in the Axisymmetric Limit

After that aside, we will now return to our task of comparing the sheet's and disc's response to an axiysmmetric perturbation. To compare the JT calculation against our linear response calculation we employ an impulsive perturbation of the form

$$\Sigma_e(x,y,t) = \delta(t)\frac{\hat{\Sigma}_e}{\sqrt{2\pi\Delta_x^2}}\mathrm{e}^{-\frac{x^2}{2\Delta_x^2}}, \tag{3.55}$$

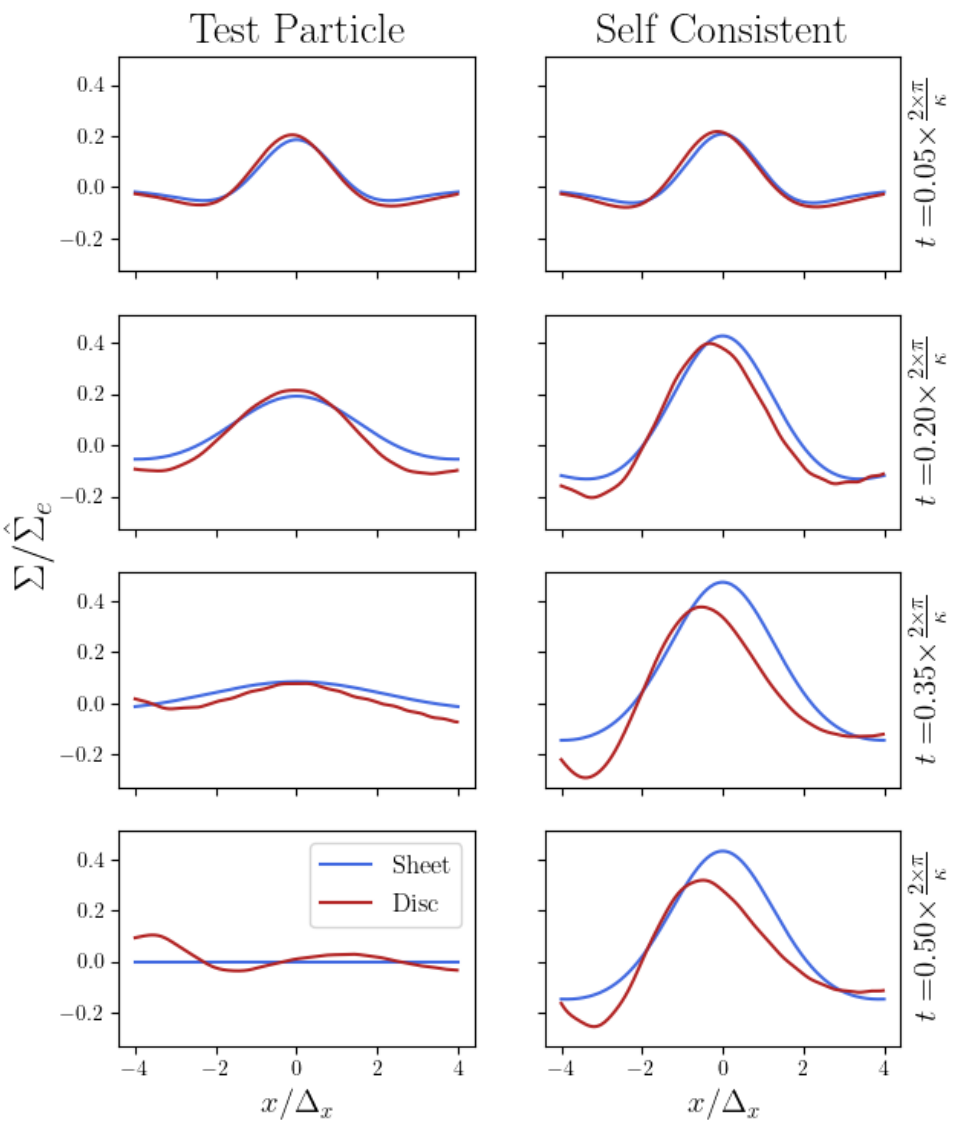


Figure 3.5: This plot compares the density response of the shearing sheet (blue line) and full disc (red line) to an impulsive perturbation. The left column shows the test particle response, while the right column shows the self consistent response. The time at which each density snapshot is taken is on the right side of the plots. In the beginning there is close agreement between the two different calculations, while at later times the two calculations start to differ. Considering the large conceptual difference in the approach taken for each calculation - a shearing sheet is a poor approximation for a whole disc - the rough order of magnitude agreement between the two approaches is encouraging.

where Δ_x sets the extent of the disturbance and $\hat{\Sigma}_e$ is the size of the initial disturbance. The Fourier transform of this perturbation is

$$\tilde{\Sigma}_e(k_x, k_y, t) = \sqrt{2\pi}\delta(t)\hat{\Sigma}_e \mathrm{e}^{-\frac{k_x^2\Delta_x^2}{2}}. \tag{3.56}$$

Due to the simple time dependence we can further reduce Equation 3.49 to

$$\tilde{\Sigma}_e(k_x, k_y, t) = \sqrt{2\pi}\hat{\Sigma}_e\kappa\mathcal{K}(t)\mathrm{e}^{-\frac{k_x^2\Delta_x^2}{2}} + \int_0^t \kappa dt' \mathcal{K}_{\mathrm{JT}}(t-t')\tilde{\Sigma}_1(t'), \tag{3.57}$$

where we have set the lower limit of the integral to zero as we assume that the sheet is in equilibrium before it is perturbed at $t = 0$. The first term on the right side of the above equation corresponds to the test particle response, whereas the second term contains information about the self consistent response. If only the test particle response is required we set the self consistent term to zero.

Once we have solved Equation 3.57 the inverse Fourier transform must be taken with respect to k_x and k_y to obtain the density, Σ. As $k_y \to 0$ for the axisymmetric case, the inverse Fourier transform of k_y multiplies the density by a constant.

For the comparisons to linear response, we choose $\Omega = 1/8$ and $\sigma_R = 0.24$ with an active fraction $\xi = 1/2$. When numerically integrating Equation 3.57 we take a time step $\Delta T = 2\pi/(100\kappa)$. We take $\Delta_x = 1/2$ to ensure a local perturbation, while remaining resolvable with the number of basis functions.

For the full disc the equivalent of Equation 3.55 for the linear response calculation is simply

$$\rho^e(R, t) = \delta(t)\frac{\hat{\Sigma}_e}{\sqrt{2\pi\Delta_x^2}}\mathrm{e}^{-\frac{(R-\tilde{R})^2}{2\Delta_x^2}}. \tag{3.58}$$

In order to compare between our linear response and the shearing sheet we must identify a patch of the disc that satisfies two conditions: that the patch is sufficiently far from the centre so that the radial scale of the perturbation is much less that the co-rotation radius of the patch. As our outer taper is at $R_{\mathrm{o}} = 11.4$ we choose $\tilde{R} = 8$ (hence the choice of $\Omega = 1/8$ in the shearing sheet calculations).

To generate the kernel we use 48 basis functions with their upper radius $R_{\mathrm{ka}} = 15$. When calculating $\hat{\psi}^{(q)}_{\mathbf{m}}(\mathbf{J})$ we expand on a uniform grid of 250 points in (R_+, R_-)

space. As we are in the axisymmetric limit, we need only consider the $\ell = 0$ harmonics. We take the same time step as that used for the shearing sheet, namely $\Delta T = 2\pi/(100\kappa)$. When solving Equation 2.16 with an impulsive perturbation we analytically evaluate the time integral over the perturbative term to obtain,

$$B_p(t) = \mathcal{K}_{pq}(t)A^q + \int_0^\infty dt' \mathcal{K}_{pq}(t-t')B^q(t'), \tag{3.59}$$

where A^q are the expansion coefficients of the perturbation in the Kalnajs basis.

In Figure 3.5 the density response of the shearing sheet (blue line) and full disc (red line) to an impulsive perturbation is plotted. The left column shows the test particle response, while the right column shows the self consistent response. The time at which each density snapshot is taken is on the right side of the plots. As we have taken $\Omega = 1/8$, the time step between snapshots equates to $15 \times \Delta T = 5.33$. Naively, one might expect the shearing sheet approximation to work well for times in which the disc is yet to learn about itself, i.e. for times shorter than a rotation period. In the beginning there is close agreement between the two different calculations. However once the disc starts to complete a rotation (the final snapshot is taken at $\sqrt{2}/2$ times the full rotational period), we start to see disagreement between the two calculations. Considering the large conceptual difference in the approach taken for each calculation - a shearing sheet is a poor approximation for a whole disc - the rough order of magnitude agreement between the two approaches is encouraging.

This calculation validates the shearing sheet approximation made in Julian & Toomre (1966) in the axisymmetric limit, showing the agreement between the sheet and full disc. There are however limitations for the axisymmetric limit of the shearing sheet: the axisymmetric JT kernel (and subsequent solution of the JT equation) does not admit decaying solutions. As discussed in subsection 4.2.1 the LSK relationship – which is the dispersion relationship from the axisymmetric JT kernel – only admits real or imaginary solutions, but not complex solutions for s, i.e. $s = x$ or $s = \mathrm{i}y$, but not $s = x + \mathrm{i}y$ for $x, y \in \mathbb{R}$. As the LSK relationship is a function of s^2 only, for every solution s, $-s$ will also be a solution and hence a decaying solution cannot exist by itself.

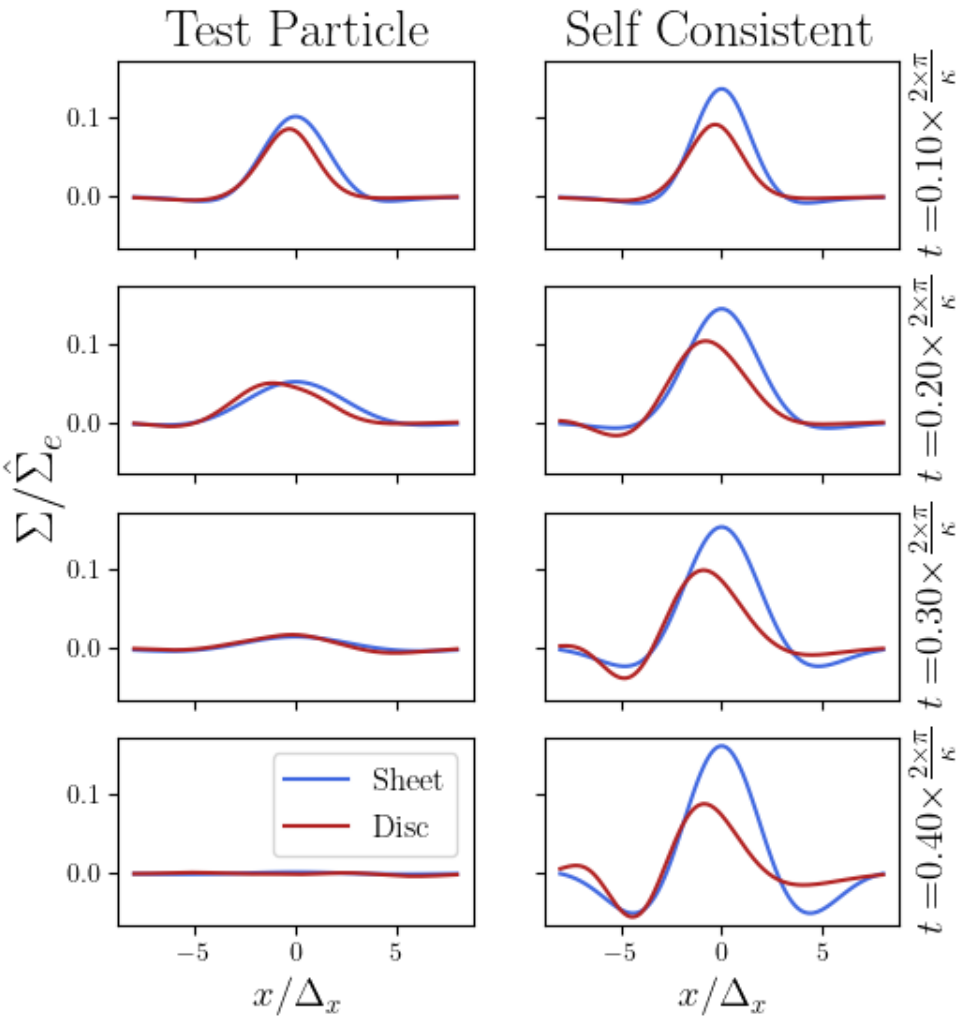


Figure 3.6: This plot is the same as Figure 3.5, however studying a non-axisymmetric response. Although the agreement is not as close for the axisymmetric case, the shearing sheet calculation does reproduce the main features of the linear response calculation and is correct within a factor of order unity.

The next task is a much harder one. The shearing sheet must be compared to the full disc response in the non-axisymmetric limit.

3.3.2 Comparison For Non-Axisymmetric Perturbations

We now compare the shearing sheet approximation to the linear response calculation for non-axisymmetric perturbations. This calculation is made harder by two added complications: the explicit time dependence of k_x in the shearing sheet;

ensuring that the two approaches have the same geometry, i.e. ensuring that k_y and ℓ correspond to one another.

The JT equation up to now has been written in such a form as to have $t = 0$ when $k_x = 0$. This makes the time at which we introduced the perturbation, t_i, dependent on the initial value of k_x. As we will now introduce a perturbation that is explicitly dependent on $\mathbf{k}$, we want to pick an initial time independent of k_x; put another way, we want waves of all wavelength to get underway when the starting gun of the perturbation is fired, rather than having staggered starts for which they are handicapped by their wavelength. To do this we introduce a new time variable, τ, and relate it to t via $\tau = t - t_\mathrm{i}$, so that our perturbation is introduced at $\tau = 0$. This will result in a time translated JT equation,

$$\tilde{\Sigma}_1(\mathbf{k},\tau) = \int_0^\tau \kappa d\tau' K_\mathrm{JT}(\mathbf{k},\tau + t_\mathrm{i},\tau' + t_\mathrm{i}) \left[\tilde{\Sigma}_e(\mathbf{k},\tau') + \tilde{\Sigma}_1(\mathbf{k},\tau')\right], \tag{3.60}$$

where we have made explicit the dependence on $\mathbf{k}$ in the density terms and the dependence of the kernel on $t_\mathrm{i} = k_x(\tau = 0)/2Ak_y$.

What complicates the matter further is that k_x grows linearly with 'time' i.e. $k_x(\tau) = k_x(0) + 2A\tau k_y$ (cf. Equation 3.23). There are two places where this effect must be accounted for: first, wherever the perturbation is dependent on k_x then $k_x(\tau)$ must be used; second, when solving the JT equation the change in k_x can be ignored - apart from when the previous complication has arisen - as it is already included in the JT kernel and because the mapping of a $k_x(\tau) \to k_x(\tau + \delta\tau)$ is one-to-one. However, if the shift of k_x is not included, when the response is Fourier transformed back to real space it must be accounted for and once again $k(\tau)$ must be used.

To compare the shearing sheet to the linear response we will again consider an impulsive perturbation of the form,

$$\Sigma_\mathrm{e}(\mathbf{x},\tau) = \frac{\hat{\Sigma}_\mathrm{e}}{\sqrt{2\pi\Delta^2}} \mathrm{e}^{-\frac{x^2}{2\Delta^2}} \cos(k_\mathrm{i} y)\delta(\tau), \tag{3.61}$$

for which the Fourier transform is,

$$\tilde{\Sigma}_\mathrm{e}(\mathbf{k},\tau) = \pi\hat{\Sigma}_\mathrm{e} \mathrm{e}^{-\frac{k_x^2\Delta^2}{2}} \left[\delta(k_y - k_i) + \delta(k_y + k_i)\right] \delta(\tau). \tag{3.62}$$

When integrating the JT equation we once again integrate the perturbative term by hand. As the density response must be real, we need only study the evolution of $k_y > 0$ as $\tilde{\Sigma}_{\rm e}(-\mathbf{k},\tau) = \tilde{\Sigma}^*_{\rm e}(\mathbf{k},\tau)$.

For the comparisons to linear response, we choose $\tilde{R} = 8$ and $\sigma_R = 0.24$ with an active fraction $\xi = 1/2$. When numerically integrating the JT equation we take a time step $\Delta T = 2\pi/(100\kappa)$. We take $\Delta_x = 0.5$. Our choice of $k_{\rm i}$ will be constrained by our choice of ℓ in our basis expansion.

In the full disc case the analog for k_y is ℓ (note that they both decouple in their respective equations, remaining constant during the evolution). The condition that the two wave in the two geometries have the same wavelengths means that we can relate them via

$$k_y = \frac{\pi}{\tilde{R}\sin(\pi/\ell)} \approx \frac{\ell}{\tilde{R}}. \tag{3.63}$$

The regime in which the shearing sheet is valid corresponds to the regime where $\pi/\ell \ll 1$. Note, as $\Omega \sim 1/\tilde{R}$, then $k_y/k_{\rm crit} = \xi\ell/2$, independent of $\tilde{R}$.

For our comparisons to the shearing sheet, we take $\ell = 8$. As a result we take $k_{\rm i} = 2$. To calculate the basis functions and the evolution kernel, we take the same parameters as section 3.3.1.

By construction, the shearing sheet is rotating with angular frequency $\Omega(\tilde{R})$. Therefore to compare with the disc's linear response calculation we must rotate our coordinate system. As our sheet's perturbation is centred on $y = 0$, the density response of the sheet is also centred on $y = 0$ (i.e. the y dependence is $\cos(k_{\rm i} y)$). Therefore, the density response of the disc along the line that goes through the centre of the sheet is

$$\rho(R, \phi = \Omega t, t) = {\rm e}^{{\rm i}\ell\Omega t} B_p(t)\rho^p(R,\phi) + {\rm c.c.} \tag{3.64}$$

In Figure 3.6 we plot the density response for the shearing sheet and the linear response calculation. In it we plot the response both including and not including self-gravity. The agreement between the two geometries is far from perfect. In particular the amplitude of the self-consistent response is off by a factor of ~ 1.5.

However, given the large conceptual difference in the two approaches there is broad agreement, which is not much worse than the axisymmetric case.

The role of the shearing sheet has been to give an order of magnitude estimate of the importance of swing arm amplification. The slight disagreement between the disc and sheet's density response, suggests that the sheet is a reasonable approximation to the disc. Ultimately given the assumptions that the shearing sheet is based on, the agreement is never going to be perfect, we merely need it to be 'good enough' (Winnicott, 1991). The role of the following sections will be to refine the estimates of amplification by swing arm amplification by removing the reliance on the shearing sheet approximation.

3.4 The Strength of Swing Amplification in Discs

We are now in a place where we can measure the strength of swing amplification in a disc without relying on the shearing sheet approximation. Unlike in the shearing sheet case $\lambda/\lambda_{\rm crit}$ is not a continuous function. Instead it is discrete, inversely depending on ℓ. If fact, for the Mestel disc this parameter can be easily written as $\lambda/\lambda_{\rm crit} = 2/(\xi\ell)$.

In order to calculate the plots of subsection 3.2.2 using our linear response machinery, we use the standard tapered Mestel disc with either $Q = 1.3$ or $Q = 1.5$ and $\xi = 0.5$. We calculate $\hat{\psi}_{\mathbf{m}}(\mathbf{J})$ on a uniform grid with 251 points between $R = 0$ and $R = 15$ and for indices $-4 \leq m_1 \leq 4$ and $0 \leq m_2 = \ell \leq 14$. For the kernel we use 500 uniformly spaced time steps with a distance of 0.3. The first forty Kalnajs basis functions are used. When looking at the $\ell = 1$ response it is worth noting that the indirect term in the Hamiltonian, that due to the bulk motion of the disc, has not been included (e.g. Murali, 1999).

In order to study the response we must generate either leading or trailing spiral waves in our disc. Externally imposing such waves in a self-consistent manner is a complex task, instead we will use an external perturbation to generate either leading or trailing waves. To do so we will perturb our disc with a delta function perturbation at a radius $\tilde{R}$ that rotates with the circular speed, $\Omega(\tilde{R})\tilde{R}$, in either

the clockwise or anti-clockwise direction. In order to study swing amplification, and not the effect of continually perturbing a disc with a rotating perturbation, the perturbation is removed after a quarter of a rotation, i.e. at $t = 0.25 \times 2\pi/\Omega(\tilde{R})$. Mathematically speaking this means that the perturbation can be written as

$$\rho^{\mathrm{e}}(R,\phi,t) = \begin{cases} \frac{M\delta(R-\tilde{R})\delta(\phi \pm \ell\Omega(\tilde{R})t)}{\tilde{R}}, & 0 \leq t \leq \frac{2\pi}{4\Omega(\tilde{R})}, \\ 0, & t > \frac{2\pi}{4\Omega(\tilde{R})}, \end{cases} \tag{3.65}$$

where M is the mass scale of the perturbation and the inverse factor of $\tilde{R}$ has come from the representation of a delta function in cylindrical coordinates. As a delta function decomposes into a sum of angular harmonics with equal weight, it allows for an easy comparison of the amplification of each harmonic. The angular expansion is curtailed at $\ell = 14$ due to the minimal amplification at higher harmonics (c.f. Figure 3.4).

To calculate the response of the shearing sheet we follow the same approach as that outlined in subsection 3.3.2. While the perturbation is 'turned on' we can consider it as a succession of delta function kicks, which are then turned off once it is removed (c.f. Binney (2020)). Once the JT equation is solved, we invert the Fourier transform to calculate the maximum density response.

3.4.1 Comparison in the Cold Disc Limit

The aim of section 3.4 is to compare amplification in the shearing sheet to the amplification in the closest realistic disc model. Even our most realistic disc model will have several distinct differences to the shearing sheet, namely: a cylindrical, rather than a Cartesian, geometry; a varying, rather than a constant, density and frequency profile; an anharmonic, rather than a harmonic, effective potential in the co-rotating frame; and, finally, a gamma distribution, rather than a Gaussian distribution, in radial velocity. The first two of these differences are fundamentally different between the disc and the sheet, however the second two are not – no matter how contrived it might be, one could construct a disc model that satisfied these two properties. In the limit that $\sigma_R \to 0$, the radial motion will decrease improving the harmonic effective potential approximation and gamma distribution

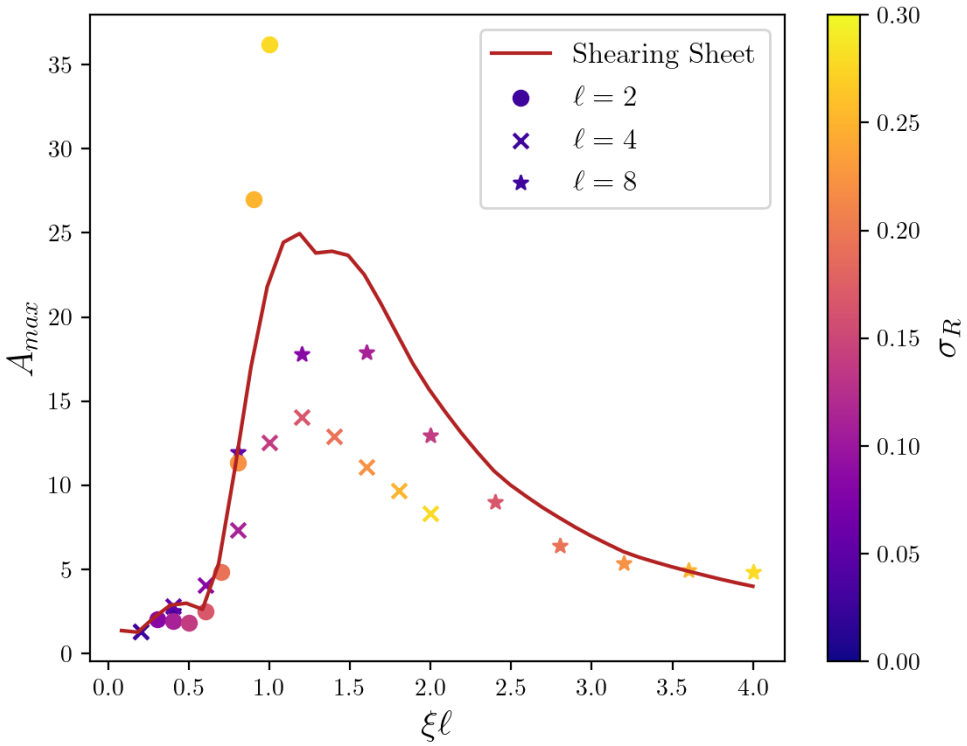


Figure 3.7: The amplification (Equation 3.44) due to self gravity in the cold disc limit. The red line shows the amplification experienced by the shearing sheet, while the points correspond to different harmonics of the disc. We plot as a function of $\xi\ell$ as for the sheet the amplification only depends on $\lambda/\lambda_{\text{crit}} \propto \xi\ell$. The colour of the points corresponds to the disc's temperature. We see good agreement in the cold disc limit for all harmonics. For warmer disc, there is a larger disagreement for the lower angular harmonic (as the geometric approximation is worse) than the higher harmonic.

for the disc and the Gaussian distribution for the sheet will both approach the same distribution, namely a delta distribution. There is, however a catch: to ensure that the disc remains stable as we take $\sigma_R \to 0$, we will have to decrease the active fraction, ξ, as well. This will result in an unrealistic model of a disc. However in this limit, we will be able to test the geometric differences most cleanly.

In Figure 3.7 we compare the role of self-gravity for the shearing sheet (line) and the full disc (dots) for different disc temperatures, where A_{max} is the ratio of absolute values of the maximum density response with and without self-gravity (Equation 3.44). Each mark represents a different ℓ and the colour corresponds to the disc temperature. To calculate the disc's amplification we generated the

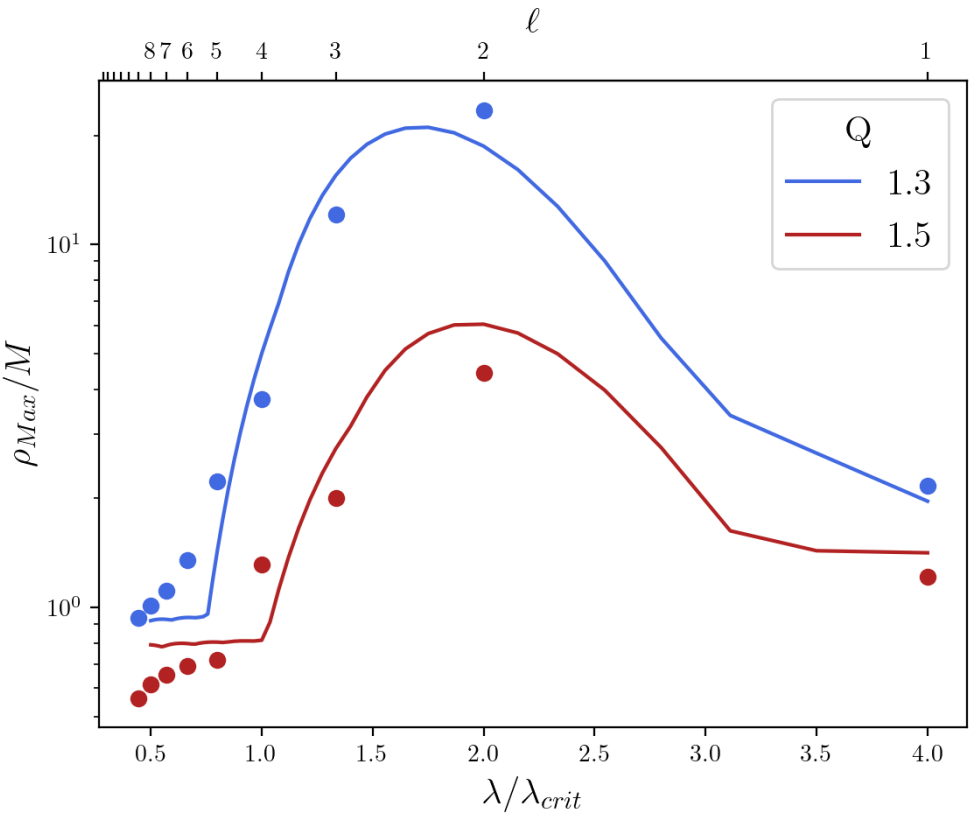


Figure 3.8: The maximum density response of a disc and the shearing sheet to a delta function perturbation for different $\lambda/\lambda_{\rm crit}$, which corresponds to different ℓ for the full disc. The dots represent the measurements from the full disc, whereas the lines represent the response of the shearing sheet. Despite the different approaches, there is surprisingly good agreement between the two methods.

kernels using the same parameters as outlined at the beginning of this section, but with different radial dispersions. We picked the radial dispersion such that $Q \propto \sigma_R/\xi$ remains constant. In Figure 3.7 we took $Q = 1.3$. We scaled the active fraction by a factor of ℓ on the x-axis as for the shearing sheet the amplification depends only on $\lambda/\lambda_{\rm crit} = \xi\ell/2$. We see good agreement in the cold disc limit, but as the temperature increases the predictions of the disc and the sheet diverge. In particular, the lower harmonics have a worse agreement due to the questionable nature of the geometric assumptions at low ℓ.

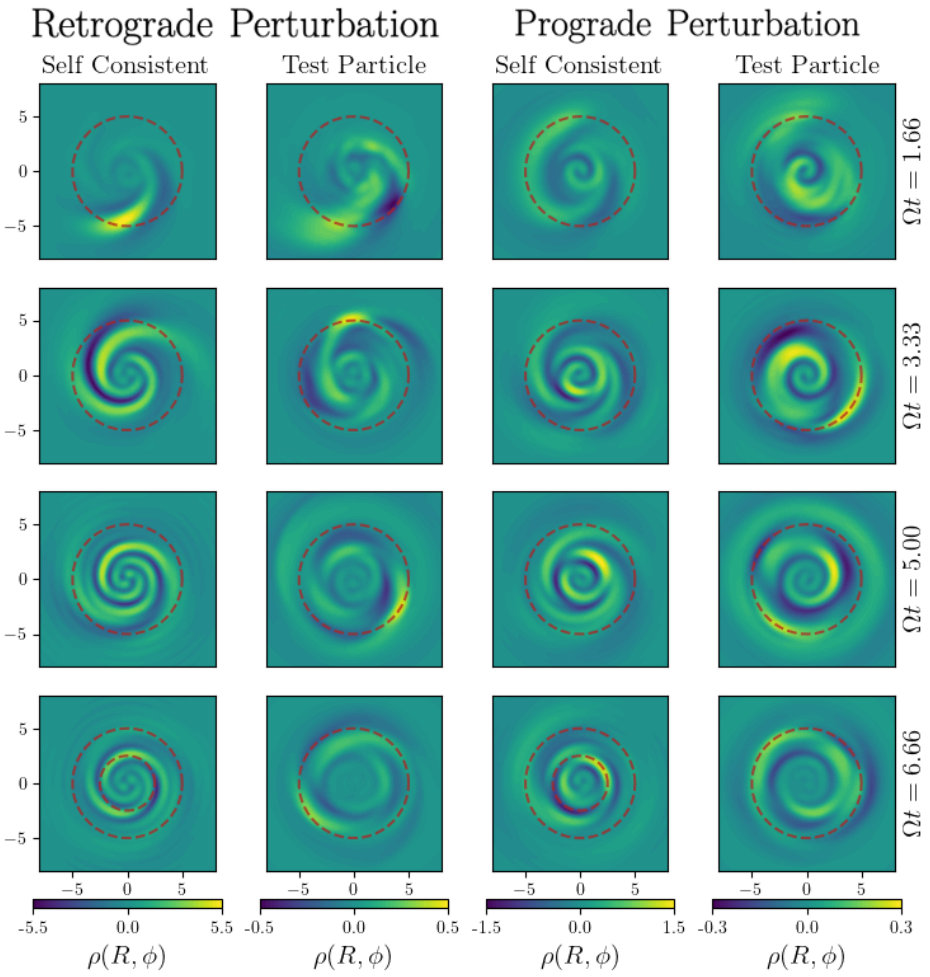


Figure 3.9: The density response of a $Q = 1.5$ disc to a point perturbation rotating at $\tilde{R} = 5$. The dotted red line shows the CR of the perturbation and, when shown, the inner dotted line shows the ILR of the $\ell = 2$ harmonic. The density response is a sum over angular harmonics from $\ell = 0$ to $\ell = 14$. The time at which each snapshot is taken is shown on the right-hand side of the plots in units of the rotational time period at $\tilde{R}$. We see the transition from leading to trailing clearly between the first and second row in the first and third column. From the scale of the colour bars, we can see that initially leading waves, are amplified relative to the trailing waves.

3.4.2 Comparison of Density Response

In Figure 3.8 we plot the maximum density as a function of angular harmonic. That calculated for the shearing sheet is shown as a full line, whereas the maximum density of the disc are points due to the discrete values which $\lambda/\lambda_{\rm crit} \sim 1/\ell$ can take. Rather surprisingly given the different approaches of the two methods, the shearing sheet and linear response calculations show reasonable agreement, validating the conclusions drawn from the shearing sheet in Julian & Toomre (1966).

Rather unsurprisingly, it is the $\ell = 2$ harmonic that shows the greatest amplification via the swing amplification mechanism (and is greatest affected by self-gravity, which will be discussed in the next section). This is not only suggested from Julian & Toomre (1966), but is in accordance with the observation that many galaxies exhibit two-arm spiral structure. To illustrate this point, in Figure 3.9 we plot the density response of all the harmonics where each column corresponds to either a different initial perturbation (leading or trailing) or if self-gravity is included. The time at which each snapshot is shown is on the right-hand side and is expressed as a fraction of the rotation time period at $R = \tilde{R}$. This can be clearly seen in Figure 3.9, where the late time initially leading self-consistent evolution leads to a clear quadrupole term.

In the shearing sheet, with our assumed shear and flat density profile, there is symmetry about $y = 0$. A real galaxy model, or even our highly idealised Mestel disc, will have a density profile that falls with radius. This would suggest that any perturbation will be able to find more recruits with which it can dress itself in its interior, leading to a larger density response at smaller radii. This can be clearly seen in Figure 3.9, where most of the density response is in the interior of the perturber's orbit (the dotted red circle). In the self-consistent case this will result in the centre of mass of the cloud which dresses the perturbation being at $R < \tilde{R}$, such that the overdensity will start to migrate inwards; such an effect can be clearly seen in the motion of the density peak in the self-consistent leading case. By this point the density response has become tightly wound two-arm spiral, and therefore proceeds to the ILR (the inner dotted red line on the bottom row) as expected by LSK and

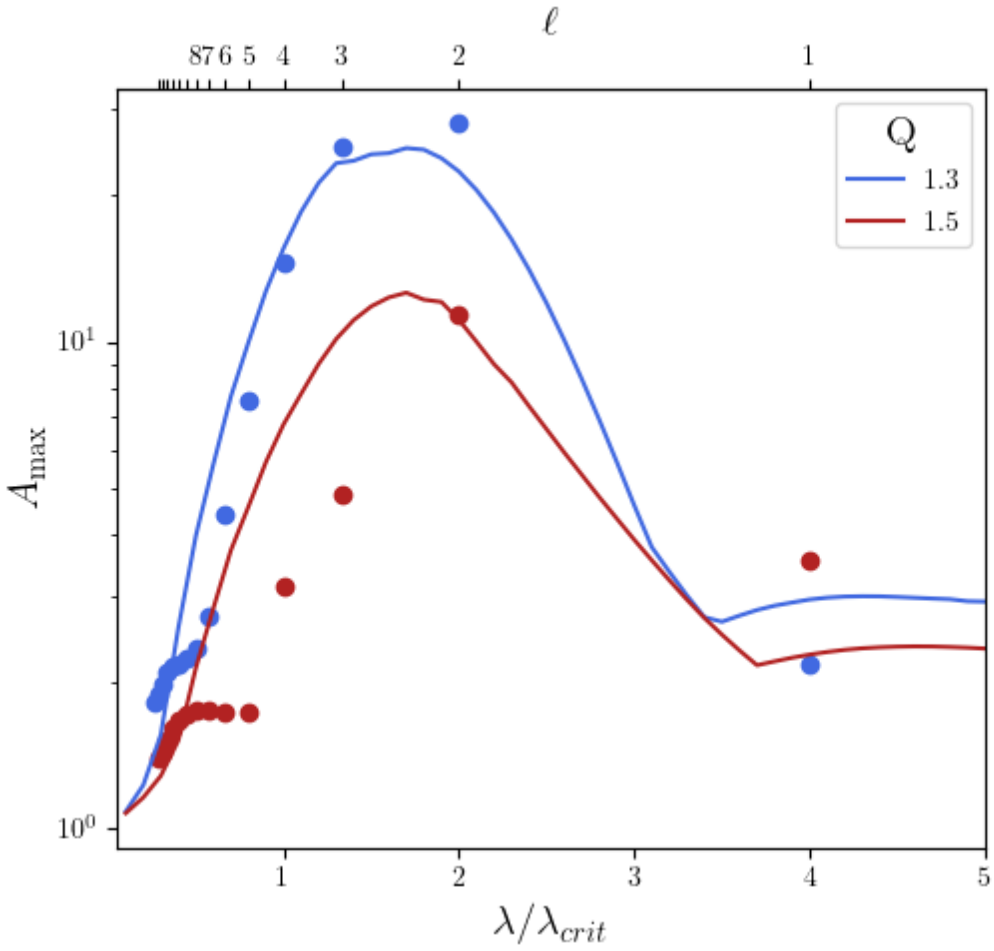


Figure 3.10: The difference that self-gravity plays in swing arm amplification for two different disc temperatures. The dots represent the measurements from the full disc, whereas the lines represent the response of the shearing sheet. For small $\lambda/\lambda_{\text{crit}}$, the region in which the response of the disc can best be approximated by the shearing sheet, there is reasonable agreement. For large $\lambda/\lambda_{\text{crit}}$ the agreement is poor, however the full linear response calculation still suggests that self-gravity is crucial – in the most unfavorable case ignoring self-gravity in a warm disc could lead to an order of magnitude sized error.

Landau damp away. If self-gravity is not included, no such dressing occurs, meaning that the maximum density fluctuation remains slightly interior to the radius of the perturbation. From the scale of the colour bars, we clearly see that a initially leading wave is turned into a much stronger trailing wave than an initially trailing wave.

3.4.3 The Role of Self-Gravity

In Figure 3.10 we compare role of self-gravity for the shearing sheet and the full disc, where $A_{\max}$ is the ratio of absolute values of the maximum density response with

and without self-gravity. The amplification of the shearing sheet, which is calculated using the JT equation, is represented by the solid line. The disc's amplification is plotted using dots as $\lambda/\lambda_{\rm crit} \sim 1/\ell$ is discrete.

Although the disagreement is notable, it does not affect the original conclusions drawn from Figure 3.4: self-gravity is important. Even when studying the response of a warm disc, Figure 3.10 suggests self-gravity can amplify the response by an order of magnitude more than the test particle response.

3.5 Conclusion

In this chapter we compared the effect of swing amplification in a full disc and the shearing sheet (Julian & Toomre, 1966). We first studied the density response to a perturbation using both the linear response method of chapter 2 and the JT equation (Julian & Toomre, 1966; Binney, 2020). The response was similar for both methods, despite the difficulties in creating comparable geometries for the two different approaches. We then showed that the level of amplification of the shearing sheet is comparable to the amplification in a full disc, despite the very different assumptions that the two approaches entail. We finally used our linear response calculation to see the response to a point perturbation and see the formation of a two arm spiral structure while using multiple angular harmonics.

The country ... demands bold, persistent experimentation. It is common sense to take a method and try it: if it fails, admit it frankly and try another. But above all, try something.

— F. D. Roosevelt, *Oglethorpe University Address*

4 Density Waves

Contents

4.1 Introduction 73
4.2 WKB Waves 74
4.2.1 LSK Dispersion Relation 74
4.2.2 Density Waves in WKB Limit 77
4.3 Kinematics of Waves in Cylindrical Coordinates 79
4.4 Test Application of a Logarithmic Spiral 86
4.5 Putting LSK to the Test 90
4.6 Conclusion 93

4.1 Introduction

The theory of tightly wound density waves is built upon the Lin-Shu-Kalnajs (LSK) dispersion relation (Lin & Shu, 1966; Kalnajs, 1965). To obtain this relation a WKB approximation is made, where the density waves are assumed to be tightly wound. Despite the limitations of the LSK relation, it is an important part in understanding the motion of density waves in disc and the formation of spirals (see Pasha, 2004a,b; Dobbs & Baba, 2014, for reviews), as it allows for a simple intuition about the motion of these waves. An understanding of these waves will be important in the following chapter, in particular when we look to study the

swing amplifier feedback cycle (Toomre, 1977, 1981; Bertin et al., 1989). In this chapter we will look to complete two tasks: we will review LSK theory and compare the predictions of LSK theory to the predictions made by linear theory, using the machinery of chapter 2. In order to study the motion of density waves in cylindrical geometries we will need to decompose them into in-going and out-going waves. To do this we will develop a method that projects density onto the eigenfunctions of the cylindrical wave operator.

In the following section we will review the WKB theory of density waves. We will then develop a method to decompose 2D waves in cylindrical geometry into in-going and out-going waves, which we will then apply to the unwinding of a logarithmic spiral. We will then finally look to reproduce the qualitative results of WKB theory.

4.2 WKB Waves

4.2.1 LSK Dispersion Relation

If we want to study the motion of a density wave then the name of the game is to find a self-consistent dispersion relationship. In chapter 3 we found this relationship by Laplace transforming the JT kernel in the axisymmetric limit (Equation 3.48). The Lin-Shu-Kalnajs (LSK) relation describes a wave, for which we assume a radial and temporal functional form $\sim \exp[\mathrm{i}(kR - \omega_0 t)]$,

$$\kappa^2(1 - s^2) = 2\pi G\Sigma(R)|k|\mathcal{F}(s, \chi)\mathrm{e}^{-\epsilon|k|}, \tag{4.1}$$

where $s = (\omega_0 - \ell\Omega(R))/\kappa(R)$ and $\chi = (k\sigma_R/\kappa)^2$ with $\Omega(R)$ and $\kappa(R)$ as the disc circular and epicycle frequencies (Lin & Shu, 1966; Kalnajs, 1965). In anticipation of chapter 5 we have also included a softening, ϵ (De Rijcke et al., 2019b). The exponential term acts to smooth out short range interactions. This is the same expression as Equation 3.54, but our definition of s is different to that used in chapter 4. The reason for this is simple: the analysis in chapter 3 is in the corotating frame and so we must include the $-\ell\Omega(\tilde{R})$ shift in our definition of s here. For a stellar disc $|s| \leq 1$ (Binney & Tremaine, 2008). The reduction factor, $\mathcal{F}(s, \chi)$,

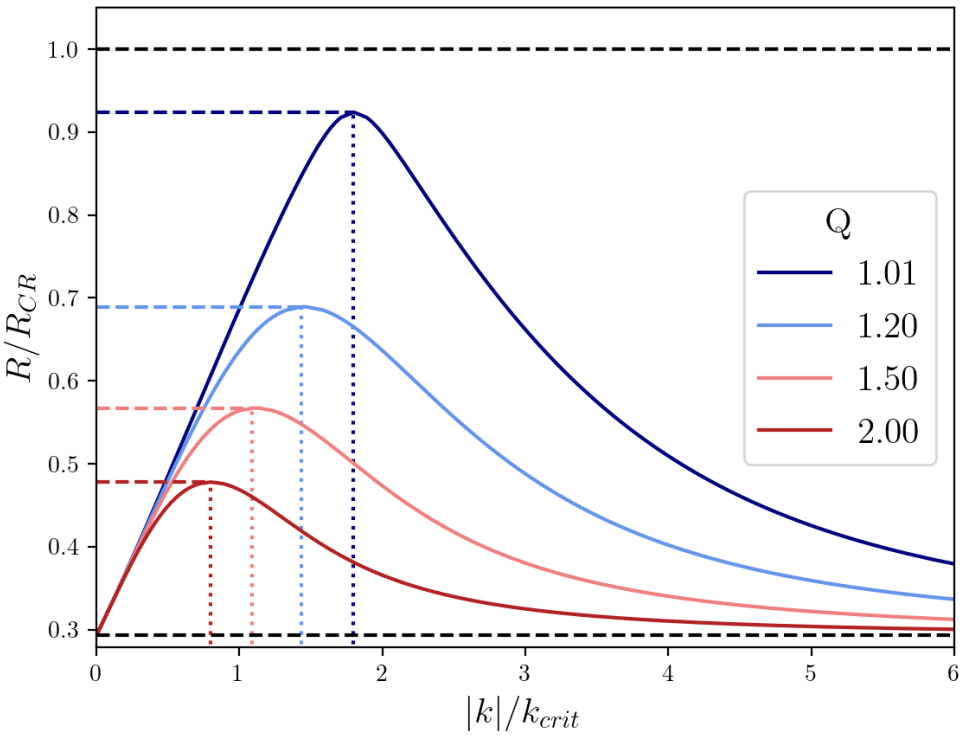


Figure 4.1: The LSK dispersion relationship for different temperature discs within CR, with the axis normalised by the critical wavenumber (c.f. Equation 3.40) and the CR radius. For a given mode ω_0 is fixed, therefore s depends solely on R via the epicycle and angular frequencies. The dashed black lines represent the ILR and CR. For discs with $Q > 1$ the LSK does not admit solutions in the 'forbidden region' around the CR, the inner boundary of which are represented by dashed coloured lines. For each R, the LSK relationship admits two solutions: the 'long' solution (small k) and the 'short' solution (large k). The point between the long and the short branches, k_*, are marked by vertical dotted lines. In normalised coordinates the plot is independent of ω_0.

which represents the disc's ability to overcome gravitational collapse due to the random thermal motion of the stars is given by

$$\mathcal{F}(s,\chi) = \frac{2(1-s^2)\mathrm{e}^{-\chi}}{\chi} \sum_{n\geq 1} \frac{I_n(\chi)}{1-(s/n)^2}, \tag{4.2}$$

where $I_n(\chi)$ are modified Bessel functions of the first kind. Despite the worrying $1/\chi$ factor in this expression, the reduction factor does not diverge and satisfies the property $\mathcal{F}(s,\chi) \leq \mathcal{F}(s,0) = 1$.

Care must also be taken with $\Sigma(R)$, the background density of the disc. By definition, $\Sigma(R) = \int \mathrm{d}^2\mathbf{v} F_0(\mathbf{x},\mathbf{v})$, which can be analytically evaluated for the

Mestel disc if the tapers are ignored. This yields the approximate form of the background density,

$$\Sigma \approx \frac{\xi}{2\pi G R_0}\frac{1}{R}, \tag{4.3}$$

which is valid in the region $R_\mathrm{i} \ll R \ll R_\mathrm{o}$, where R_i and R_o are the positions of the inner and outer taper respectively.

If the right-hand side of Equation 4.1 is less than (greater than) the epicycle frequency, then s is a purely real (imaginary) solution for s. The boundary between stable solutions, $s \in \mathbb{R}$, and unstable solutions, $s \in \mathbb{I}$, can be obtained by setting $s = 0$, yielding

$$\frac{|k|\sigma_R^2 \mathrm{e}^{\epsilon|k|}}{2\pi G\Sigma} = \left[1 - \mathrm{e}^{-\chi} I_0(\chi)\right]. \tag{4.4}$$

Unsurprisingly, there is no analytic solution to this condition. Solving it numerically, however, produces the result

$$Q = \frac{\sigma_R \kappa}{3.36 G\Sigma} > 1 \text{ for a stable disc,} \tag{4.5}$$

which is precisely the Toomre stability criteria (Toomre, 1964) with no softening. With softening this condition is still sufficient, but not necessary.

As can be seen from Figure 4.1, discs with $Q > 1$ do not have solutions to the LSK dispersion relation at all radii, but rather only outside of a 'forbidden region' which is centred on the CR. The reason for this name will become clear in the next section. From Figure 4.1 we can see that at each radius there are two solutions to the LSK relationship: the short wavelength solution (large k) and the long wavelength solution (small k). The value of k for which this we transition from the long to the short branch, k_*, is marked by a vertical dotted line. There is no analytic expression for k_* and the edges of the forbidden region, so they must be found numerically.

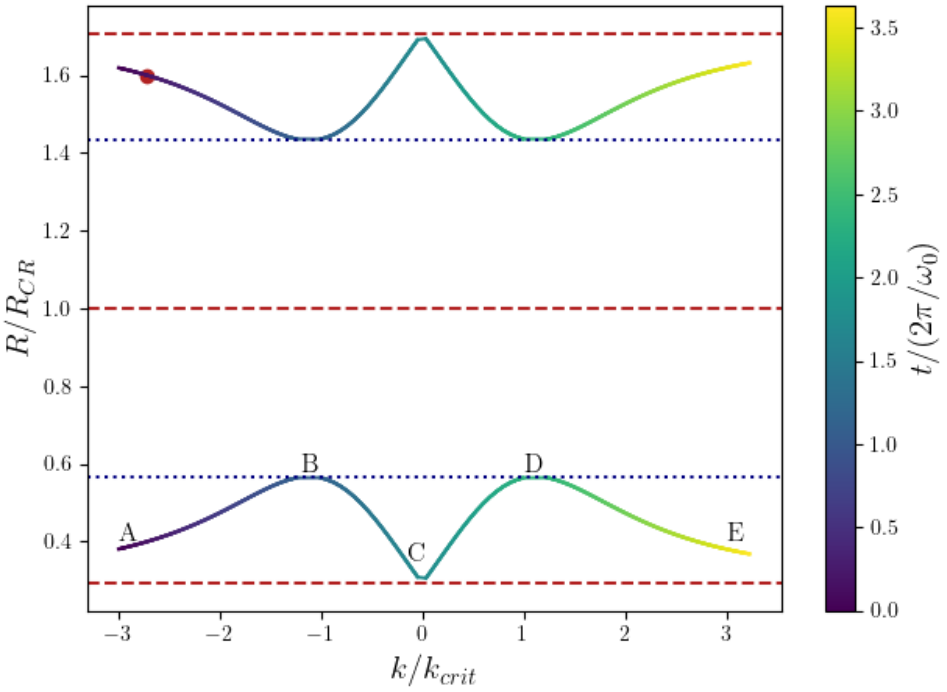


Figure 4.2: Motion of an initially leading WKB wave packet in a disc with $Q = 1.50$. The red dashed lines represent, from bottom to top
, the ILR, CR and OLR respectively, while the blue dotted lines represent the edges of the forbidden zone. As the wave unwinds it moves to the CR, where it is reflected off the forbidden region (B). The red point indiciates the intial conditions of the wave used in section 4.5. We do not include softening in this plot.

4.2.2 Density Waves in WKB Limit

In order to understand the propagation of a density waves one must obtain the group velocity from the LSK dispersion relation. The group velocity of a wave is

$$v_g(R) = \left(\frac{\partial \omega_0(k,R)}{\partial k}\right)_R, \tag{4.6}$$

where we have assumed that the distance over which ω_0 varies is much longer than the wavelength (e.g. Whitham, 1974). For the LSK dispersion relation we obtain the following,

$$\left[\frac{2s}{1-s^2} + \frac{\mathrm{e}^{\epsilon|k|}}{\mathcal{F}(s,\chi)}\frac{\partial \mathcal{F}}{\partial s}\right] v_g = -\frac{\kappa}{k}\left[1 - \epsilon|k| + 2\frac{\chi}{\mathcal{F}(s,\chi)}\frac{\partial \mathcal{F}}{\partial \chi}\right]. \tag{4.7}$$

where the derivatives of the reduction factor can be expressed in terms of

$$\begin{aligned}\frac{\partial \mathcal{F}}{\partial \chi} &= \frac{2(1-s^2)\mathrm{e}^{-\chi}}{\chi}\sum_{n\geq 1}\frac{I_{n-1}-2\left[1+1/\chi\right]I_n+I_{n+1}}{1-(s/n)^2},\\ \frac{\partial \mathcal{F}}{\partial s} &= \frac{-4s\mathrm{e}^{-\chi}}{\chi}\sum_{n\geq 2}\frac{1-(1/n)^2}{1-(s/n)^2},\end{aligned} \tag{4.8}$$

where for brevity's sake we use $I_n = I_n(\chi)$ (Toomre, 1969; De Rijcke et al., 2019b).

To illustrate the motion of a wave from both ILR or OLR, we plot the change in its wavenumber and radius as a function of time in Figure 4.2. We do this in a disc with $Q = 1.50$. In this plot and the rest of this chapter we ignore softening ($\epsilon = 0$) and only take it as non-zero in section 5.3. In order to calculate the forbidden radius, we numerically search for the largest value of $R < R_{\mathrm{CR}}$ that solves Equation 4.1. For a disc with $Q = 1.50$ the inner forbidden radius, which is calculated numerically, is $R/R_{\mathrm{CR}} \approx 0.55$. There are several key moments in the motion of such a wave which we illustrate by discussing the motion of a leading wave launched from ILR at point A in Figure 4.2:

A Initially at point A on Figure 4.2, the leading wave packet ($k < 0$) propagates towards the CR due to its positive group velocity, v_g. As it does so the wave starts to unwind, increasing its wavenumber, k.

B Once at the edge of the forbidden zone, at point B, the group velocity changes sign and the wave packets starts to propagate away from the forbidden zone, back towards the Lindblad resonance from which it came. The wave can be considered to 'reflect' off the forbidden region - the band in which the LSK relation only admits complex solutions (Mark, 1974). This process is akin to the reflection off a potential barrier experienced by a wave packet in quantum mechanics or by an electromagnetic evanescent wave.

C In the region around point C the progression of the wave in this region is in blatant violation of the tight winding approximation upon which the WKB approximation is based. In our derivation of the axisymmetric JT equation, we took $k_y \to 0$, while ensuring k_x, the wavenumber in the sheet that corresponds

to k in the disc, remained non-zero. However, this contradicts the LSK picture where k goes to zero. To study this process using a local theory, one must revert to the full dynamics of the shearing sheet, making use of the full JT equation. If a non-local linear model is wanted, as was required in chapter 3, then the techniques of chapter 2 can be utilized.

D Once again the wave, which is now trailing as $k > 0$ at point D, winds up as it travels towards CR. Its progress is once again stifled as it reflects off the forbidden zone and starts its final progression back towards the ILR.

E As the wave packet approaches the ILR at point E, it will do so as a trailing wave. Under the WKB approximation, this wave will be continually wound up to smaller wavelengths. In a real disc, however, the wave will ultimately be absorbed via a stellar equivalent of Landau damping (Toomre, 1969; Mark, 1974; Sellwood & Carlberg, 1984), where the structure in real space will 'damp' away by generating structure in velocity space (Sridhar, 2019).

A similar motion is true for a wave launched, with the direction of travel reversed (c.f. Figure 4.2). To study density waves further, a problem that will be important in chapter 5, we will want to be able to decompose waves into in-going and out-going waves. This surprisingly complex problem is one to which we now turn.

4.3 Kinematics of Waves in Cylindrical Coordinates

In order to study the evolution of density waves in a disc we want to decompose the density response of our disc, calculated using our linear response machinery, into in-going and out-going waves. If we were on a one dimensional line this would be a straight forward exercise: by projecting onto the eigenfunctions of the wave equation we could identify waves as left-moving ($kx + \omega t$) and right moving ($kx - \omega t$). Once our solution is projected into (k, ω) space we could split the solution into its two regimes and transform each part back separately, yielding the left and right moving

wave. This approach relies on the linearity of the wave equation, so that the total solution is the sum of the left and right moving waves. Our approach for the disc will be the same, but with the wave equation in cylindrical coordinates.

It is worth highlighting that the linearised CBE is an inhomogeneous wave equation

$$\frac{\partial f}{\partial t} + \boldsymbol{\Omega}(\mathbf{J}) \cdot \frac{\partial f}{\partial \boldsymbol{\theta}} = \frac{\partial F_0(\mathbf{J})}{\partial \mathbf{J}} \cdot \frac{\partial \psi}{\partial \boldsymbol{\theta}}, \tag{4.9}$$

with the left-hand side a first order wave operator in $\boldsymbol{\theta}$ and the right-hand side an inhomogeneity from the external perturbation and self consistent response. Expressing this as a second order wave equation for density is a complicated task (e.g. Lau & Binney, 2021).

Wave Equation in Cylindrical Coordinates

In order to separate the waves we first must obtain the eigenfunctions of the cylindrical wave equation. In cylindrical geometry with no z dependence the wave equation becomes the simplest form of the wave equation is

$$\frac{1}{R}\frac{\partial}{\partial R}\left[R\frac{\partial \rho}{\partial R}\right] + \frac{1}{R^2}\frac{\partial^2 \rho}{\partial \phi^2} = \frac{\partial^2 \rho}{\partial t^2}. \tag{4.10}$$

where we have chosen our units such that $c = 1$. We have chosen to use this form of the wave equation not because it precisely describes the linearised CBE, but rather because it serves as a useful guide for interpreting our solutions. By separation of variables we obtain the general solution,

$$\rho(R, \phi, t) = \sum_{\ell=-\infty}^{\infty} \mathrm{e}^{\mathrm{i}\ell\phi} \int_{-\infty}^{\infty}\int_{0}^{\infty} C_\ell(\omega)\mathrm{e}^{-\mathrm{i}\omega t}\left[D_\ell(k)H_\ell^{(1)}(kR) + E_\ell(k)H_\ell^{(2)}(kR)\right] dk d\omega, \tag{4.11}$$

where $C_\ell(\omega)$, $D_\ell(k)$ and $E_\ell(k)$ are the expansion coefficients of our solution and $H_\ell^{(1)}$ and $H_\ell^{(2)}$ are the Hankel functions of the first and second kind respectively. These Hankel functions can be considered the cylindrical analog of the complex exponential, which can clearly be seen when they are defined in terms of Bessel functions of the first kind, $J_\ell(x)$, and second kind, $Y_\ell(x)$,

$$\begin{aligned} H_\ell^1(x) &= J_\ell(x) + \mathrm{i}Y_\ell(x), \\ H_\ell^2(x) &= J_\ell(x) - \mathrm{i}Y_\ell(x). \end{aligned} \tag{4.12}$$

From Equation 4.11 we can see that the eigenfunctions of the cylindrical wave equation in 2D are $\mathrm{e}^{\mathrm{i}\omega t}\mathrm{e}^{\mathrm{i}\ell\phi}H_\ell^n(kR)$ where $n \in [1,2]$. If we were to follow the Cartesian case we would now calculate the necessary integrals to project our density wave onto the basis functions in Equation 4.11; for the t and ϕ dependence this would be straightforward: for the time dependence we would FFT and for the angular dependence we would use

$$\int_0^{2\pi} e^{-\mathrm{i}\phi\ell} e^{\mathrm{i}\phi\ell'} d\phi = 2\pi\delta_{\ell'}^{\ell}. \tag{4.13}$$

This approach cannot be taken for the radial dependence, as the second order Bessel functions do not obey an orthogonality relationship due to their divergent nature at the origin. To see this, consider two Bessel functions, $B_\ell(ax')$ and $B_\ell(bx')$, where B_ℓ are either first order or second order Bessel functions and a and b are zeros of the Bessel function. By definition they satisfy the Sturm–Liouville operator

$$\frac{x'}{a}\frac{d}{dx'}\left[x'\frac{dB_\ell(ax')}{dx'}\right] + ax'B_\ell(x') = \frac{\ell^2}{ax'}B_\ell(ax'), \tag{4.14}$$

and similarly for $B_\ell(bx')$. If we multiply this ODE for $B_\ell(ax)$ by b^2 and subtract it from a similar expression with a and b swapped we obtain

$$\begin{aligned}&\left[x'B_\ell(bx')\frac{dB_\ell(ax')}{dx'} - x'B_\ell(ax')\frac{dB_\ell(bx')}{dx'}\right]_{x'=0}^{x'=1} \\ &+ (a^2-b^2)\int_0^1 B_\ell(ax')B_\ell(bx')x'dx' = 0,\end{aligned} \tag{4.15}$$

after we have integrated with respect to x'. Now consider the case where B_ℓ is either the first order or second order Bessel function:

- Bessel Functions of the first kind, $B_\ell = J_\ell$. As these Bessel functions and their derivatives are well defined at the origin, the lower limit of the first bracket is zero. The upper limit is also zero, as a and b are the roots of the Bessel functions. As a result we obtain the standard first order Bessel orthogonality relationship, namely for $a \neq b$

$$\int_0^1 J_\ell(ax')J_\ell(bx')x'dx' = 0 \tag{4.16}$$

- Bessel Functions of the second kind, $B_\ell = Y_\ell$. The argument is the same as above, however now as the Bessel functions are undefined at the origin. In particular in the limit that $x \to 0$, $Y_\ell(ax) \sim 1/(ax)^\ell$ for $\ell \in \mathbb{Z}^+$. Apart from the case in which $a = b$, which trivially solves Equation 4.15, the divergence at $x = 0$ from the square bracket will never cancel. Hence for the Bessel functions of the second kind, an orthogonality relationship cannot be found.

As we cannot project onto Bessel functions of the second kind, we will not be able to project onto the Hankel functions. To proceed with the decomposition onto the radial eigenfunctions, we will take a different approach.

WKB Basis

As we cannot project the radial part of the response onto the Hankel functions, we must instead look for a set of basis functions that solve the radial part of the cylindrical wave equation that we can easily project onto. If we consider the Hankel functions in the limit $x = kR \gg 1$ then to first order we can expand them as

$$\begin{aligned} H_\ell^{(1)}(x) &\approx \sqrt{\frac{2}{\pi x}} \mathrm{e}^{ix}. \\ H_\ell^{(2)}(x) &\approx \sqrt{\frac{2}{\pi x}} \mathrm{e}^{-ix}, \end{aligned} \tag{4.17}$$

ignoring any global phases that are independent of x. We will now construct a basis – which, crucially, we will be able to project onto using a orthogonality relation – using these approximate forms of the Hankel functions. We can once again take inspiration from the Cartesian case and introduce the basis functions

$$\mathrm{e}_k(R) = \frac{1}{\sqrt{2\pi|kR|}} \mathrm{e}^{ikR}. \tag{4.18}$$

If we want to project a density response that is zero at the origin and at some outer value R_o, we choose k such that $R_\mathrm{o} k/\pi = n \in \mathbb{Z}\backslash 0$. Note that the Kalnajs basis has this property with $R_\mathrm{o} = R_\mathrm{Ka}$ (see section 2.3 for details). We define the basis of the dual space as

$$\mathrm{e}_k^*(R) = \frac{1}{\sqrt{2\pi|kR|}} \mathrm{e}^{-ikR}. \tag{4.19}$$

	$k>0,\ \ell>0$	$k<0,\ \ell>0$	$k>0,\ \ell<0$	$k<0,\ \ell<0$
$\omega<0$	In-going Clockwise	Out-going Clockwise	In-going Anticlockwise	Out-going Anticlockwise
$\omega>0$	Out-going Anticlockwise	In-going Anticlockwise	Out-going Clockwise	In-going Clockwise

Table 4.1: A table showing the motion of waves with different values associated with the different motion: radially in-going clockwise, out-going clockwise, out-going anti-clockwise and in-going anti-clockwise.

This allows an inner product to be defined, doing so via its action on two basis functions

$$\langle \mathrm{e}_{k'}(R), \mathrm{e}_k(R)\rangle = \int_0^{R_\mathrm{o}} \frac{e^{-\mathrm{i}k'R}\mathrm{e}^{\mathrm{i}kR}}{2\pi\sqrt{|kR||Rk'|}} RdR = \frac{R_\mathrm{o}}{k}\delta_k^{k'}, \tag{4.20}$$

where we have introduced R as the natural weight factor for cylindrical geometry. Note that this weight factor is the same as for Bessel Functions, an unsurprising consequence of where we started.

This orthogonality relationship implies that $\{\mathrm{e}_k(R)\}$ are linearly independent. For $\{\mathrm{e}_k(R)\}$ to constitute a basis, we must show that they are complete. Completeness follows because we can approximate any well-behaved piecewise continuous function, $f(R)$, to arbitrary accuracy measured by the mean squared error. We do this by choosing expansion coefficients

$$a_k \frac{k}{R_\mathrm{o}} = \langle e_k(R), f(R)\rangle = \int_0^{R_\mathrm{o}} e_k^*(R) f(R) RdR. \tag{4.21}$$

In essence what we have done here is to introduce a cylindrical equivalent of the Fourier transform. We also note that these basis functions are only a solution of the wave equation in the limit $kR \gg 1$. Therefore we will find it hard to study long wavelengths close to the origin. However, as we only care about the motion of density waves outside of ILR we will not run into this complication as long as ω_0 is small enough.

Separating the Waves

Armed with our basis functions we express our density response as

$$\begin{aligned}\rho(R,\phi,t) &= \sum_{k=-\infty}^{\infty}\int_{-\infty}^{\infty} C_\ell(\omega,k)\mathrm{e}_k(R)\exp\left[i(-\omega t+\ell\phi)\right]d\omega + \text{c.c.},\\ &= \sum_{k=-\infty}^{\infty}\int_{-\infty}^{\infty} C_\ell(\omega,k)\frac{\exp\left[i(-\omega t+kR+\ell\phi)\right]}{\sqrt{2\pi|kR|}}d\omega + \text{c.c.},\end{aligned} \tag{4.22}$$

where we have assumed that our density response is composed entirely of one angular harmonic, ℓ, and that any unwanted normalisation factors have been absorbed into C_ℓ. From Equation 4.22 we can see that our density solution is a weighted sum of plane waves that have the form

$$\mathrm{e}_k(R)\exp\left[i(-\omega t+\ell\phi)\right] = \frac{\exp\left[i(-\omega t+kR+\ell\phi)\right]}{\sqrt{2\pi|kR|}}. \tag{4.23}$$

From this expression we can now see the four different motions of a wave: in-going clockwise, out-going clockwise, out-going anti-clockwise and in-going anti-clockwise. The values of parameter space which represent one of these four options are laid out in Table 4.1.

Now that we have figured out the eigenfunctions of the wave equation, we must project our density response calculated via our linear response machinery onto them, i.e. we must calculate $C_\ell(\omega,k)$. To do this write our density response as

$$\rho(R,\phi,t) = \rho(R,t)\cos\left[\ell\phi+\psi(R,t)\right], \tag{4.24}$$

where $\rho(R,t)$ is the density amplitude and ψ the phase of the response. The phase is related to the radial wavenumber, $k(R,t)$, via

$$k(R,t) = \left.\frac{\partial\psi(R,t)}{\partial R}\right|_t. \tag{4.25}$$

If we define

$$\begin{aligned}I_\mathrm{c}(R,t) &= \frac{1}{2\pi}\int_0^{2\pi}\rho(R,\phi,t)\cos(\ell\phi)d\phi,\\ I_\mathrm{s}(R,t) &= \frac{1}{2\pi}\int_0^{2\pi}\rho(R,\phi,t)\sin(\ell\phi)d\phi,\end{aligned} \tag{4.26}$$

then we obtain $\rho(R,t)^2 = I_\mathrm{c}(R,t)^2 + I_\mathrm{s}(R,t)^2$ and $\tan\psi(R,t) = I_\mathrm{s}(R,t)/I_\mathrm{c}(R,t)$. From this $C_\ell(t,k)$ is

$$C_\ell(t,k) = \int_0^{R_\mathrm{o}} \left[\frac{2\pi \mathrm{e}^{\psi(R)}}{2} \frac{\rho(R,t)\mathrm{e}^{-ikR}}{\sqrt{2\pi|kR|}}\right] R dR, \tag{4.27}$$

where the first factor is from the phase integral and the second factor from the inner product with the dual basis. For this integral to converge, we assume that $R\rho(R,t) \to 0$ faster than the denominator. For densities that have been projected into real space from their Kalnajs representation this must be true, as for small R the radial basis functions $\sim R^\ell$. As $C_\ell(\omega,k)$ is the Fourier transform of $C_\ell(t,k)$, we can numerically calculate it by using a standard fast Fourier transform method. We calculate $C_\ell(\omega,k)$ on a rectangular grid.

We split $C_\ell(\omega,k)$ into four parts, $C_\ell^\mathrm{IC}(\omega,k)$, $C_\ell^\mathrm{OC}(\omega,k)$, $C_\ell^\mathrm{OA}(\omega,k)$ and $C_\ell^\mathrm{IA}(\omega,k)$, which are equal to $C_\ell(\omega,k)$ in the region laid out by Table 4.1 and zero elsewhere. For example, for in-going clockwise waves

$$C_\ell^\mathrm{IC}(\omega,k) = \begin{cases} C_\ell(\omega,k), & \text{if } k,\ell > 0 \text{ and } \omega < 0 \quad (4.28a) \\ 0, & \text{otherwise.} \quad (4.28b) \end{cases}$$

To obtain the real space representation of these waves we first invert the FFT of each part separately and then project back into real space from the WKB basis, i.e.

$$\rho^\mathrm{IC}(R,\phi,t) = \sum_{k=-\infty}^{\infty} C_\ell^\mathrm{IC}(t,k) \frac{\mathrm{e}^{\mathrm{i}(\ell\phi+kR)}}{\sqrt{|kR|}} + \text{c.c.}, \tag{4.29}$$

and similarly for C^OC, etc. In LSK theory the direction of these waves will also depend on which branch of the dispersion relation the wave is on. When projecting back into real space we can restrict the the summation domain in order to obtain short waves ($|k| > k_* k_\mathrm{crit}(R)$) or long waves ($|k| < k_* k_\mathrm{crit}(R)$) where k_* is the value of k at the turning point of the dispersion relation in units of k_crit (c.f. Figure 4.1). Explicitly, for example, we can write the in-going clockwise long and short waves as

$$\begin{aligned} \rho^\mathrm{IC}_\mathrm{long}(R,\phi,t) &= \sum_{k=-k_*(R)}^{k_*(R)} C_\ell^\mathrm{IC}(t,k) \frac{\mathrm{e}^{\mathrm{i}(\ell\phi+kR)}}{\sqrt{|kR|}} + \text{c.c.}, \\ \rho^\mathrm{IC}_\mathrm{short}(R,\phi,t) &= \sum_{\substack{k=-\infty \\ k\notin[-k_*(R),k_*(R)]}}^{\infty} C_\ell^\mathrm{IC}(t,k) \frac{\mathrm{e}^{\mathrm{i}(\ell\phi+kR)}}{\sqrt{|kR|}} + \text{c.c.}, \end{aligned} \tag{4.30}$$

respectively, where we abbreviate $k_*(R) = k_* k_{\rm crit}(R)$, and similarly for OC, etc.

In this section we have developed a kinematic method to analyse the dynamical response of a disc. In the next section we will use it to study the evolution of a toy problem – the unwinding of a logarithmic spiral – before testing the LSK dispersion relation in section 4.5.

4.4 Test Application of a Logarithmic Spiral

We now test two of the key assumptions that were made in the previous section: that the density response is described by the physics captured in the cylindrical wave equation and that our decomposition is calculated correctly. We will show this by comparing the original density response, calculated using the machinery of chapter 2, to the sum of the decomposed density responses calculated as in the previous section. If our assumptions are correct, then once summed over the split response will be equal to the original density response. It is worth highlighting that these tests are not particularly profound, rather they allow us to us to test the assumptions of this method and serve as an instructive example to show how the decomposition works in practice.

To do this we will study the evolution of an initial perturbation in the form of a logarithmic spiral,

$$\rho_{\rm s}(R, \phi) = \rho_{\rm s}(R) \cos(\ell\phi + kR). \tag{4.31}$$

For the radial density of the spiral we take

$$\rho_{\rm s}(R) = \frac{2M(R/R_{\rm s})^{\mu/2}}{1 + (R/R_{\rm s})^{\mu}}, \tag{4.32}$$

with M setting the maximum density of the spiral, μ setting the sharpness of the inner and outer taper and $R_{\rm s}$ setting the radial scale. In the calculations that follow we take $\mu = 6$ and $R_{\rm s} = 5$. Note that when assuming an initial perturbation, the solution of the linearised CBE is the sum of the particular integral (due to any external perturbation and self-gravity of the initial disturbance) and the homogeneous solution, $f_{\mathbf{m}}(\mathbf{J}, t = 0) \exp(-{\rm i}\mathbf{m} \cdot \mathbf{\Omega})$. The homogeneous solution

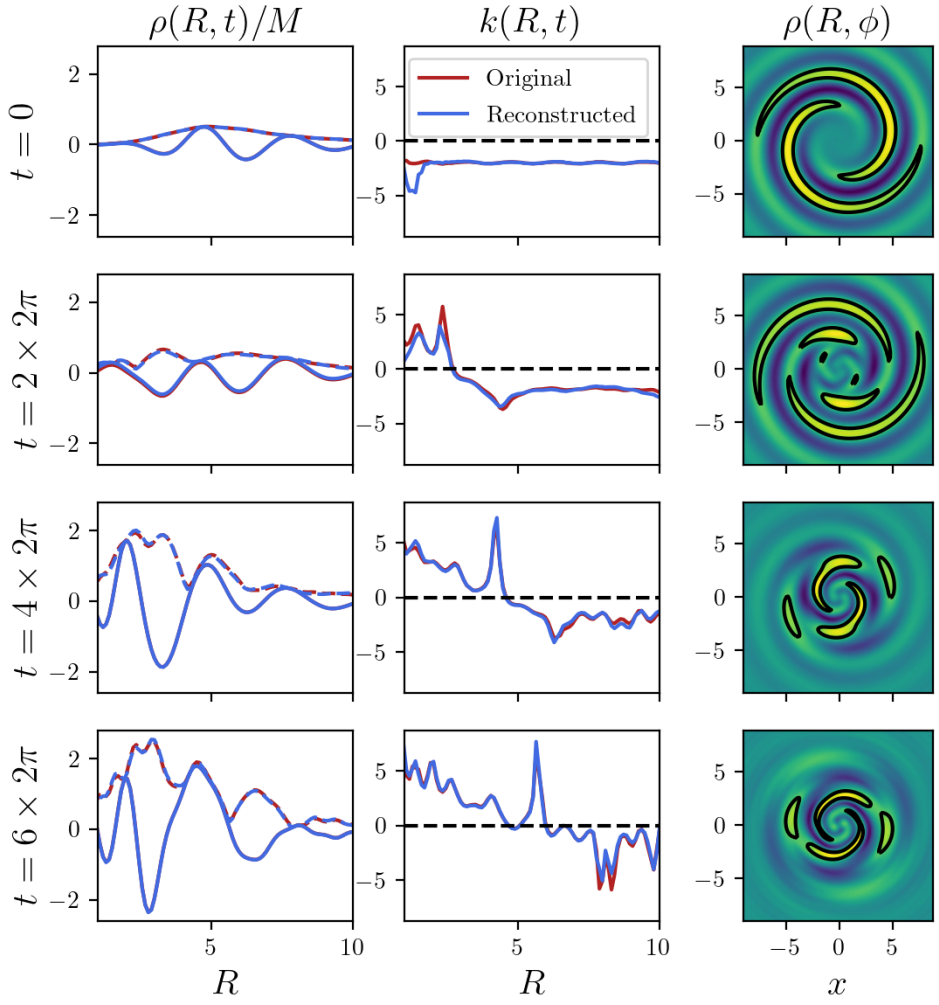


Figure 4.3: A comparison between the original and reconstruction of an unwinding logarithmic spiral. *Left hand column*: the amplitude of the density response (dashed) using the original response (red) and its reconstruction after the decomposition process (blue). *Middle column*: the wavenumber, $k(R)$, as a function of radius for the original and reconstruction. *Right hand column*: the density response of the spiral with a black contour at 50% the maximum value of the density snapshot. As expected the original density evolution and the reconstruction, which is calculated by summing the in-going and out-going density wave response, agree closely.

corresponds to the kinematic evolution of the spiral, while the particular integral corresponds to the evolution driven by self-gravity. As we will ultimately only be interested in the response of a disc driven by self-gravity (and possibly some external perturbation) when solving Equation 2.16 only the particular integral is considered. As this is a toy example, designed to test and illustrate the mathematics laid out in the previous section, this complication is not required.

In order to calculate the response we use the standard tapered Mestel disc with $Q = 1.5$ and $\xi = 0.5$. We calculate $\hat{\psi}_{\mathbf{m}}(\mathbf{J})$ on a uniform grid in apocentre and pericentre coordinates with 251 points between $R = 0$ and $R = 15$ and for indices $-4 \leq m_1 \leq 4$ (c.f. chapter 2), as a balance between computational speed and numerical convergence. For the kernel we use 500 uniformly spaced time steps with a distance of 0.3, to ensure that all perturbations have died away by the end of the calculation. The first forty Kalnajs basis functions are used. As the Kalnajs basis is orthonormal, we can easily calculate $B_p(t = 0)$ using Equation 4.31 and, as we are not considering the action of an external perturbation, we do not require A_p.

In Figure 4.3 we plot the evolution of the density response in two ways. The red lines in the first two columns represent the direct solution of the Volterra equation. The blue lines in the first two columns represents the sum of these four different possibilities (in-going, out-going, long and short), which recover the response initially fed in (i.e. the red line). The slight discrepancy is due to the number of basis functions which we project onto when separating the in-going and out-going: we use 500 time steps with our FFT and 300 for our radial projection. The number of basis functions used could be increased leading to a slight improvement in agreement, but a much longer computational time.

In the left hand column of Figure 4.3 we plot the envelope of the density response (dashed line) and the density along $\phi = 0$ (solid line). The ability of the reconstruction to reproduce the original is very strong away from the centre. However closer to the origin the $1/\sqrt{kR}$ in the WKB, will always diverge and so the density at the centre is poorly reproduced.

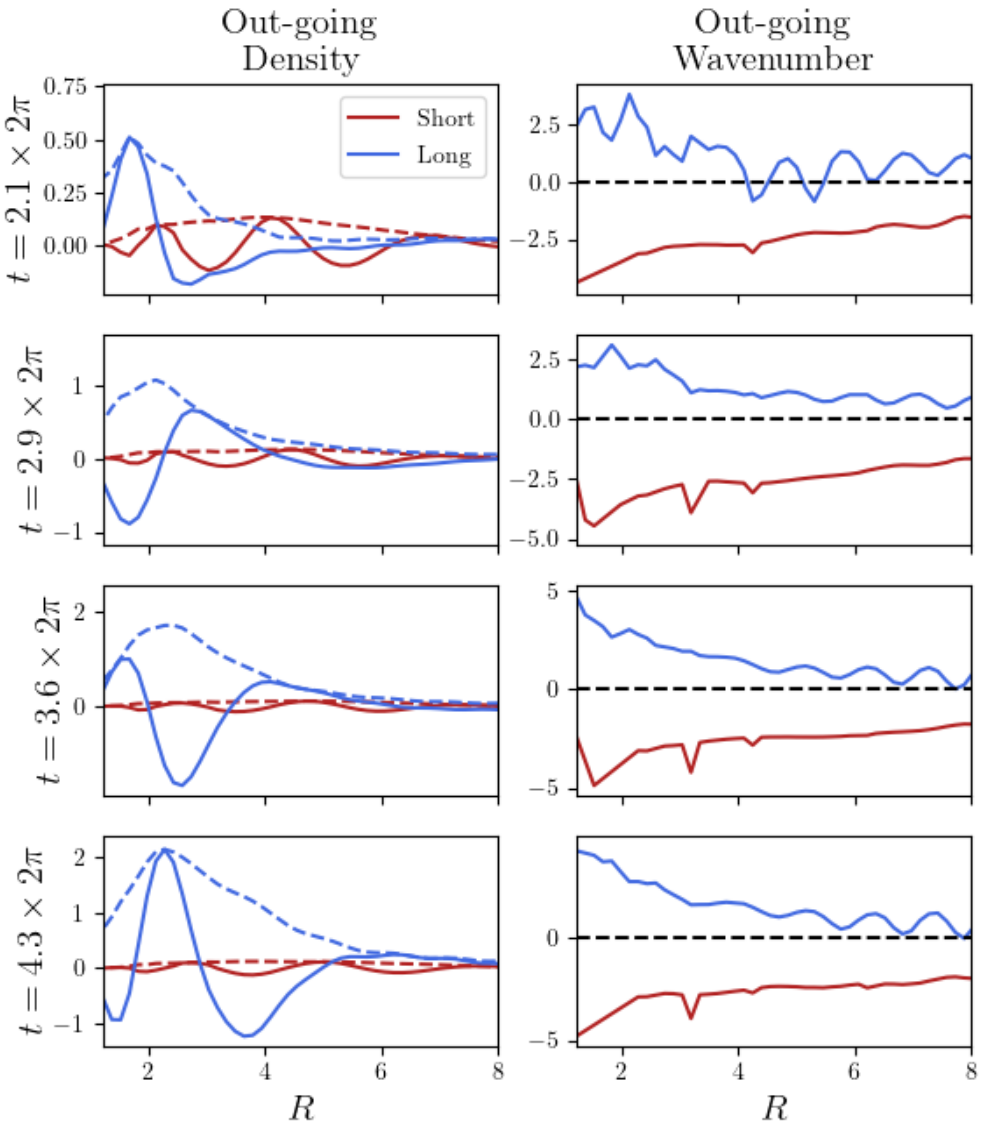


Figure 4.4: The out-going component of the density response spilt into short wavelength (red) and long wavelength (blue). *Left panel*: the density amplitude (dashed line) and the density along the x-axis (solid line) which corresponds to the density wave with a given phase, which highlights the wave like nature. *Right panel*: the wavenumber for the long and short waves.

In the middle column of Figure 4.3 we plot the evolution of $k(R,t) = \partial\psi(R,t)/\partial R$. Again there is good agreement between the original and the reconstructed, however it is not perfect, as result of projecting onto finite incomplete bases (i.e. the radial and temporal basis). As we chose our initial perturbation to have the form of a logarithmic spiral, $k(R, t = 0)$ is independent of radius. However as time evolves, the spiral will unwind from its initially leading configuration ($k(R,t) < 0$) and eventually start to wrap itself up as a trailing spiral ($k(R,t) > 0$). In the right hand column we show the density response of the spiral to indicate its transition from leading to trailing. To aid with this, we plot a black contour at 50% of the maximum value of the density snapshot. It is curious to note that we see spikes in the wavenumber at different radii. When the density is low, the wavenumber, which is a derivative of the phase, can change very rapidly, leading to these spikes.

In Figure 4.4 we show the out-going component of the density response split into short wavelength (red) and long wavelength (blue), via the condition laid out in Equation 4.30. In the right panel, we can clearly see the short out-going waves are leading, while the long outgoing waves are trailing. This is superficially reminiscent of the intuition gained in subsection 4.2.2. We also clearly see that red solid line has a sorter wavelength than the blue line (the dashed line just shows the envelope of the wave packet). However in order to make a strict comparison to LSK one must know the position of the wave relative to CR. Therefore the question becomes: what is ω_0? For the logarithmic spiral the response is made up of a superposition of modes, each with a different ω_0, making a comparison to LSK difficult. To be able to make a comparison to LSK we must generate a wave with a fixed frequency, ω_0.

4.5 Putting LSK to the Test

In this section are going to perturb our disc and generate a self-consistent density response using the machinery developed in chapter 2. We will then use the techniques developed in the previous two sections to decompose the response into in-going and out-going waves, which we will then compare the LSK theory. There are two qualitative predictions that are an important part of LSK theory: (i) waves on the

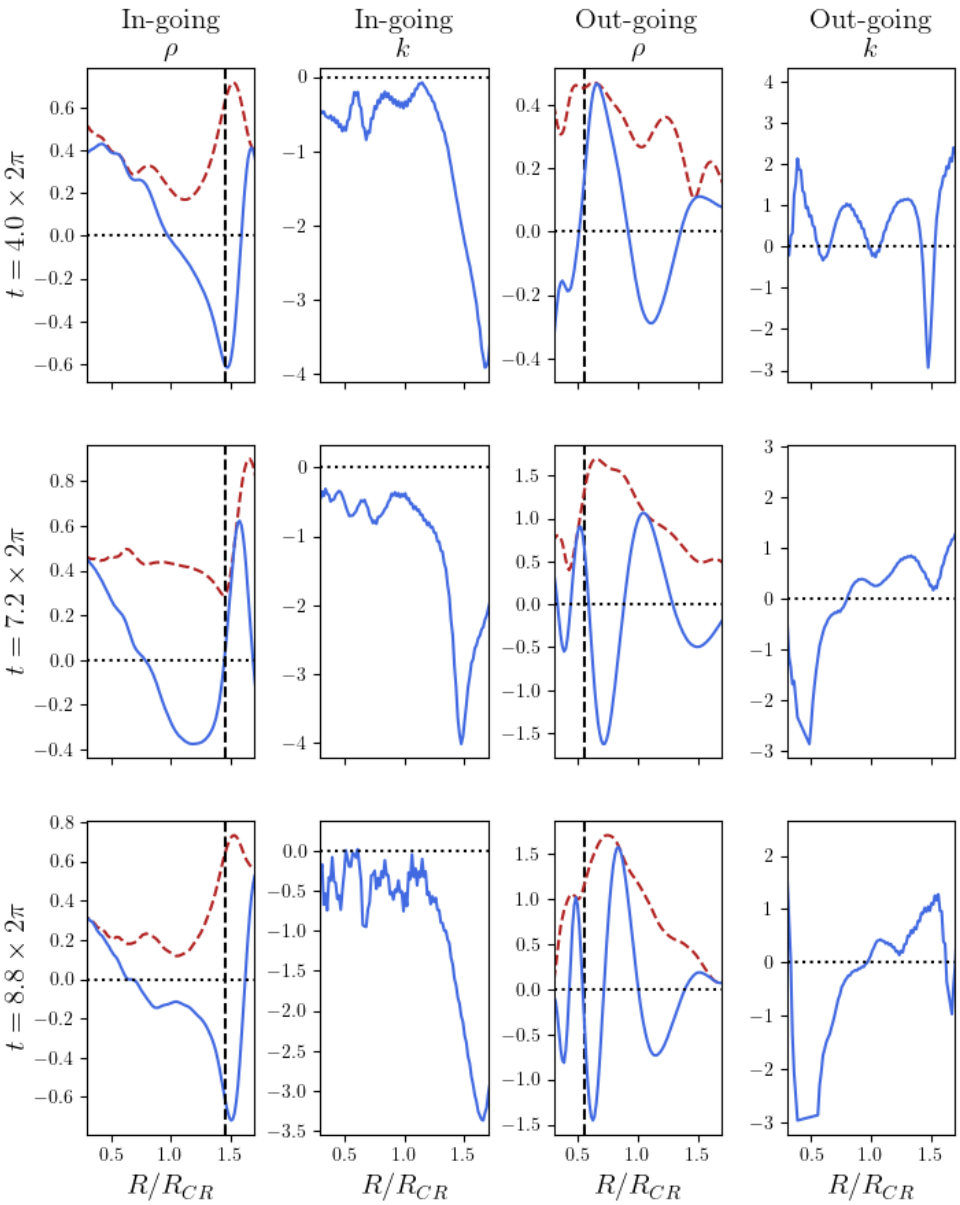


Figure 4.5: The density response to the perturbation in Equation 4.33 split into in-going and out-going waves. Counting from the left, we plot: *First column*, the density amplitude, $\rho(R)$ (red) and the density along the x-axis, $\rho(R, \phi = 0)$ (blue) for in-going waves; *Second column*, the wavenumber, $k(R)$, for the in-going waves; *Third column*, same as the first column, but for out-going waves; *Fourth column*, same as the second column, but for out-going waves. The vertical dotted line in the first (third) columns shows the position of the outer (inner) edge of the forbidden radius. The results shown are in qualitative agreement with LSK theory.

short branch of the LSK dispersion relationship will move towards (away from) CR if leading (trailing); (ii) waves decay as they pass through the forbidden region (Mark, 1974). As highlighted in the previous section a direct comparison is rather difficult (for a given perturbation what is ω_0?).

Instead we will impose an initial perturbation of the form given by the LSK relation and will decompose the resulting evolution into in-going and out-going waves, i.e. the wave's wavelength will be calculated using LSK and we will assume a form of the envelope. For this initial perturbation we will assume that $\omega_0 = 0.4$ (so that $R_{\rm CR} = 5$) and that our wave will start at $R = 8$ (this position is marked by the red dot on Figure 4.2). We can change these values without qualitative change as long as the response is away from the inner taper, at $R = 1$, and outer taper, at $R = 11.4$. We start from closer to OLR due to its larger radius so that our approximation of the Hankel functions, $|kR| \gg 1$ (c.f. section 4.3), is better satisfied. From the LSK relation then the initial wavenumber, i.e. the x-axis value of the red dot in Figure 4.2, is $k = -1.36$, where we assume an initially leading wave. To these assumptions we will add that our wave packet has an envelope of the form

$$\rho(R) = \frac{M}{\sqrt{2\pi}\sigma_{\rm w}} {\rm e}^{-\frac{(R-R_{\rm s})^2}{2\sigma_{\rm w}}}, \tag{4.33}$$

where we will take $\sigma_{\rm w} = 0.3$ and $M = 1$ as the size of the initial condition. For this wave the group velocity calculated via Equation 4.7 is 3.8×10^{-2}. For the calculation of the kernel we will use the same parameters as the previous section, but will increase our number of basis functions to 72, in order to allow the narrow initial perturbation to be resolved. Importantly this means that the forbidden radius for a disc with this temperature, $Q = 1.5$, is at $R \approx R_{\rm CR}(1 \pm 0.45)$.

In Figure 4.5 we plot the density response to this perturbation, calculated using Equation 2.16, separated into in-going and out-going waves using the techniques outlined in section 4.3. Counting from the left, we plot: *First column*, the density amplitude, $\rho(R)$ (red) and the density along the x-axis, $\rho(R, \phi = 0)$ (blue) for in-going waves; *Second column*, the wavenumber, $k(R)$, for the in-going waves; *Third column*, same as the first column, but for out-going waves; *Fourth column*, same as

the second column, but for out-going waves. The vertical dotted line in the first (third) columns shows the position of the outer (inner) edge of the forbidden radius.

It can be quickly seen that the real world is not as clean as LSK theory might initially suggest. Despite that, this figure still illustrates the predictions of LSK theory. From the in-going density plots we observe that the waves are short at large radius, but long at small radius. The sign of k is that predicted via LSK theory using these observations (and the fact that the waves are in-going). The picture is simple: as the wave moves inwards from large R it unwinds, but by the time it has crossed the forbidden region it has not had sufficient time to unwind, so it continues its travels inwards while unwinding, but now on the long branch of the LSK dispersion relationship. From the third column we see that all the out-going density waves are short. Again this observation and the sign of k agrees with LSK theory, for short waves to travel outwards they must have $k < 0$ within CR and $k > 0$ outside of CR as seen in the fourth column.

Where LSK fails is in the notion of a forbidden zone. In the first column we mark the outer edge of the forbidden zone, while in the third column we mark the inner edge of the forbidden zone. In the third column, although as a function of radius the amplitude decreases, as a function of time the density is increasing. This is unsurprising, in-fact we spent chapter 3 studying what is going on here, namely swing amplification. As LSK assumes tightly wound spirals, it will never be able to explain the amplification that a wave undergoes during its least wound phase. This process happens in the forbidden zone and so the picture from LSK, of short waves being reflected off the forbidden zone is simply not true.

4.6 Conclusion

In this chapter we turned our attention to LSK theory and the study of density waves. We started by reviewing the LSK dispersion relation (Lin & Shu, 1966; Kalnajs, 1965) and the motion of density waves in the WKB limit (Toomre, 1969). In particular we highlighted the relation between the direction of winding and the direction of travel. We then developed a method to decompose waves in cylindrical

geometry into in-going and out-going waves. The main complication was finding a basis for the radial decomposition as the Hankel functions, the cylindrical equivalent of complex exponential in the Cartesian case, cannot easily be projected onto. To solve this problem we constructed a basis based on the Hankel functions in the limit $|kR| \gg 1$. By studying the unwinding of a logarithmic spiral we showed that this approach could accurately reproduce the input evolution. We finally used this method to study the motion of WKB waves. We confirmed the relationship between the direction of travel and the sign of the wavenumber on the short branch of the LSK relation, however due to the number of assumption behind LSK theory a clear comparison was not possible.

Except for the fact there was no solution, the problem was simple.

— M. Kishlansky, *A Monarchy Transformed*

5 Scars

Contents

5.1 Introduction 95
5.2 Modes of the Tapered Mestel Disc 97
5.2.1 Volterra Method for Response Matrix 98
5.2.2 Coefficient Fitting 100
5.2.3 Modes of the Tapered Mestel Disc 102
5.2.4 Active Fraction & Modes of the Mestel Disc 103
5.3 Modes of Scarred Discs - Angular Momentum Scars . 103
5.3.1 WKB Cavity Modes 104
5.3.2 Scar Model 106
5.3.3 Cavity Modes of Scarred Discs 109
5.4 Conclusions 112

5.1 Introduction

Spirals are common in disc galaxies, yet as was outlined in chapter 1 understanding the processes that drive their formation is complicated. Mechanisms by which spiral structures can be formed by external perturbations, such as bars (Schwarz, 1981; Athanassoula, 2012) or external satellites (e.g. Holmberg, 1941; Toomre & Toomre, 1972; Dobbs et al., 2010) which act as giant sticks stirring the disc, are straightforward but can spirals be formed independent of external meddling? From studying the stability of disc models, many are susceptible to non-asymmetric

instabilities (e.g. Weinberg, 1991; Vauterin & Dejonghe, 1996; Pichon & Cannon, 1997; Evans & Read, 1998; Jalali & Hunter, 2005; Khoperskov et al., 2007; Fridman & Polyachenko, 2012; Polyachenko & Just, 2015; De Rijcke et al., 2019a), many of which lead to spiral instability. Therefore if we want to form spirals independent of external influence the question then becomes: how does an initially stable disc become unstable?

One such mechanism is the swing amplifier feedback cycle (Toomre, 1977, 1981; Bertin et al., 1989), established when two 'barriers' create a resonance cavity. In this cycle an in-going trailing wave is reflected into an out-going leading wave at an inner 'barrier' before being reflected back into a in-going trailing wave at an outer 'barrier', creating a 'resonant' cavity. As these waves bounce back and forth they can amplify via their own self-gravity. In particular the out-going leading wave will unwind, amplifying via swing amplification as it does so (chapter 3, Julian & Toomre, 1966; Binney, 2020). The motion of density waves is well described by LSK theory, but how does one construct a resonant cavity – and in particular the inner and outer barrier – in a disc?

From LSK theory the only possible outer wall of the cavity is simply the inner forbidden radius around CR as it will reflect any wave that is incident upon it (c.f. Figure 4.2). The inner cavity wall cannot be the ILR, because rather than reflect waves, it absorbs them[1] a change in impedance that could act as an inner barrier (Sridhar, 2019). One way to create a reflective barrier is to change the impedance of the field through which the density wave travels. A process that does this, without the need for external intervention, is secular scarring (Sellwood & Carlberg, 2014; Fouvry et al., 2015b). In this process the disc is heated and orbits radially diffuse leaving a 'scar' behind. De Rijcke et al. (2019a) studied the modes of a scarred disc and confirm them as cavity modes by using an elegant WKB argument. This builds on the N-body work of Sellwood & Kahn (1991), who show that sharp features in background DFs lead to instabilities. We want to expand this work by using more realistic density models to confirm De Rijcke

[1]Note that Landau damping of these waves at ILR will lead to heating (Mark, 1974) and thus trigger

et al. (2019a) findings. To do this we will need to calculate the modes of a disc, a task we will turn to in the following section, before applying that machinery to study scarred disc and the existence of cavity modes.

5.2 Modes of the Tapered Mestel Disc

In this chapter we will use two different methods to calculate the response matrix, $\hat{\mathcal{M}}(\omega)$:

1. Volterra Method (subsection 5.2.1): following Magorrian (2021) we will assume that the response coefficient takes the form $B_p = \hat{B}_p e^{-\mathrm{i}\omega t}$ and use this time dependence to numerically calculate the response matrix using the kernel and Volterra equation, Equation 2.16. This method can run into numerical difficulties for slowly decaying modes. We will discuss this subsection 5.2.4.

2. Laplace Method (Appendix C): we will follow the approach laid out in Fouvry et al. (2015b), where the response matrix is calculated by integration over action space. This method is numerically correct for all positive decay rates, no matter how small.

In essence the difference between these two methods is the order of the time and action space integration: in the Volterra Method we integrate over actions space first (to calculate the kernel) and then calculate the time integral, whereas in the Laplace method we integrate over time first and then calculate the action space integrals. As it is the time integral that introduces ω, we find that the Volterra approach is much quicker when searching for modes as the action-space integral does not need to be recalculated for each ω.

We will finish this section by calculating the modes of tapered Mestel disc using these two methods for the response matrix. We find close agreement between our predictions and those of Evans & Read (1998). We also calculate the least stable mode dynamically, by fitting the response coefficients to a small perturbation to our disc in subsection 5.2.2.

5.2.1 Volterra Method for Response Matrix

By definition for a mode we want $B_p(t) = \hat{B}_p e^{-i\omega t}$, where $\omega = \omega_0 + \mathrm{i}\eta$. For an unstable mode the growth factor, η, is greater than zero and ω_o is related to the pattern speed via $\omega_0 = \ell\Omega_p$. Putting this into Equation 2.16 and setting the external perturbation, $A_p = 0$, to zero we get

$$\begin{aligned}\hat{B}_p = &-\mathrm{i}(2\pi)^2 \sum_{\mathbf{m}} \int \mathrm{d}^2\mathbf{J} \Bigg\{ \int_{-\infty}^{t} \mathrm{e}^{\mathrm{i}\omega(t-t')} \times \\ &\left[\mathbf{m} \cdot \frac{\partial F(\mathbf{J})}{\partial \mathbf{J}} \mathrm{e}^{-\mathrm{i}\mathbf{m}\cdot\boldsymbol{\Omega}(\mathbf{J})(t-t')} \left(\hat{\psi}^{(p)}_{\mathbf{m}}(\mathbf{J})\right)^* \hat{\psi}^{(q)}_{\mathbf{m}}(\mathbf{J}) \right] dt' \Bigg\} \hat{B}_q,\end{aligned} \tag{5.1}$$

were we have swapped the order of integration. When calculating modes we will use the orthonormal Kalnajs basis function. Therefore we will drop any explicit reference to $\mathcal{E}_{pq}$ as it can be reduced to the identity. To progress we relabel our integration variable as $\tau = t - t'$. Using this substitution and by comparing to the modal equation,

$$\hat{B}_p = \hat{\mathcal{M}}(\omega)^{pq} \hat{B}_q, \tag{5.2}$$

we obtain an expression for the response matrix

$$\begin{aligned}\hat{\mathcal{M}}^{pq}(\omega) = &-\mathrm{i}(2\pi)^2 \sum_{\mathbf{m}} \int \mathrm{d}^2\mathbf{J} \times \\ &\left\{ \int_0^{\infty} \left[\mathbf{m} \cdot \frac{\partial F(\mathbf{J})}{\partial \mathbf{J}} \mathrm{e}^{\mathrm{i}(\omega - \mathbf{m}\cdot\boldsymbol{\Omega}(\mathbf{J}))(\tau)} \hat{\psi}^{(p)*}_{\mathbf{m}} \hat{\psi}^{(q)}_{\mathbf{m}} \right] d\tau \right\}, \text{for } \eta > 0.\end{aligned} \tag{5.3}$$

One can integrate over τ to obtain the standard expression for the response matrix (e.g. Binney & Tremaine, 2008)

$$\hat{\mathcal{M}}^{pq}(\omega) = -\mathrm{i}(2\pi)^2 \sum_{\mathbf{m}} \int \mathrm{d}^2\mathbf{J} \left[\mathbf{m} \cdot \frac{\partial F(\mathbf{J})}{\partial \mathbf{J}} \frac{\hat{\psi}^{(p)*}_{\mathbf{m}} \hat{\psi}^{(q)}_{\mathbf{m}}}{\omega - \mathbf{m} \cdot \boldsymbol{\Omega}} \right]. \tag{5.4}$$

If Appendix C we outline a method to calculate the remaining integrals. A slightly easier approach, however, is to evaluate the time integral second and express the response matrix in terms of the evolution kernel (Equation 2.17)

$$\hat{\mathcal{M}}^{pq}(\omega) = \int_0^{\infty} d\tau \mathcal{K}^{pq}(\tau) \mathrm{e}^{\mathrm{i}\omega\tau}. \tag{5.5}$$

Care must be taken when calculating this integral: if $\eta \leq 0$ then this method will not work as the integral will not converge and hence another approach must be taken; if $0 < \eta \ll 1$ then the kernel must be evaluated for large τ to ensure convergence. This can require many short time steps if $\eta \ll \omega_0$, making the computation of the kernel costly.

When calculating the response matrix we limit our potential-density pair expansion to the first 40 terms of the Kalnajs basis functions (section 2.3) and our sum over $-7 \leq m_1 \leq 7$. We expand each $\hat{\psi}^{(q)}_{\mathbf{m}}(\mathbf{J})$ on a uniform grid of 250 points in (R_+, R_-), which results in a spacing of 0.06 in R as the scale of the Kalnajs basis function is set as $R_{\rm ka} = 15$ (section 2.3 for more details). These parameters ensure convergence to three significant figures in the calculation of the response matrix. Choosing the step size and the upper limit of the τ integral is more complex as they will depend on ω: a quickly oscillating mode will need a finer step size, but a more slowly decaying mode will require a longer integration time. For the comparison in subsection 5.2.3 we integrate up to $\tau = 400$ with a step size of 0.5.

Once we have calculated the response matrix – either via the Volterra approach (subsection 5.2.1) or Laplace approach (Appendix C (Fouvry et al., 2015b)) – the Landau modes of the disc can be found. A mode of a disc must satisfy Equation 5.2. Therefore to calculate ω we must solve

$$\det\left[\hat{\mathcal{M}}(\omega) - \mathcal{I}\right] = 0, \tag{5.6}$$

where $\mathcal{I}$ is the identity matrix. To solve this equation we calculate $\det\left[\hat{\mathcal{M}}(\omega) - \mathcal{I}\right]$ on a uniform grid and search for any roots. In Figure 5.1 we plot the LHS of Equation 5.6 using the kernel method to calculate the response matrix for a disc with $\sigma_R = 0.377$ and inner taper index $\nu_t = 4$. From this figure we can clearly see four different modes, the most unstable of which has $\eta \approx 0.14$. Once we have found an eigenfrequency of the disc we can then find structure of the mode by solving Equation 5.2 for the coefficients that represent the mode, $\hat{B}_p$. In Figure 5.2 we plot the real space representation that corresponds to the most unstable mode of Figure 5.1.

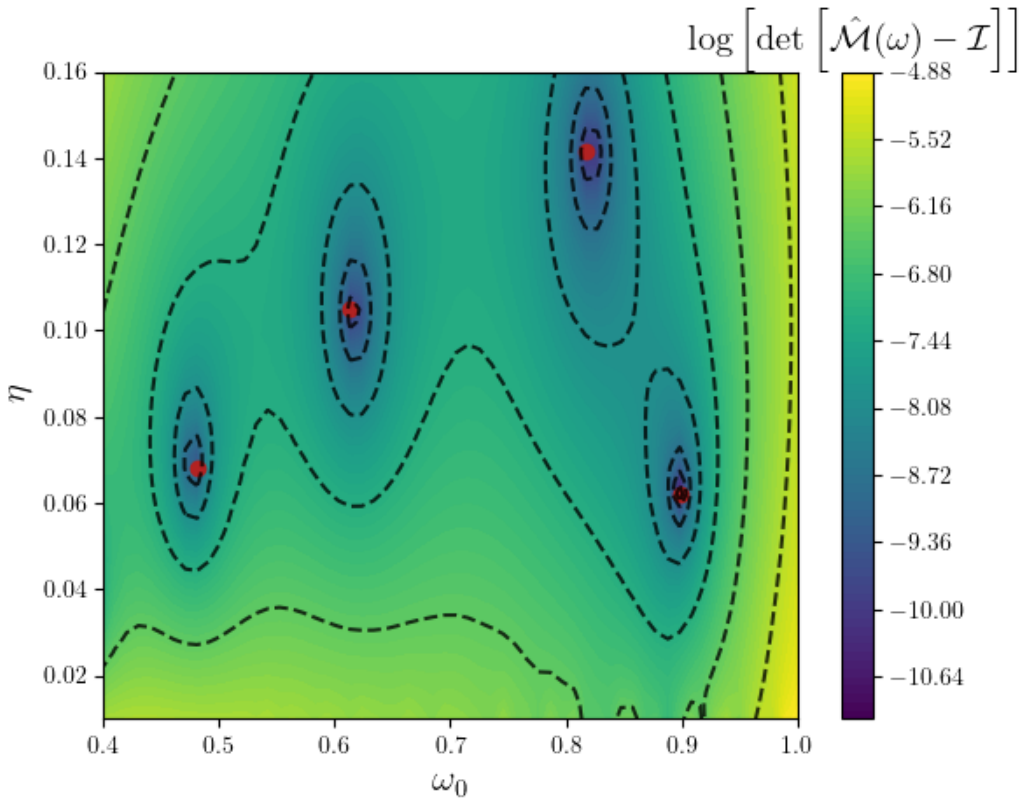


Figure 5.1: A contour map of det $\left[\hat{\mathcal{M}}(\omega) - \mathcal{I}\right]$ with an inner taper index $\nu_t = 4$. We see four roots (marked by the red dots), the most unstable of which (largest η) corresponding to $\omega = 0.82 + 0.14\mathrm{i}$, in agreement with Evans & Read (1998).

5.2.2 Coefficient Fitting

If a disc is unstable, after any perturbation one would expect this least stable mode to dominate the response. One could calculate the frequency of this mode by perturbing the disc with an impulse and fitting the late time evolution of the response coefficients to the form $\hat{B}_p \exp(-\mathrm{i}\omega t)$. If an unstable (stable) disc is perturbed the response will grow (decay) exponentially, with the most unstable mode dominating the response at late times. By fitting each coefficient with a global amplitude and phase of the response coefficients at late times we can estimate ω.

To do this we use an impulse perturbation at $t = 0$ with only the lowest Kalnajs coefficent non-zero, i.e

$$A_q(t) = \begin{cases} \delta(t),\ q = 0, \\ 0,\ \text{otherwise.} \end{cases} \tag{5.7}$$

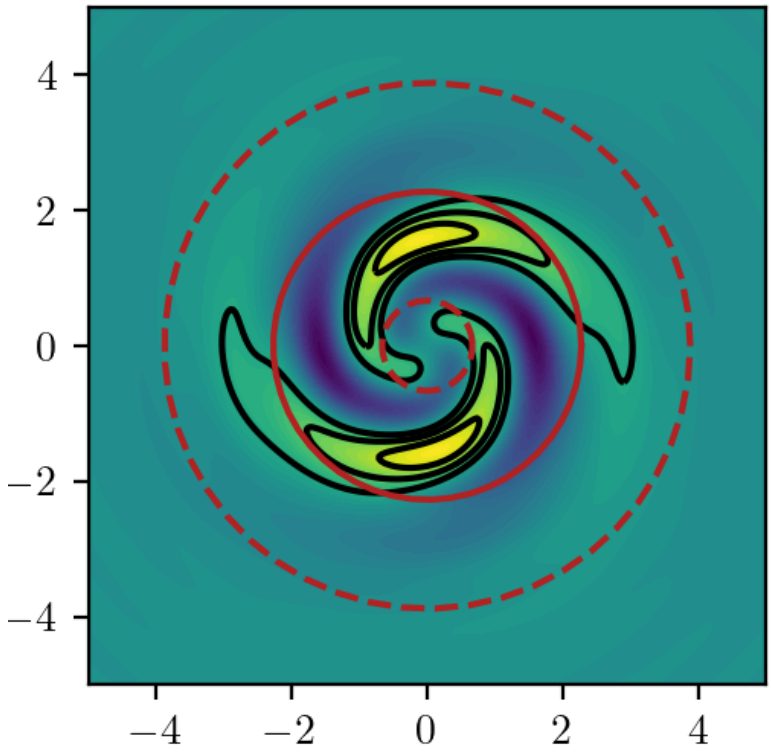


Figure 5.2: The real space representation of the most unstable mode in Figure 5.1 ($\omega = 0.83 + 0.14$i) for the $\nu_t = 4$ full mass tapered Mestel disc ($\xi = 1$). The dotted red circle shows the position of the ILR and OLR and the solid red circle shows the position of CR.

The results are independent of which q we choose to be non-zero. In Figure 5.3 we plot the evolution of amplitude (right panel) and the derivative of the argument with respect to time (left panel) of a selection of coefficients. At late times the amplitude tends towards an exponential form, given by the growth rate of the most unstable mode, while at late times the derivative of the argument tends to a constant value given by $-\omega_0$. This plot is generated using the test for inner taper $\nu_t = 6$ in subsection 5.2.3. When calculating ω we average the results from each coefficient. To calculate the response we use the same kernels as outlined in subsection 5.2.1.

This method has two limitations: it can only find the most unstable mode and if the growth rate (be it positive or negative) of the most unstable mode is not well separated from the next most unstable then calculating the most unstable mode is difficult. Of course the evolution could always be run for a longer time, allowing the most unstable mode to dominate, but this is bound to introduce numerical errors.

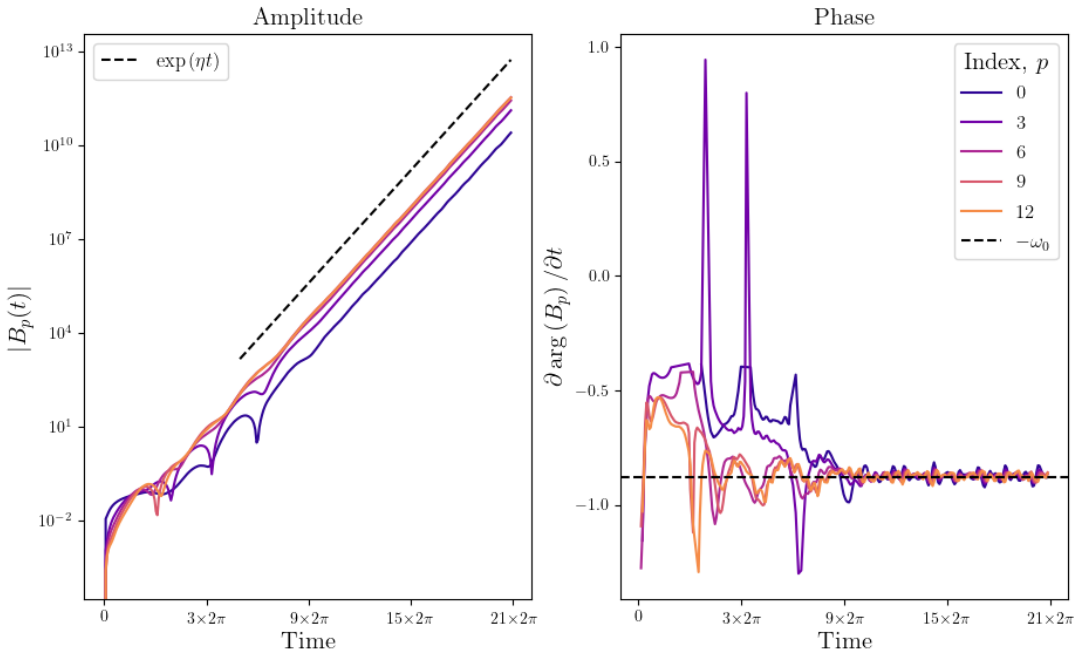


Figure 5.3: The response coefficients when the disc has been perturbed with an impulsive perturbation at $t = 0$. The left hand panel shows the evolution of the magnitude of a selection of coefficients which at late time tend to the same exponential growth shown by the black dashed line ($\exp(\eta t)$). The right hand panel shows the derivative of the argument of a selection of coefficients which at late time tend to constant value ($-\omega_0$).

Method	$\nu_t = 4$		$\nu_t = 6$		$\nu_t = 8$	
	ω_0	η	ω_0	η	ω_0	η
Volterra Method	0.83	0.14	0.88	0.22	0.90	0.26
Laplace Method	0.82	0.13	0.90	0.22	0.92	0.27
Coefficient Fitting	0.82	0.14	0.88	0.22	0.90	0.26
Evans & Read (1998)	0.88	0.13	0.90	0.22	0.92	0.27

Table 5.1: A comparison between the different methods for calculating the most unstable modes in disc with inner taper ν_t. The Volterra approach calculates the response matrix using the evolution kernel and integrating over τ (subsection 5.2.1). The Laplace approach calculates the response matrix via its Laplace transformed expression outlined in Appendix C. The coefficent method fits the response coefficients to the form $\propto \mathrm{e}^{-\mathrm{i}\omega t}$.

5.2.3 Modes of the Tapered Mestel Disc

In order to compare our three approaches (kernel response matrix, coefficient fitting and direct integration) and to validate our implementation against the external work of Evans & Read (1998) we will calculate the most unstable mode of the full mass ($\xi = 1$) tapered Mestel disc. For this disc we will take $\sigma_R = 0.377$ ($q = 6$).

We set the radius of the inner taper at $R_i = 1$ with a power law index ν_t. We set the radius of the outer taper as $R_o = 11.5$ with power law $\eta_t = 5$. We give explicit functional form of these tapers in subsection 2.3.2.

In Table 5.1 we present our results for the most unstable modes of the tapered Mestel disc with different inner tapers powers. Each of the three approaches are consistent with one another and show strong agreement with the previous work of Evans & Read (1998).

Two clear trends can be seen in Table 5.1: first, although there is a slight positive correlation between ω_0 and ν_t, changing ν_t does not make a large difference; second, there is a strong positive correlation between η and ν_t, i.e. the sharper the taper the stronger the growth rate.

5.2.4 Active Fraction & Modes of the Mestel Disc

We can check convergence of the Volterra method (subsection 5.2.1) by comparing its results to coefficient fitting (subsection 5.2.2). In Figure 5.4 we plot the least stable mode of the tapered Mestel disc while varying the active fraction ξ. The dashed line corresponds to $\nu_t = 4$ and the solid line for $\nu_t = 8$, both calculated by fitting coefficients. The dots corresponds to $\nu_t = 4$ and the crosses for $\nu_t = 8$, both found using the response matrix calculated using the kernel. Despite the two different approaches, there is strong agreement suggesting that the integrals converge.

As we approach the stable modes, both methods fail: the coefficient fitting method fails as there are multiple stable modes with similar decay rates, making it hard to identify a single least stable mode; likewise, the kernel method no longer works as the integrals become prohibitively expensive to compute to convergence.

5.3 Modes of Scarred Discs - Angular Momentum Scars

Now that we have developed a method to calculate the modes of a disc, we will return to the task we set ourselves at the beginning of this chapter: to identify cavity modes in a scarred disc.

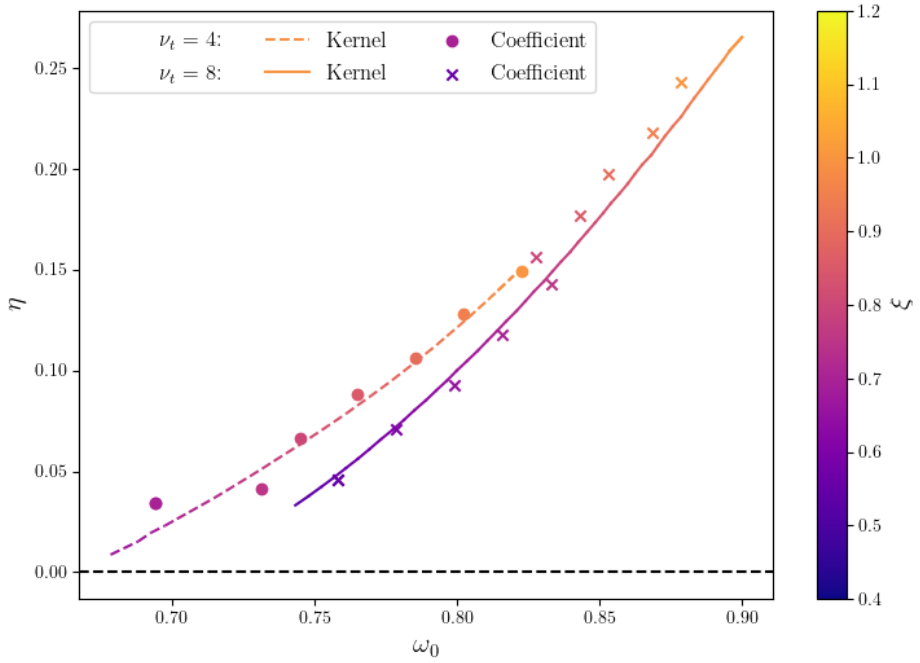


Figure 5.4: The least stable mode as a function of the active fraction of the disc, ξ. The dashed line corresponds to $\nu_t = 4$ and the solid line for $\nu_t = 8$, both found using the response matrix calculated using the kernel
. The dots corresponds to $\nu_t = 4$ and the crosses for $\nu_t = 8$, both calculated by fitting coefficients. Despite the two different methods, there is good agreement between the two approaches.

5.3.1 WKB Cavity Modes

If a mode is a standing wave between two boundaries then the fundamental harmonic must obey the quantum condition (e.g. Jackson, 1999),

$$\oint_{R_{\rm in}}^{R_{\rm out}} k(\omega_0, R) dR = \pi, \tag{5.8}$$

where $k(\omega_0, R)$ is the radial wave number, $R_{\rm in}$ is the inner and $R_{\rm out}$ is the outer boundary of the cavity. The cavity's inner boundary will either be due to the ILR (set by ω_0) or the inner scar, whereas the outer boundary will be the forbidden zone around CR. To calculate $k(\omega_0, R)$ we will use the LSK dispersion relation outlined in subsection 4.2.1. To solve Equation 5.8 to obtain ω_0 we split the solution of the LSK dispersion relation into two branches: the long branch, to the left of the turning point in Figure 4.1, with solution $k_{\rm long}$; and the short branch, to the right of the turning

point in Figure 4.1, with solution $k_{\rm short}$. The path integral in the quantum condition, which is along a path from $R_{\rm in}$ to $R_{\rm out}$ and back again, can then be into two parts,

$$\int_{R_{\rm in}}^{R_{\rm out}} \left[k_{\rm short}(\omega_0, R) + k_{\rm long}(\omega_0, R)\right] dR = \pi. \tag{5.9}$$

The lower limit, $R_{\rm in}$, is the larger of the ILR or the inner scar. The upper limit, $R_{\rm out}$, is the inner forbidden radius.

Once the spatial frequency is obtained we can calculate the growth rate of the cavity mode, using eq. 17.68 from Bertin (2014),

$$\int_{R_{\rm in}}^{R_{\rm out}} \left[\frac{1}{|v_{\rm g,\ short}(R)|} + \frac{1}{|v_{\rm g,\ long}(R)|}\right] dR = \frac{\log 2}{2\eta}, \tag{5.10}$$

where the integral can be thought of as the time for the wave to propagate between the two boundaries of the cavity.

To validate the implementation of Equation 5.8 we first compare our results to De Rijcke et al. (2019b). In De Rijcke et al. (2019b) the modes of a tapered Mestel DF that had been scarred by secular diffusion (Sellwood, 2012; Fouvry et al., 2015b) were calculated. They compared the eigenfrequencies of the modes with the same WKB approximation that we use. They found strong agreement between the prediction and true modal frequency, indicating the existence of resonance modes.

For comparison to De Rijcke et al. (2019b) we take their parameter values, namely: the softening length as $\epsilon = 0.125$, $q = 11.4$ and inner scar radius of $R_{\rm in} = 1.2$. We use the same surface density as De Rijcke et al. (2019b), which ignores the tapers (i.e. using Equation 4.3 to approximate the surface density). We find a cavity mode of frequency $\omega = 0.544 + 0.012$i which agrees with De Rijcke et al. (2019b)'s estimate of $\omega = 0.510 + 0.017$i. Reassuringly, however, our estimate is in better agreement with their true calculation of the mode, $\omega = 0.597 + 0.013$i , than their estimate, giving support to their conclusions and our own implementation of Equation 5.9 and 5.10.

Note, that as the WKB assumption is the limiting approximation in the LSK dispersion relation, we do not include the scars or tapers in the calculation of the reduction factor, $\mathcal{F}$.

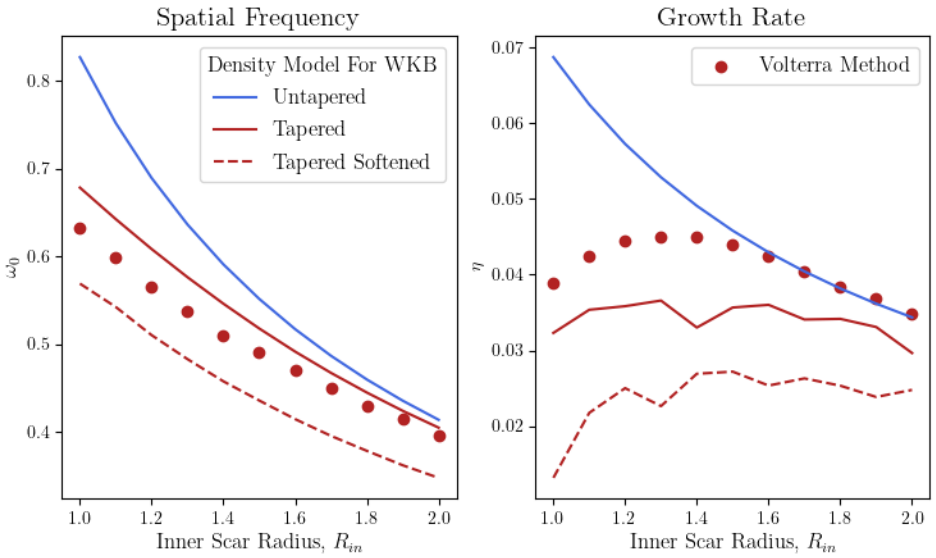


Figure 5.5: A comparison between the modal frequencies calculated via the response matrix using the Volterra method, which are represented by the red dots, and the WKB approximation, represented by lines. *Left Panel*: Comparison of the spatial frequencies, ω_0. *Right Panel*: Comparison of the growth rates, η. We estimate the eigenfrequencies using three different models in our WKB approximation: an untapered Mestel Disc (solid blue line), a tapered Mestel disc for (solid red line) and a tapered softened Mestel disc, with $\epsilon = 1/8$ (dashed red line).

5.3.2 Scar Model

For our scarred background DF, $F_{\rm s}(E, L)$, we simply take,

$$F_{\rm s}(E, L) = \xi T_{\rm inner}(L) T_{\rm outer}(L) \mathcal{S}(L) F_m(E, L), \tag{5.11}$$

where $\mathcal{S}(L)$ is our scar function. In De Rijcke et al. (2019b) a scar model was chosen to best reproduce the secular scarring found in Sellwood (2012) and Fouvry et al. (2015b). Instead, we take an easier approach and assume that the scar only depends on angular momentum. This will allow us to control the position of the inner cavity wall more closely, and hence better test for the existence of a resonant cavity. For this model we assume a scar function

$$\mathcal{S}(L) = 1 + A_{\rm s} {\rm e}^{-\frac{(L - v_c R_{\rm s})^2}{2(v_c \Delta_{\rm s})^2}}, \tag{5.12}$$

where $R_{\rm s}$ and $\Delta_{\rm s}$ is the central location and width of the scar and $A_{\rm s}$ is its depth. When calculating the integrals in Equation 5.8 we must integrate from the inner

boundary of the cavity, $R_{\rm in}$. We set this inner limit as the position of the scar i.e. $R_{\rm in} = R_{\rm s}$. Unless otherwise stated we take $\Delta_{\rm s} = 0.25$, chosen to be small enough to make the scar local while also being resolved by the number of basis function, and $A_{\rm s} = 0.95$, so that the groove is deep enough to excite a strong response. For the background tapered Mestel disc we take the same parameters are De Rijcke et al. (2019b); Sellwood (2012), namely: $q = 11.4$, $R_{\rm i} = 1$, $R_{\rm o} = 11.4$, $\nu_t = 4$, $\mu_t = 5$ and $\xi = 1/2$. We ignore gravitational softening when calculating the WKB approximation by setting $\epsilon = 0$ in Equation 4.1. We return to the effect of softening in the following section.

When calculating the evolution kernel we limit our potential-density pair expansion to the first 48 terms of the Kalnajs basis functions (section 2.3) and our sum over $-4 \leq m_1 \leq 4$. We expand each $\hat{\psi}^{(q)}_{\mathbf{m}}(\mathbf{J})$ on a uniform grid of 250 points in (R_+, R_-), which results in a spacing of 0.06 in R as the scale of the Kalnajs basis function is set as $R_{\rm ka} = 15$. Due to the relatively slow growth rate of grooved modes (e.g. De Rijcke et al., 2019b), we calculate the kernel up to $\tau = 400$, with a time step of 0.5. In searching for the modes we calculate the response matrix using the kernel method on a square grid with $\omega_0 \in [0.3, 0.6]$, $\eta \in [0.02, 0.05]$ with sixty points each.

Changing the position of the scar will change the mass of the disc, which will have its own affect on the modes of the disc (in essence by changing the active fraction in subsection 5.2.4, this is what we were studying). In order to negate this fact so that we can compare the response of each disc with a different scar radius, we set the mass of each scarred disc to the mass of the unscarred, but still tapered, Mestel disc. For the parameters used, the mass of the unscarred disc is 11.7.

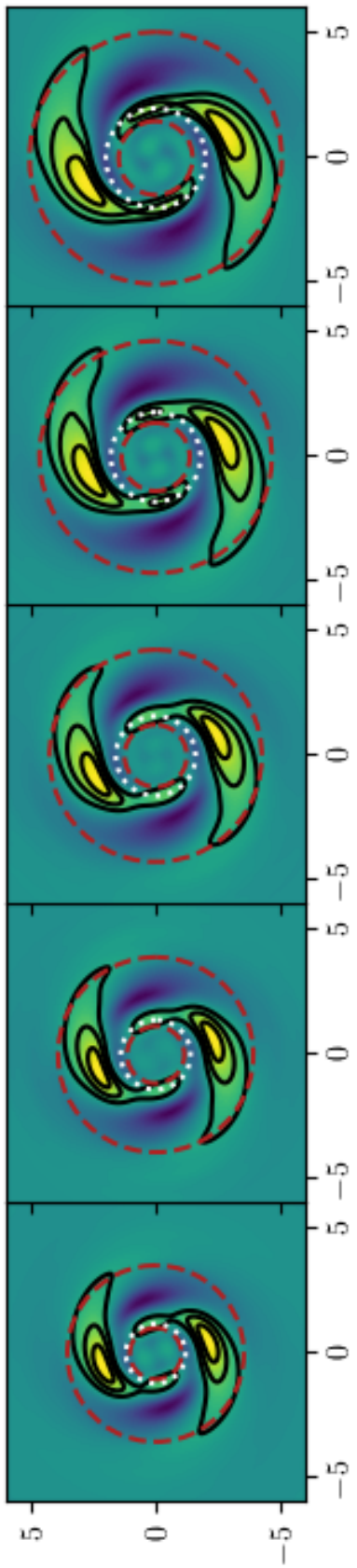


Figure 5.6: The cavity modes of the tapered Mestel disc with different scar radii. The dashed red lines indicate the position of the ILR and CR. The OLR is not plotted as the density perturbation is effectively zero at this large radii. The dotted white line indicates the position of the inner scar, R_s. The contours are linearly spaced between 20% and 80% of the maximum values of the mode for the positive density only.

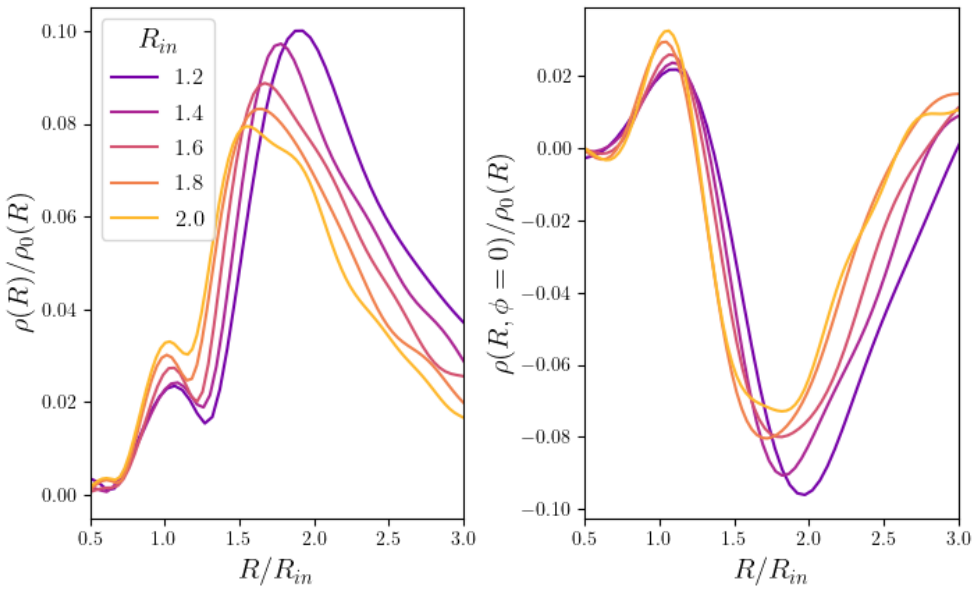


Figure 5.7: We plot the density profile of modes with different scar radius, R_{in}, in radial units of the scar radius. *Left panel*: the mode's density amplitude normalised by the Mestel background. *Right panel*: the mode's density along the x-axis normalised by the Mestel background. For larger scar radii, the mode profiles tend towards one another suggesting that the scar radius is setting the scale length.

5.3.3 Cavity Modes of Scarred Discs

Figure 5.5 shows a comparison between the modal frequencies calculated via the response matrix, which are represented by the red dots, and the WKB approximation using different models, represented by lines. We estimate the eigenfrequencies using three density models: first, an untapered Mestel Disc that has a density profile produced by Equation 4.3, represented by a solid blue line; second, a tapered Mestel disc for which we calculate the density by $\Sigma = \int d^2\mathbf{v} F_0$ where F_0 includes the inner and outer taper, represented by a solid red line; finally, a tapered softened Mestel disc, with $\epsilon = 1/8$ which corresponds roughly to the softening calculated in section 2.5, represented by a dashed red line.

In the left hand panel of Figure 5.5, where we plot the spatial frequency, there is reasonable agreement between the tapered Mestel disc WKB estimate and the Volterra method, but we do see a constant overestimation of the WKB frequency. The dashed line uses the tapered density profile, but includes softening, the effect of

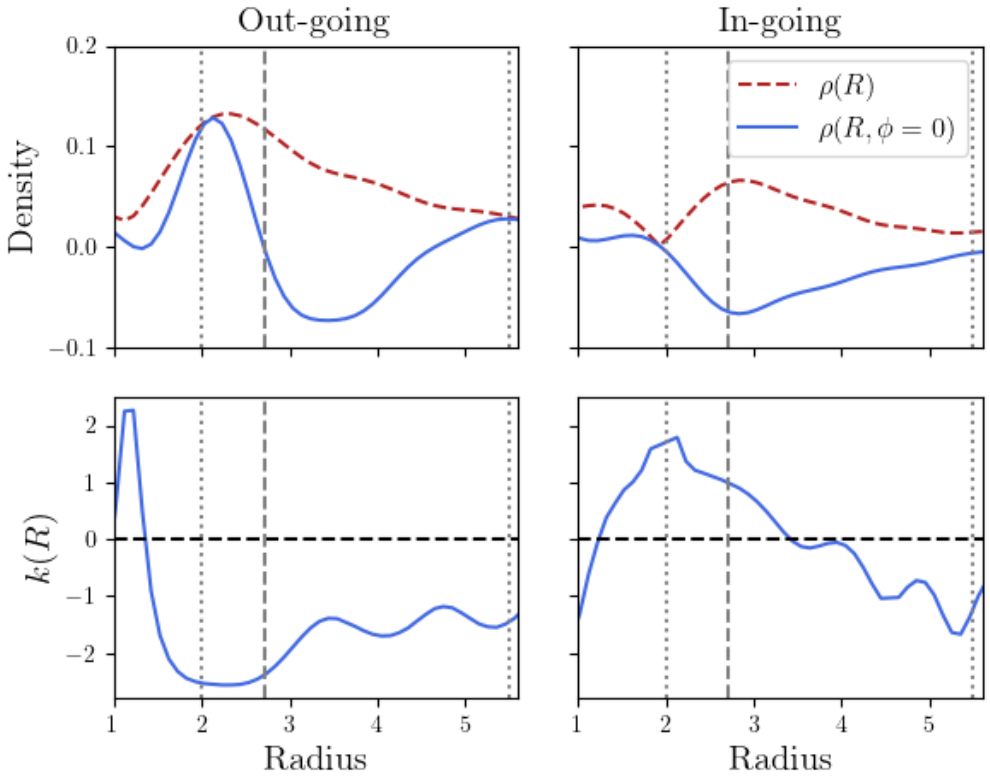


Figure 5.8: Decomposition of the cavity mode with $R_s = 2$ into out-going and in-going waves. The upper row shows the density amplitude $\rho(R)$ (red dashed) and the density along the x-axis (blue solid), while the bottom row shows the wavenumber, $k(R)$. The left hand column shows the out-going waves whereas the right hand column shows the in-going waves. The vertical lines show the positions of scar radius
, the forbidden radius and CR from left to right. As expected for the swing amplifier feedback cycle (Toomre, 1977, 1981; Bertin et al., 1989), the out-going waves are leading $k < 0$ whereas the in-going waves are trailing ($k > 0$).

which is to decrease the estimated frequency, suggesting that softening could be the cause of this systematic shift. For the softening we take $\epsilon = 1/8$ a rough estimate from section 2.5 given the number of basis functions. A direct inclusion of softening is hard, however, as the WKB approximation includes conventional 'Plummer softening', where the Green's function of Poisson's equation is modified, whereas the softening in the Volterra method is due to the finite basis expansion (c.f. section 2.5).

In the right hand panel, where we plot the growth rate, there is reasonable agreement between the tapered Mestel disc WKB estimate and the Volterra method. However we do see a constant underestimation of the WKB frequency. This is

due to our overestimation of the spatial frequencies: we calculate ω_0 first using Equation 5.9 then use this estimate to calculate η using Equation 5.10, so any error in ω_0 is liable to propagate through. It is worth noting, however, that given the questionable assumptions behind the WKB approximation, the fact the agreement is this good is rather surprising. This agreement is a good indicator that these modes are due to cavity mechanism outlined in subsection 5.3.1.

Further evidence for this claim comes when the cavity modes are plotted in their density representation in Figure 5.6. The close correspondence between the scar radius (plotted as a dotted white line) and the ILR of the mode is indicative of a cavity mode, while limited extent beyond CR also indicates the constrained nature of a cavity mode. To show that it is the position of the scar, rather than the inner taper, setting the scale of the mode we plot the density amplitude, $\rho(R)$, as a function of $R/R_{\rm in}$ in Figure 5.7. For scars further away from the inner taper ($R_{\rm i} = 1$), their radial profile is independent of $R_{\rm in}$, whereas those closer to the inner taper feel its effects. Note that we normalise the density by the background Mestel density so that we can compare the absolute values.

In Figure 5.8 we plot the decomposition of the cavity mode with $R_s = 2$ into out-going and in-going waves. The upper row shows the density amplitude, $\rho(R)$ (red dashed line), and the density along the x-axis, $\rho(R, \phi = 0)$ (blue solid line), while the bottom row shows the wavenumber, $k(R)$. The left hand column shows the out-going waves whereas the right hand column shows the in-going waves. The vertical lines show the positions of ILR, the forbidden radius and CR from left to right. To calculate the response, we obtain the eigenvector of the response matrix, $\hat{B}_q$, corresponding to an eigenvalue of unity once ω has been found. The evolution of the mode is then simply $\hat{B}_q \exp(-\mathrm{i}\omega t)$. Once expressed as a density response in real space, we can decompose the response into out-going and in-going waves using the method outlined in chapter 4. We can clearly see that the out-going waves are leading $k < 0$ whereas the in-going waves are trailing ($k > 0$), agreeing with the prediction of the swing amplifier feedback cycle (Toomre, 1977, 1981; Bertin et al., 1989). We can also observe that the scar (the inner dotted line) works well as a

barrier for in-going waves as the amplitude there approaches zero. The forbidden zone, however, works less well as it leaks out-going waves in agreement with the angular momentum predictions from Mark (1974).

5.4 Conclusions

Our aim in this chapter was to confirm the existence of the swing amplifier feedback cycle (Toomre, 1977, 1981; Bertin et al., 1989) in a scarred disc. In particular cavities provide an answer to an important question: how are leading density waves formed? As these waves are particularly susceptible to swing arm amplification, a subject that we studied in chapter 3, their production is an important stage in the production of large scale spiral structure.

We first developed a method to calculate the modes of a disc using the evolution kernel outlined in chapter 2. The modes calculated using our kernel method agrees well with the previously calculated modes of the tapered Mestel disc (Evans & Read, 1998). However in the marginally unstable and stable case our method fails due to numerical convergence issues. We illustrated this by decreasing the active fraction of the disc and comparing our response matrix method to the dynamical response of the disc.

In order to study the swing amplifier feedback cycle we added a scar to the background DF that depends only on angular momentum. We chose this, rather than a more complicated scar based on Jacobi-Integral (De Rijcke et al., 2019b), as it gave us clear control over the position of the scar, allowing us to test the wave cavity hypothesis further. To confirm the existence of the feedback cycle we first compared the eigenfrequencies of a scarred disc calculated via the response matrix to a WKB based approximation, that assumed the inner scar and forbidden radius act as the walls of a density wave cavity, supporting a standing wave between them. The strong agreement indicates the existence of such a cavity. We further confirm that the inner scar reflects the in-going density wave, protecting it from absorption at the ILR, by decomposing the density response into in-going trailing

and out-going leading density waves, agreeing with the predictions of the swing amplifier feedback cycle (Toomre, 1977, 1981; Bertin et al., 1989).

"I'm not great at happiness, but am excellent in relief."

— B. Nighy

6 Bars

Contents

6.1 Introduction 115
6.2 Bar Models 116
6.3 Linear Torque Calculation 119
6.3.1 An expression for the steady-state torque 121
6.3.2 Comparison with N-body simulation 123
6.4 Limitations Of Linear Theory 124
6.4.1 Varying the Strength of the Bar 124
6.4.2 Orbital Sections of Trapped Orbits 125
6.5 Conclusion 128

6.1 Introduction

In the final chapter of this thesis, we use the formalism developed in chapter 2 to study dynamical friction within the disk: we perturb the disc by a rotating bar-like potential and calculate the frictional torque that the disc's response exerts on the imposed bar.

Variants of this problem have been already studied in great detail, making use of both linear response theory and N-body simulations (e.g. Tremaine & Weinberg, 1984; Sellwood, 2006; Banik & van den Bosch, 2021; Chiba & Schönrich, 2022, among others). Lynden-Bell & Kalnajs (1972) first studied the response of a disc to

a bar with constant pattern speed, deriving the LBK formula which gives the torque acting on a bar once any transient responses have died away. Banik & van den Bosch (2021) applied a similar formalism while relaxing the adiabatic assumption and found that the torque is a combination of two different parts: a transient component and a steady-state component. The transients are damped, not due to a dissipative process, but rather due to the phase mixing of different orbits. As the transients are inherently a time dependent phenomena, we can use the formalism laid out in the previous chapter to study more closely the torque they apply to the bar.

When studying a perturbation rotating at a constant pattern speed, linear theory quickly runs in to the problem of trapped orbits. The standard approach to deal with the divergences at resonance is to transform into a frame that rotates at the resonant frequency (pp. 109-117 Lichtenberg & Lieberman, 2013), which allows for the construction of a new coordinate system: the so-called 'slow' and 'fast' actions (e.g. Lynden-Bell & Kalnajs, 1972; Tremaine & Weinberg, 1984; Chiba et al., 2021). Tremaine & Weinberg (1984) focused on friction in the secular adiabatic limit and showed that only trapped orbits contribute to the total torque on the bar in this regime. As trapped orbits play a key part in the story of bar formation, justifying the linear assumption via comparison to a non-linear prediction is crucial. In this chapter we will aim to understand the limitations of linear theory when studying the exchange of angular momentum between a disc and bar.

In the following section we will outline our bar model. Following that, in section 6.3, we will compare linear response theory, and its steady-state torque prediction, to non-linear N-body simulations in the weak bar regime. In section 6.4 we will increase the size of the bar and study the non-linear effects that are introduced in N-body simulations – most notably orbital trapping.

6.2 Bar Models

To calculate the potential profile of our bar we employ the model used in Sormani et al. (2015). In Sormani et al. (2015) a quadrupole potential was generated by

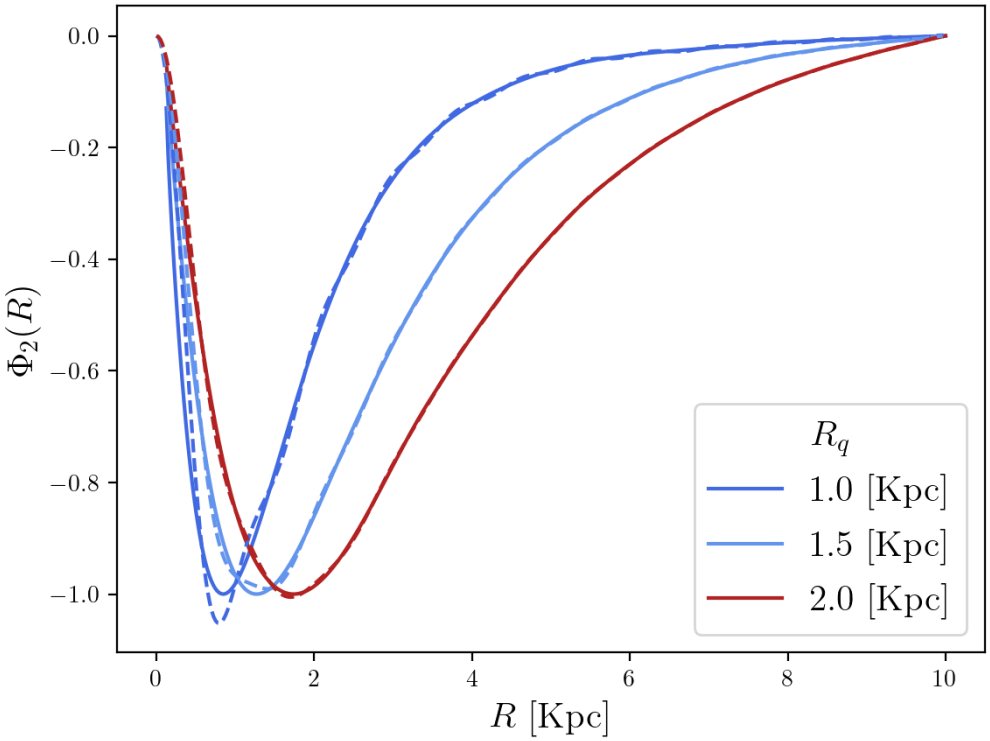


Figure 6.1: The radial potential functions, $\Phi_2(R)$, as a function of radius for different scale lengths, $R_{\rm q}$. The solid lines are the radial potential functions calculate solving Poisson's equation with density given in Equation 6.1. The dotted lines represent the Kalnajs basis function fit using 48 basis functions. We normalise $\Phi_2(R)$ such that the maximum absolute value of the potential is one so that the strength is all contained in ϵ.

a density distribution of the form,

$$\rho_2(r,\phi,\theta) = \frac{\epsilon}{r_{\rm q}^2}\exp\left(-\frac{2r}{r_{\rm q}}\right)\sin^2\theta\cos 2\phi, \tag{6.1}$$

where $\{r,\phi,\theta\}$ are spherical polar coordinates with $\theta = 0$ along the cylindrical z axis and ϵ sets the strength of the bar. Sormani et al. (2015) took the length scale of the bar as $r_{\rm q} = 1.5$ [Kpc]. The best fit parameters where found by comparing the gas at the centre of the Galaxy in the ℓ-velocity plane to fluid simulations. As the trigonometric terms in Equation 6.1 are an eigenfunction of the Laplacian in spherical polar coordinates, solving Poisson's equation for the potential becomes an ODE that can be solved numerically using an 'off the shelf' boundary value problem

solver. Doing so allows the potential sourced by the bar to be expressed as,

$$\Phi_2(r, \phi, \theta) = \epsilon M(t) \Phi_2(r) \sin^2\theta \cos 2\phi, \tag{6.2}$$

where $\Phi_2(r)$ is calculated numerically by solving

$$\frac{1}{r^2}\frac{d}{dr}\left[r^2 \frac{d\Phi_2(R)}{dr}\right] - \ell(\ell+1)\Phi_2(r) = 4\pi G\left[\frac{\epsilon}{r_{\rm q}^2}\exp\left(-\frac{2r}{r_{\rm q}}\right)\right], \tag{6.3}$$

subject to the boundary conditions $\Phi_2(r=0) = \lim_{r\to\infty}\Phi_2(r) = 0$.

In order to 'turn the bar on' we let $M(t)$ be a slowly increasing function, $M(t) \in [0, 1]$. We set $\theta = \pi/2$ to obtain the potential in the plane, so that

$$\Phi_2(R, \phi) = \epsilon M(t) \Phi_2(R) \cos 2\phi, \tag{6.4}$$

where we have replaced r with R to highlight we are once again in polar coordinates. In Figure 6.1 we plot $\Phi_2(R)$ for different scale lengths (solid lines) and their reconstruction using the first seventy-two Kalnajs basis functions (dashed lines). The fit works very well, even at small radii. We normalise $\Phi_2(R)$ such that its maximum value is unity, so that the strength of the potential perturbation is solely contained in ϵ.

For our bar model we will assume that it rotates at a constant pattern speed, $\Omega_{\rm p}$, so that the density (and hence potential) can be written as a simple rotation, i.e.

$$\Phi_2(R, \phi, t) = \Phi_2(R, \phi - \Omega_{\rm p} t). \tag{6.5}$$

(Sormani et al., 2015) fit to observational data gave a scale length 1.5 Kpc and pattern speed 40 $\mathrm{Kms}^{-1}\mathrm{Kpc}^{-1}$, which in our units, with $v_c = 1$, are: $\epsilon = 0.14$, $R_q = 1.5$ and $\Omega_p = 0.18$.

As explained above, we will use $M(t)$ to turn the bar on. Explicitly it will take the form

$$M(t) = \begin{cases} \sin^2\left(\dfrac{t}{4}\right), \, t \leq 2\pi \\ 1, \text{ otherwise,} \end{cases} \tag{6.6}$$

so that the bar grows over a single dynamical time at the position of the inner taper. We chose this short growth timescale to excite a strong transient response of the disc.

The torque acting on the bar is given by

$$\tau_{\rm b}(t) = \int d^3\mathbf{x}\rho^{\rm e}\left[\mathbf{x}\times(-\nabla\psi^f)\right] = \int d^3\mathbf{x}\rho^f\left[\mathbf{x}\times(\nabla\psi^{\rm e})\right] \tag{6.7}$$

where we have used Newton's third law in the final equality to swap the position of the response and the stimulus. As our disc is razor thin, we can represent its 3D response as $\rho^f(R,\phi,z) = \rho^f(R,\phi)\delta(z)$ (c.f. subsection 2.3.1), allowing us to trivially evaluate the integral with respect to z. From the special harmonic structure of the bar (Equation 6.4), the z component of the gravitational acceleration due to the bar's potential is zero in the plane of the disc. As the position vector in cylindrical coordinates is $\mathbf{x} = R\hat{\mathbf{R}} + z\hat{\mathbf{z}}$, the torque can be simply expressed as

$$\tau_{\rm b}(t) = \int \rho^f(R,\phi)\frac{\partial\psi(R,\phi)^{\rm e}}{\partial\phi}RdRd\phi. \tag{6.8}$$

By expressing the stimulus in potential density pairs, we can analytically evaluate the derivative with respect to ϕ, which results in replacing the derivative with a factor $\mathrm{i}\ell$. By expanding the response in potential density pairs, and remembering that the density and potential are real quantities resulting from the sum of $+\ell$ and $-\ell$ terms, we obtain

$$\tau_{\rm b}(t) = -\int d\mathbf{x}\left(\mathbf{x}\times\frac{\partial\psi^f}{\partial\mathbf{x}}\right)\rho^{\rm e} = -\mathrm{i}\ell\mathcal{E}^{\mu\nu}A_\mu(t)B^*_\nu(t), \tag{6.9}$$

where we have used the ortho-normality relation of the potential density pairs to obtain the final equality. As $A_p \sim \epsilon$ (c.f. Equation 2.16) and $B_p \sim \epsilon$ (by definition), from Equation 6.9 we see that the torque scales as ϵ^2. Note that as $A_{n,\ell} = A^*_{n,-\ell}$ and the same for $B_{n,\ell}$, where $p = [n,\ell]$, when correctly summed over the torque returns a real value.

6.3 Linear Torque Calculation

For the linear calculation we use the Kalnajs basis functions with 72 terms in the basis expansion. As we are evolving mean quantities, unlike N-body simulations which evolves individual particles, much larger time steps can be taken with the linear response methods, while retaining the same accuracy. As such we calculate

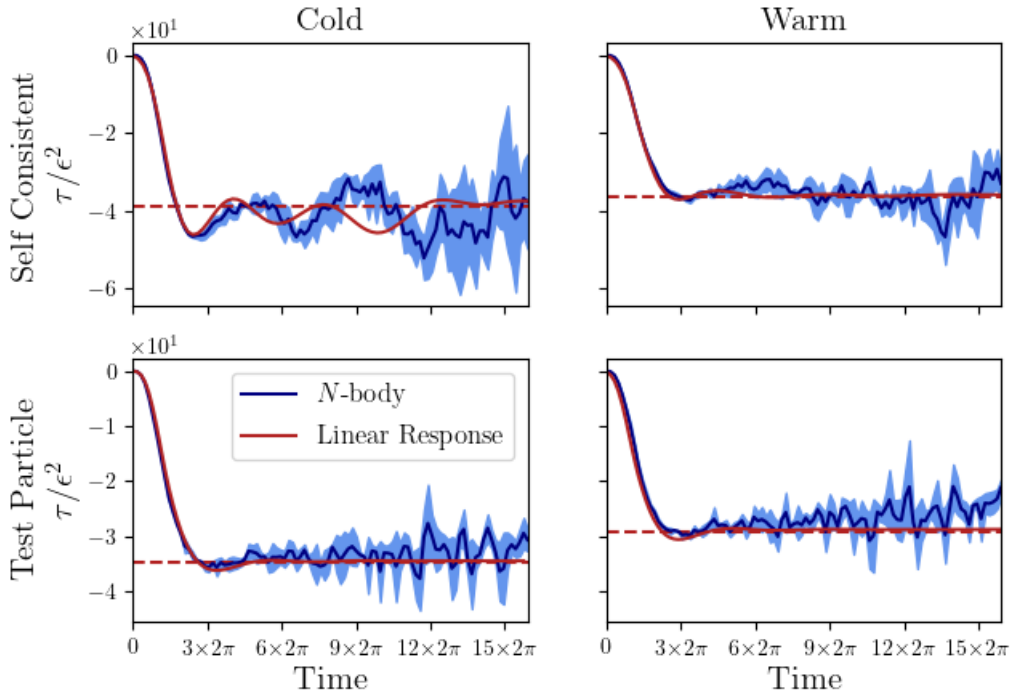


Figure 6.2: The frictional torque (red curve) exerted by discs of different temperatures, $\sigma_R = 0.35, 0.45$, when the Sormani bar with $R_{\rm q} = 1.5$ is imposed with pattern speed $\Omega_{\rm p} = 0.18$. The top row plots results when the self-gravity of the disc's response is included, while the bottom row shows corresponding results without self gravity for comparison. The blue curve shows torques measured from three N-body realisations with 10^6 particles, the blue shaded area indicating the standard deviation of these measurements. The horizontal dashed line is the steady-state torque calculated using Equation 6.13.

the kernel for 100 steps with a time step of 1. The resulting torque is plotted in Figure 6.2 for two different disc temperatures, $\sigma_R/v_{\rm c} = 0.35$ and 0.45 (left and right columns) and with (top row) and without (bottom row) the self-gravity of the response. We take the pattern speed as that which correspond to our own galaxy $\Omega_p = 0.18$. We deliberately take ϵ to be much smaller than the value for the real galaxy, to ensure that the response is in the linear regime, i.e. so that the terms of $\sim \epsilon^2$ in Equation 2.4 remain negligible compared to the linear terms. We will take $\epsilon = 5 \times 10^{-3}$ for the calculations in this section. We will discuss the issues with setting ϵ at a realistic value in section 6.4.

The response plotted in Figure 6.2 shows that the torque from the time-dependent linear response is a combination of two parts: a transient response that damps out and a longer-term, constant contribution from the steady-state density wake

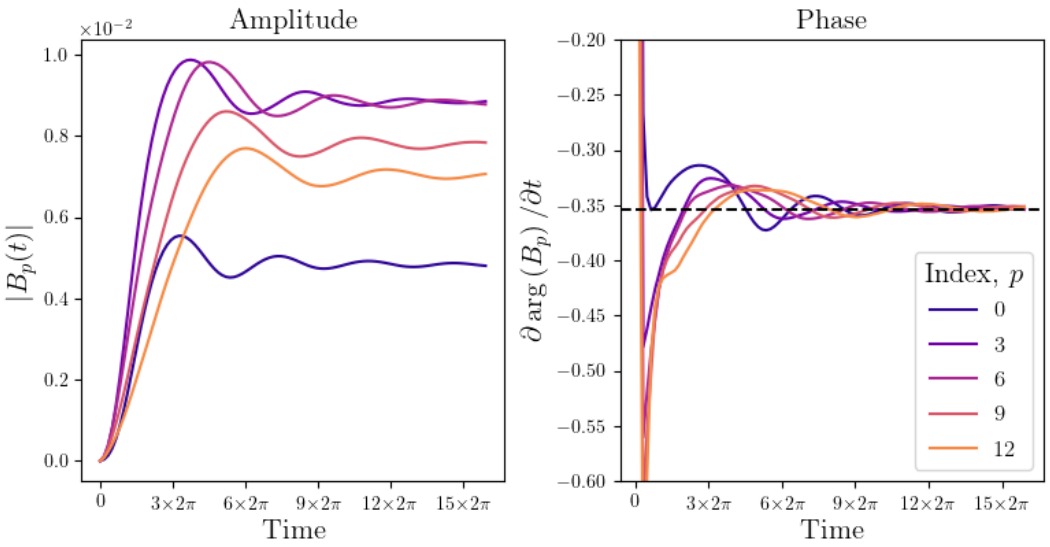


Figure 6.3: The evolution of the response coefficients to a steadily rotating bar. The left hand panel shows $|B_p|$, while the right hand panel shows the derivative of $\arg B_p(t)$ with respect to time. As the amplitude approach constant values and the derivative of the argument approach a constant value set by $-\ell\Omega_p$ (shown by the dashed black line), we can justify assuming a response of the form Equation 6.11.

that trails behind the bar. These two contributions have already been found for bars in halos (Banik & van den Bosch, 2021) and for the drag acting on a satellite pulled through the periodic cube (Magorrian, 2021). As expected self-gravity is more important for cooler discs; with less random motion, stars find it easier to 'unionise' and work together creating a coherent response, and hence a large torque.

Note that despite the weak strength of the bar, one can start to see the beginnings of non-linear behaviour. In particular, the slight drift back towards zero of the test particle simulations is indicative of the disc being spun up by the bar.

6.3.1 An expression for the steady-state torque

We can apply the method used in Magorrian (2021), chapter 4 and chapter 5 to calculate this longer term, steady-state torque. After the bar has developed the stirring potential is simply

$$A_p(t) = \hat{A}_p \mathrm{e}^{-\mathrm{i}\ell\Omega_\mathrm{p} t}, \tag{6.10}$$

in which the coefficients $\hat{A}_p$ are constant. We assume that the disc is stable, so that, after the transients have died away and the dynamical equilibrium established, its response will be

$$B_p(t) = \hat{B}_p \mathrm{e}^{-\mathrm{i}\ell\Omega_\mathrm{p} t}, \tag{6.11}$$

where the coefficients $\hat{B}_p$ are independent of time and dependent linearly on the imposed $\hat{A}_p$ [1]. Although this is a reasonable assumption, we can add further justification by plotting the evolution of the response coefficients. In the left hand panel of Figure 6.3 we plot the absolute value of different coefficients. It can clearly see that $|\hat{B}_p|$ tend towards a constant value. In the right hand panel we plot the derivative of the phase of each coefficient with respect to time, as expect this gradient tends to a constant with value $-\ell\Omega_p$ (represented by the black dashed line).

Making the change of variable $\tau = t - t'$ in Equation 2.16 gives

$$\mathcal{E}^{pq}\hat{B}_q = \mathcal{M}^{pq}(\Omega_\mathrm{p})\left[\hat{A}_q + \hat{B}_q\right]. \tag{6.12}$$

where $\mathcal{M}^{pq}(\Omega_\mathrm{p})$ is the response matrix (Equation 2.19 and chapter 5). It is unsurprising that the response matrix should appear, as by making the assumptions in Equation 6.10 and 6.11 we have carried out a modal analysis.

We can now combine our expression for the response in Equation 6.12 with our general expression for the torque on the bar in Equation 6.9 to obtain,

$$\tau_\mathrm{b}(t \to \infty) = -i\ell\mathcal{E}^{pq}\hat{A}_p\left[(\mathcal{E} - \mathcal{M})^{-1}_{qs}\mathcal{M}^{st}\hat{A}_t\right]^*. \tag{6.13}$$

an expression that has already been obtained by (Weinberg, 1989, equation 53) taking a slightly different route.

As the method outlined in section 2.2 gives the evolution of potential-density coefficients (rather than the perturbation of the DF), we naturally express the torque in these coefficients, leading to a concise expression for the torque. Unlike Banik

[1] Notice that an imposed potential of the form (6.10) would be continually imparting angular momentum to the disc, thereby spinning it up. We follow the usual justification in frequency-based analyses by adding a vanishingly small imaginary component to Ω_p so that $A_p(t)$ vanishes as $t \to -\infty$. We can then ignore the angular momentum imparted to the disc, because it scales as $|A_p|^2$: it is a second-order effect.

& van den Bosch (2021), however, but like Weinberg (1989), the expression (6.13) accounts for the self-gravity of the disc's response via the resonant denominator. We discuss the calculation of the response matrix in chapter 5 using the 'Laplace Method' (Fouvry et al., 2015b), with the same values for the basis function expansion as for the kernels.

The horizontal dashed lines in Figure 6.2 plot the torque obtained from Equation 6.13, showing that it agrees very well with the fully time-dependent linear response calculations. We stress, however, that the real strength of our Volterra method is to calculate the latter.

6.3.2 Comparison with N-body simulation

The results we have just presented are based on a number of approximations. For example, we had to introduce a truncated basis expansion to expand the density and potential response and we neglect the second-order $[f, \psi]$ term in the CBE (c.f. Equation 2.4). To test these assumptions we use N-body simulation to provide an independent measurement of the torque on the bar in which we isolate and evolve the $l = 2$ response of the disc only. This scheme and the basic operation of our N-body simulations are discussed in Appendix B.

These N-body measurements are included in Figure 6.2. In each panel the Figure plots the mean and and standard deviation of the torque measured from three realisations of the disc's response with 10^6 particles. The close agreement between the two different methods justifies the assumptions that we have made in our linear response calculations.

Although the linearity assumption works well for the problem of calculating the net torque on the bar in this particular case, there are nevertheless underlying nonlinear effects that have not been accounted for, most notably orbit trapping. In order to avoid these effects, we have set the strength of the bar to an artificially weak value. In the following section, we study the discrepancy between linear and non-linear theory when the strength of the bar is increased to a more realistic value.

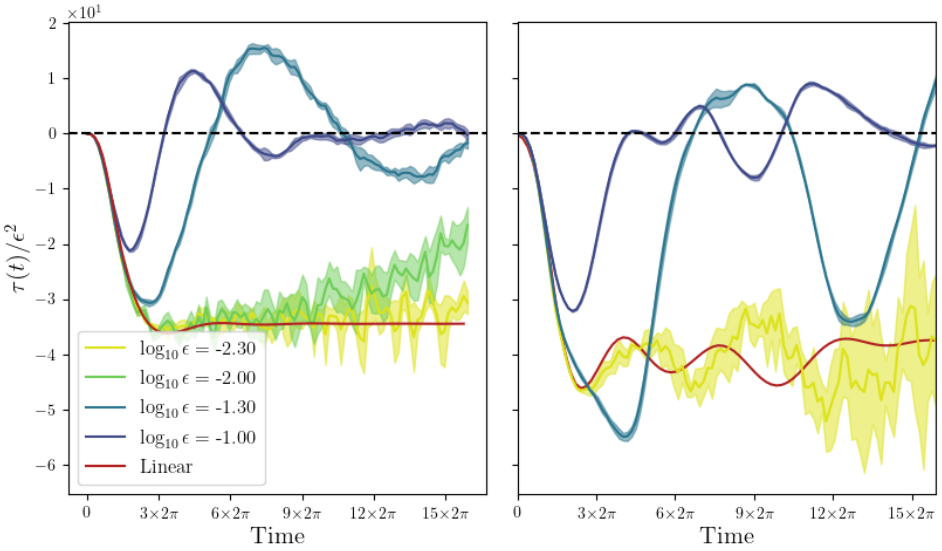


Figure 6.4: The torque acting on the bar due to the response to different strength bars. The left panels shows the test particle response whereas the right panel shows the self-consistent response. The red line shows the linear response calculation. The other lines show the mean N-body response for different strength bars with three realisations and the shaded area giving one standard deviation. As linear theory predicts a that the torque $\sim \epsilon^2$, we normalise by this so that any linear response is independent of ϵ. The weak bar agrees well with linear theory, however the response for the larger, more realistic bars quickly become non-linear.

6.4 Limitations Of Linear Theory

6.4.1 Varying the Strength of the Bar

In our previous calculations of the torque we have taken $\epsilon = 0.005$, despite the realistic value being $\epsilon = 0.14$. The reason for this is simple: we wanted to ensure that non-linear effects could be neglected and hence test our linear theory. We will now drop this selfish assumption and study the limitations of linear theory when applied to a bar of realistic strength.

In Figure 6.4 we plot the torque acting on the bar due to the response to different strength bars. We take the same cold bar model as the previous section with the pattern speed set at a realistic value, $\Omega_p = 0.18$. We use three realisation of the N-body simulations, from which we calculate a standard deviation, which is represented by the shaded region around the mean. The values of ϵ are given in the legend. We normalise the torque by a factor of $1/\epsilon^2$ as linear theory predicts $\tau_b \sim \epsilon^2$.

Initially, when non-linear effects have not had time to develop, the normalised torques agree with one another, indicative of a linear response. However, as time progresses, starting with the strongest bar, the N-body torque depart from the linear response calculation as non-linear effects start to develop. As expected, the non-linear torques will eventually return to zero, whereas the linear torque approaches a non-zero constant.

One way of estimating the timescale on which this effect *must* become important is by comparing the torque predicted by linear theory to the disc's total angular momentum: how long before the disc's response becomes comparable to the initial background DF? The initial angular momentum of the cold disc used above is $L_{\rm disc} = \int \mathrm{d}^2\mathbf{x}\mathrm{d}^2\mathbf{v} L_z F_0 = 40.4$. For the largest bar ($\epsilon = 0.1$), the torque at the first turning point is $\tau \sim 0.45$. Therefore the time taken for the bar to add 25% of the disc's starting angular momentum back into the disc is $0.25 \times L_{\rm disc}/0.45 \sim 3 \times 2\pi$, as this is less than the time of our simulation we expect to see non-linear effects.

6.4.2 Orbital Sections of Trapped Orbits

If the discrepancy in subsection 6.4.1 between linear response and N-body is due to orbital trapping, then we will be able to identify these orbits by plotting surface of sections. In the co-rotating frame we define x' as the axis along the major axis of the bar. The new coordinate system is a rotation of the old coordinate system,

$$\begin{aligned} x' &= x\cos\left(\Omega_p t\right) + y\sin\left(\Omega_p t\right), \\ y' &= y\cos\left(\Omega_p t\right) - x\sin\left(\Omega_p t\right), \end{aligned} \tag{6.14}$$

and the generalised momenta in the co-rotating frame then take the form,

$$\begin{aligned} p_{x'} &= \dot{x} - y\Omega_p, \\ p_{y'} &= \dot{y} + x\Omega_p. \end{aligned} \tag{6.15}$$

In this new coordinate system the Jacobi integral

$$H_J = H - \Omega_p L_z, \tag{6.16}$$

becomes a conserved quantity.

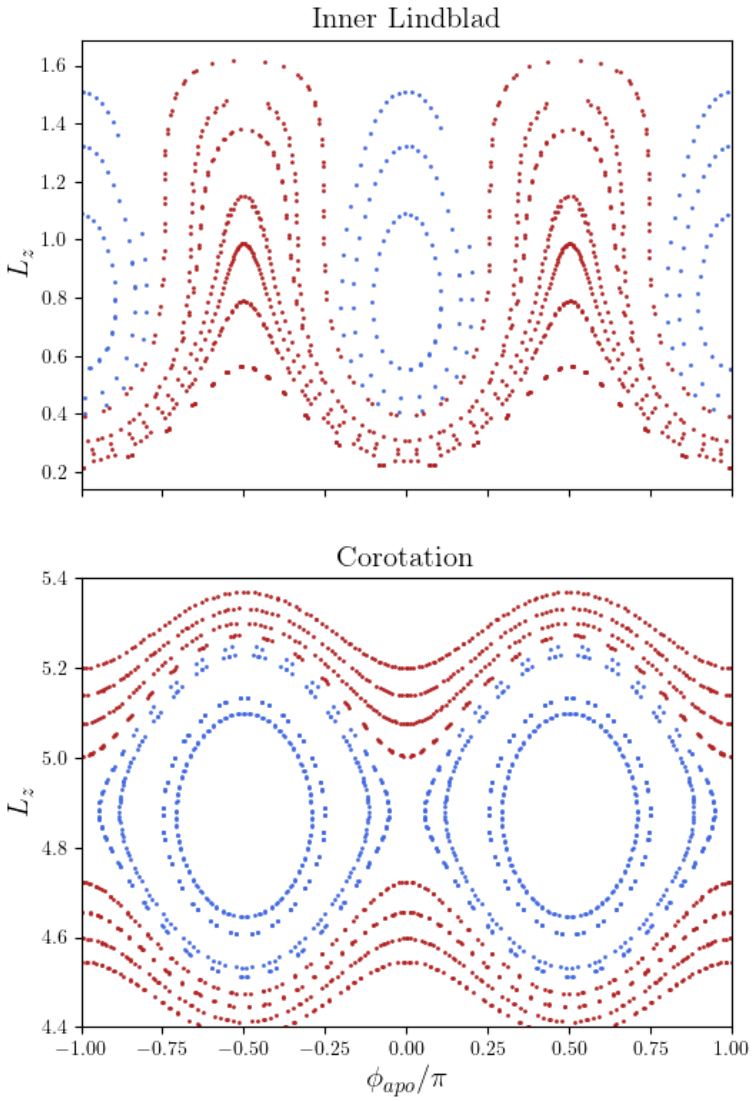


Figure 6.5: Surface of section for a bar with $\Omega_p = 0.18$. The upper hand panel shows orbits at ILR and the lower hand panel shows orbits at CR. We plot the angle relative to the bar, ϕ_{apo}, and the angular momentum, L_z, each time the particle passes through apocentre. The trapped orbits, plotted in blue, are radically altered by the perturbation, causing the assumptions of linear theory to breakdown.

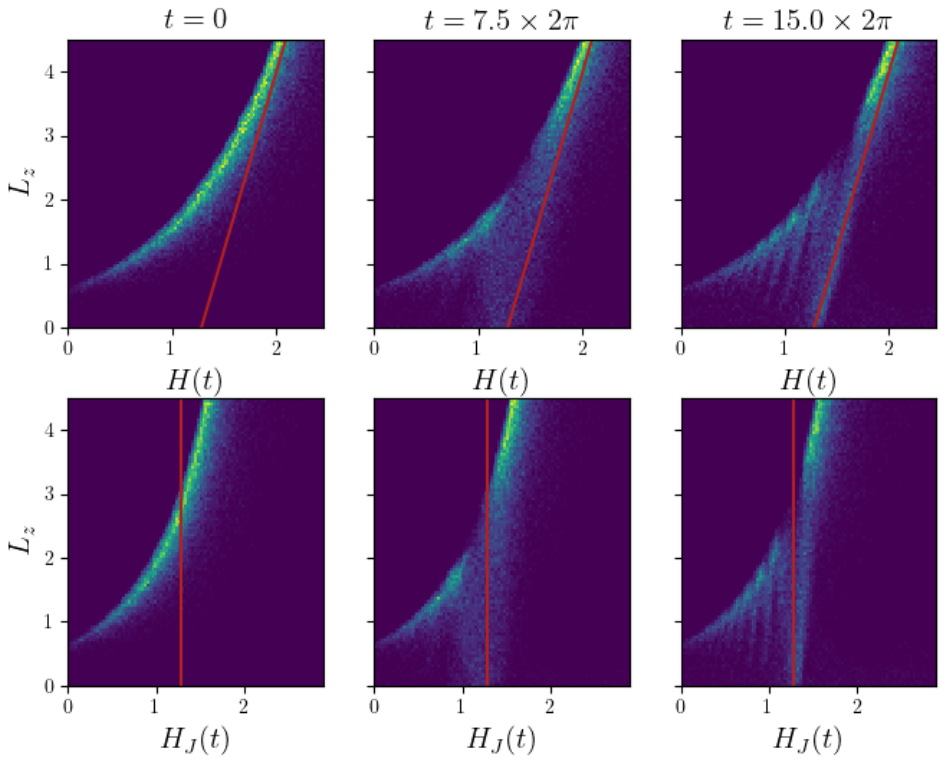


Figure 6.6: The orbits averaged phase space density of the N-body simulation with $\epsilon = 0.05$. The top row shows the phase space as a function of energy, $H(t)$, and angular momentum, L_z, while the bottom shows the phase space as a function of the Jacobi integral, $H_J(t)$, and angular momentum. Particles are driven along lines of constant Jacobi integral by the bar, with particularly large excursions along these lines for trapped orbit. We highlight the line of constant Jacobi integral, H_J associated with the ILR by a red line.

We sample a variety of orbits, from circular to more elliptical, such they all have the same Jacobi integral. We pick the Jacobi integral such that the unperturbed circular orbit's radius is at the location of the resonance. For the unperturbed circular orbit the energy and angular momentum are related via

$$H = \frac{L_z^2}{2R^2} + \Phi_0(R), \tag{6.17}$$

and the radius of such an orbit is given implicitly by

$$\frac{\partial \Phi_0(R)}{\partial R} - \frac{L_z^2}{R^3} = 0. \tag{6.18}$$

With a bar of pattern speed of $\Omega_p = 0.18$ we have $H_J = 0.69$ for the ILR and $H_J = 1.21$ for the CR. When evolving these particles these particles do not feel self-gravity.

The orbital sections are are plotted in Figure 6.5, with the upper hand plot showing those at ILR and the lower hand plot showing sections at CR. To calculate the orbital sections we use $\Omega_p = 0.18$, $\epsilon = 0.1$ and $R_{\rm q} = 1.5$. When evolving the particles we only do so in the bar's potential without the addition of the self consistent response. To make a surface of section, a mark is made every time an orbit passes through its apocentre, with its angular momentum and angle relative to the major axis of the bar, $\phi_{\rm apo}$, begin used to select the location of the mark. The two distinct orbit families can be seen: trapped (blue) and not trapped (red). The existence of these orbits, and the radical change to their initial shape, cannot be captured by linear theory. Hence why it fails to reproduce the torque of the large bars in Figure 6.4.

In linear theory we integrate along unperturbed orbits, in essence assuming that, despite any perturbation, the actions remain approximately unchanged. If we wanted to include non-linear effects we would have to update the orbits in light of their reshaping by the perturbation. In Figure 6.6 we plot the phase space of the N-body simulation with $\epsilon = 0.05$ (the livid blue line in Figure 6.4). The top row shows the phase space as a function of energy, $H(t)$, and angular momentum, L_z, while the bottom row of the shows the phase space as a function of the Jacobi integral, $H_J(t)$, and angular momentum. Each column shows a snapshot at $t = 0$, $7.5 \times 2\pi$, $15.0 \times 2\pi$ from left to right. We can see a drastic change in the DF as a function of time, questioning our linear assumption. In particular we get a striped pattern appearing, which can be seen to occur at fixed Jacobi integral. In Figure 6.6 we plot the line of constant Jacobi integral, H_J associated with the circular orbit that rotates at ILR in the unperturbed case

6.5 Conclusion

In this chapter we have studied the evolution of a disc perturbed by a bar rotating at a fixed pattern speed. In particular we showed that with a weak bar the agreement

between linear response theory and N-body simulations are in strong agreement, indicating the absence of non-linear effects. In particular we highlighted: the effect of disc temperature on the size of the response and how linear theory could easily predict the torque at late times using the response matrix. We then increased the strength of the bar making it more realistic, but this led to a disagreement between our linear response calculation and non-linear N-body simulations. By studying the phase space evolution of the N-body simulations we diagnosed the source of the disagreement: resonant orbits, whose existence was further confirmed by plotting surface of sections for particles near ILR and CR. Although our linear response calculation fails, it is not a terminal illness for analytic approaches. In particular, transforming to slow and fast actions offers a way to cure this ailment (e.g. Tremaine & Weinberg, 1984; Kaasalainen, 1994; Sridhar & Tourna, 1996; Binney, 2016, 2018; Monari et al., 2017; Chiba et al., 2021; Chiba & Schönrich, 2022)

You're gonna to have to leave me now I know,
But I'll see you in the sky above,
In the tall grass and the ones I love,

– B. Dylan, *You're Gonna Make Me Lonesome When You Go*

7 Conclusions

Contents

7.1 What Just Happened? 131
7.2 What is Left to Happen? 134

7.1 What Just Happened?

There are many methods to study the evolution of a galactic disc, but few that do so in a linear, self-consistent manner. To address this problem, this thesis developed a method to study the response of a disc to an external perturbation under such assumptions and used it to study bars and three different processes that are crucial in our understanding of spiral arm formation: density waves (Lin & Shu, 1966; Kalnajs, 1965), swing amplification feedback cycle (Toomre, 1977, 1981; Bertin et al., 1989) and swing amplification (Julian & Toomre, 1966; Binney, 2020).

In chapter 2 we developed a method to solve the linearised CBE inspired by the approach Murali (1999) used to study halos. This is a complex problem as galaxies are inhomogeneous; the trajectories of particles in phase space are complex orbits that do not have a simple form in real space coordinates. To overcome this issue, we transformed into action-angle variables – a coordinate system in

which the momenta (the 'actions') are integrals of motion and hence the generalised coordinates (the 'angles') grow lineally in time. This introduced a new complication – the Laplacian in Poisson's equation is non-local in action-angle coordinates, and hence much harder to solve – which was alleviated by the introduction of potential density pairs (Kalnajs, 1976). With these two building blocks, we recast the linear CBE as an matrix integral equation which could be solved to study the linear response of a disc to an external perturbation. We finished by comparing the density response when using two different potential-density pairs, showing that the density response was basis independent, and discussing the implict softening in the potential density pair approach.

In chapter 3 we compared the effect of swing-arm amplification in a full disc and the shearing sheet (Julian & Toomre, 1966). We first studied the density response to a perturbation using both the linear response method of chapter 2 and the JT equation (Julian & Toomre, 1966; Binney, 2020). The response was qualitatively similar for both methods, despite the difficulties in creating comparable geometries for the two different approaches. We then showed that the level of amplification of the shearing sheet is comparable to the amplification in a full disc, despite the very different assumptions that the two approaches entail. We finally used our linear response calculation to see the response to a point perturbation and the formation of a two-armed spiral structure.

In chapter 4 we turned our attention to LSK theory and the study of density waves. We started by reviewing the LSK dispersion relation (Lin & Shu, 1966; Kalnajs, 1965) and the motion of density waves in the WKB limit (Toomre, 1969). In particular we highlighted the relation between the direction of winding and the direction of travel. We then developed a method to decompose waves in cylindrical geometry into in-going and out-going waves. The main complication was finding a basis for the radial decomposition as the Hankel functions, the cylindrical equivalent of complex exponential in the Cartesian case, cannot easily be projected onto. To solve this problem we constructed a basis based on the Hankel functions in the limit $|kR| \gg 1$. By studying the unwinding of a logarithmic spiral we showed

that this approach could accurately reproduce the input evolution. We finally used this method to study the motion of WKB waves. We confirmed the relationship between the direction of travel and the sign of the wavenumber on the short branch of the LSK relation, however due to the number of assumption behind LSK theory a clear comparison was not possible.

In chapter 5 we confirmed the existence of the swing amplifier feedback cycle (Toomre, 1977, 1981; Bertin et al., 1989) in a scarred disc. We first developed a method to calculate the modes of a disc using the evolution kernel outlined in chapter 2. The modes calculated using our kernel method agreed well with the previously calculated modes of the tapered Mestel disc (Evans & Read, 1998), but in the marginally unstable case our method fails due to numerical convergence issues. By adding a scar to the background DF we confirmed the presence of a swing amplification cycle. To do so we first compared the spatial frequency, ω_0, of a scarred disc calculated via the response matrix to a WKB based approximation, that assumed the inner scar and forbidden radius act as the walls of a density wave cavity, supporting a standing wave between them. The strong agreement indicated the existence of such a cavity. We further confirmed that the inner scar reflects the in-going density wave, protecting it from absorption at the ILR, by decomposing the density response into in-going trailing and out-going leading density waves, agreeing with the predictions of the swing amplifier feedback cycle (Toomre, 1977, 1981; Bertin et al., 1989).

In chapter 6 we studied the evolution of a disc perturbed by a bar rotating at a fixed pattern speed. In order to have a realistic bar model we took the best fit from Sormani et al. (2015). When the bar was weaker than the best fit of Sormani et al. (2015) the linear response method and N-body simulation showed strong agreement, indicating the absence of non-linear effects. In particular we highlighted: the effect of disc temperature on the size of the response and how linear theory could easily predict the torque at late times using the response matrix. We then increased the strength of the bar making it more realistic, however this led to a disagreement between our linear response calculation and non-linear N-body simulations. By

studying the phase space evolution of the N-body simulations we diagnosed the source of the disagreement: resonant orbits, whose existence was further confirmed by plotting surface of sections for particles near ILR and CR. Although our linear response calculation fails, it is not a terminal illness for perturbative approaches. In particular, transforming to slow and fast actions offers a way to cure this ailment (e.g. Tremaine & Weinberg, 1984; Kaasalainen, 1994; Sridhar & Tourna, 1996; Binney, 2016, 2018; Monari et al., 2017; Chiba et al., 2021; Chiba & Schönrich, 2022).

7.2 What is Left to Happen?

Despite the contributions that this thesis has hopefully made, there is still a lot of work to be done. In particular there are two areas which call for continued study: the development perturbative methods to continue the study of specific processes that drive galactic evolution and the application to real data.

A future development, literally taking the method of chapter 2 to new heights, would be to drop the razor-thin assumption and study the vertical motion. In particular the bending modes of the disc (Hunter & Toomre, 1969) and their interplay with other linear phenomena have been used to explain otherwise unexplained phenomena (e.g. Antoja et al., 2018). There are two obstacles that make this a difficult task: first, one would have to use true action-angle variables, rather than apocentre and pericentre coordinates used chapter 2. Although computationally more difficult, this could be included by deployment of the torus machine (McGill & Binney, 1990). Second, and more importantly, the warping of a disc is much larger than its thickness (Oort et al., 1958), making the motion difficult to study with perturbation theory. Although the study of 3D effects is relatively easy using N-body simulations, we hope that this thesis has shown the need for linear tools, allowing what are complex phenomena to be stripped back and studied carefully.

One natural extension of this method would be to include non-linear effects, in particular those of resonant orbits. A standard way of dealing with trapping is to transform in to a new set of action, the so-called ‘slow’ and ‘fast’ actions (e.g. Tremaine & Weinberg, 1984; Kaasalainen, 1994; Sridhar & Tourna, 1996; Binney,

2016, 2018; Monari et al., 2017; Chiba et al., 2021; Chiba & Schönrich, 2022). Including these techniques in the method of chapter 2 would be difficult, but would allow for the inclusion of trapped orbits and hence the study of bars.

Give then complications that are quickly encountered when studying non-linear theory, but the advantages of solving the CBE directly for the DF, another approach would be to develop a method that solves the CBE in its unadulterated form. In Magorrian (2023) periodic cube, a highly idealised self-gravitating system, was studied by numerically solving the CBE using a finite difference method. A similar approach, where a finite difference method is developed to study non-linear behaviour in the razor thin disc, could become a torch illuminating the non-linear phenomena of discs. A non-linear method would be helpful in continuing the analysis of chapter 5: current understanding is there is no single groove mode that creates spiral structure, rather each groove instability cuts a new groove, modifying the spiral instability of the disc (Sellwood & Carlberg, 2014, 2019; Sellwood & Masters, 2022). As these updates are modifications to the background DF, a non-linear method is needed for further investigation.

Ultimately Astrophysics is an experimental science, yet in this thesis we have tried our hardest to stay away from the sobering reality of observations. There is, of course, an important place for the development of techniques and the study of model galaxies, yet the final test has be to comparison to observation. In particular the method outlined in chapter 2 has two advantages over their N-body counterparts: once the kernel is generated, the evolution response can be obtained very quickly, allowing a quick search of parameter space to fit predictions of observation; this method gives the evolution of the DF (via Equation 2.7, rather than, in the case of N-body simulations, the evolution of particles sampled from a DF. This method could be used in conjunction with Gaia observations to develop our understanding of the spirals in our own galaxies, e.g. Sellwood et al. (2019). Due to their inherently non-linear nature, this method would be unsuitable in a study of the effects that a bars has on its host disc. Fortunately, however, this area of study is rapidly advancing (e.g. Chiba et al., 2021; Chiba & Schönrich, 2022).

The spiral arms of disc galaxies are some of the most beautiful structures in the universe. I hope that this thesis has managed to illuminate some of the physical process that drive their development.

Appendices

A

Calculating Basis Functions via Green Function

In this appendix we discuss how to calculate the potential partner of a razor-thin density basis function. By definition, potential density pairs must pairwise solve Poisson's equation (Kalnajs, 1976). For functions that are extended in all 3D dimension (e.g. if their angular dependence is a spherical harmonic) this is a simple statement. However for those density functions that are 2D razor-thin disc embedded in a 3D space this is a complex statement: how do we deal with a delta functions when solving a PDE? For these equations it is equivalent and much easier to say that the density function is related to its partner via a convolution with a Green function, $G(\mathbf{x}, \mathbf{x}')$, i.e.

$$\psi^p(\mathbf{x}) = \int d^3\mathbf{x}' G(\mathbf{x}, \mathbf{x}')\rho^p(\mathbf{x}'), \tag{A.1}$$

where $\mathbf{x}$ is all three spatial coordinates. If we want our basis functions to solve Poisson's equation then the Green function has the form

$$G(\mathbf{x}, \mathbf{x}') = \frac{-G}{|\mathbf{x} - \mathbf{x}'|}, \tag{A.2}$$

which can be modified if softening is to be included (e.g. Binney & Tremaine, 2008; De Rijcke et al., 2019b). Before we use Equation A.1 to calculate the potential

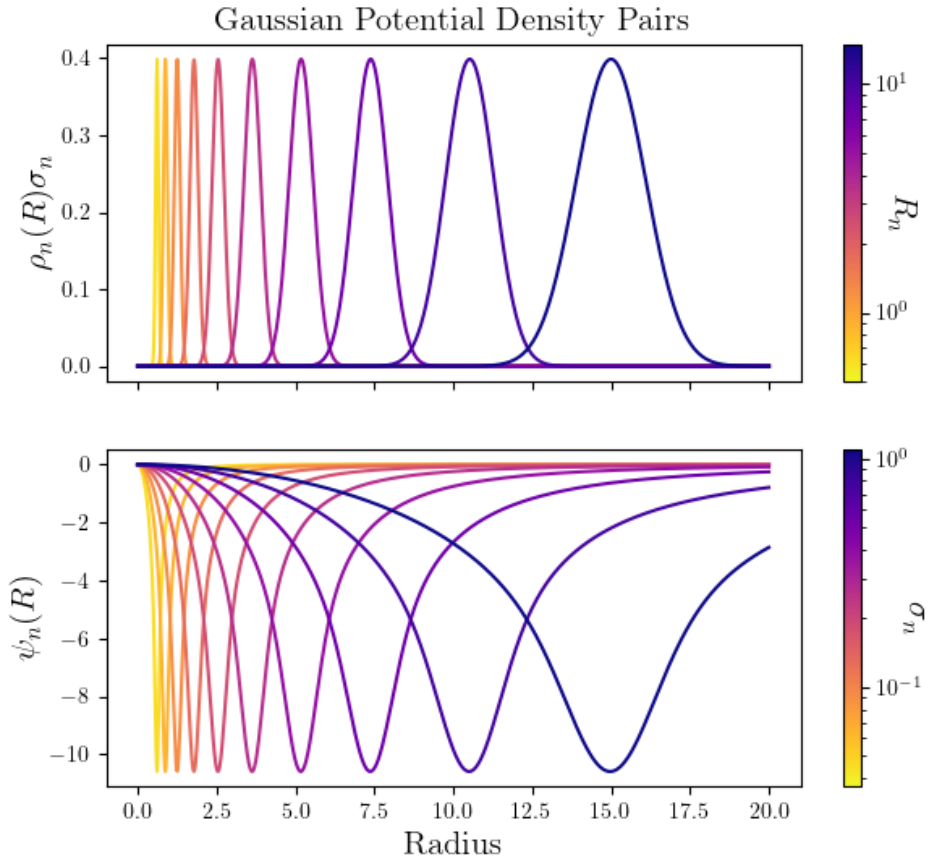


Figure A.1: The logarithmically spaced Gaussian basis functions for $\ell = 2$ with inner radius 0.5 and outer radius 15. The upper (lower) colour bar shows the central radius (width) of each basis function. *Top row*: the density functions multiplied by their width, σ_n. *Bottom row*: the potential functions calculated via the method outlined in this appendix. In order to keep the diagram uncluttered we plot every fifth basis function.

pair, we have two jobs to do: first, we must express the Green function of Poisson's equation in cylindrical coordinates; second, we must develop a judicious numerical scheme that takes due care of the singularity at $\mathbf{x}' = \mathbf{x}$ in Equation A.2.

Fortunately, after consultation §3.11 of Jackson (1999) we can express the Green function of Poisson's equation in cylindrical coordinates as

$$\frac{-1}{|\mathbf{x}' - \mathbf{x}|} = -\frac{2}{\pi} \sum_{m=-\infty}^{\infty} \int_0^{\infty} dk \cos\left[k(z - z')\right] \mathrm{e}^{\mathrm{i}m(\phi - \phi')} I_m(kR_<) K_m(kR_>), \qquad \text{(A.3)}$$

where $R_< = \min(R, R')$ and $R_> = \max(R, R')$ and I_m and K_m are modified Bessel functions of the first and second kind respectively. It is worth noting that this

expression requires the correct boundary conditions to be obtained, namely that the potential is finite at the origin and zero at infinity, the sensible choice for our disc. For a razor thin disc, the density function of the pair will be non-zero only in the plane $z = 0$, therefore we can write its z dependence as $\rho^p(R', \phi', z') = \rho^p(R', \phi')\delta(z')$. Integrating out the ϕ' and z' dependence of Equation A.1 with this density and the Green function given by Equation A.3 gives

$$\mathcal{U}_\ell^n(R) = -4 \int_0^\infty \left\{ \int_0^\infty I_\ell(kR_<) K_\ell(kR_>) dk \right\} \mathcal{D}_\ell^n(R') R' dR' \tag{A.4}$$

where we have assumed that we want the potential at $z = 0$ and that our basis functions are a singular angular harmonic.

In order to simply this expression we define

$$\mathcal{I}_\ell\left(\frac{R}{R'}\right) = \begin{cases} \displaystyle\int_0^\infty I_\ell\left(x\frac{R}{R'}\right) K_\ell\left(x\right) dx, & \text{if } \frac{R}{R'} < 1, \qquad \text{(A.5a)} \\ \displaystyle\int_0^\infty K_\ell\left(x\frac{R}{R'}\right) I_\ell\left(x\right) dx, & \text{if } \frac{R}{R'} > 1. \qquad \text{(A.5b)} \end{cases}$$

For small arguments the integrand of this integral is well defined. However for large values there is a problem. Using the expansion of the modified Bessel functions for $x \sim x\frac{R}{R'} \gg 1$ the integrand of both cases becomes

$$I_\ell\left(x\frac{R}{R'}\right) K_\ell\left(x\right) = K_\ell\left(x\frac{R}{R'}\right) I_\ell(x) \sim \frac{\mathrm{e}^{-x|R/R'-1|}}{2x\sqrt{R/R'}}. \tag{A.6}$$

For R/R' close to unity this integral will converge slowly and, more worryingly, $\mathcal{I}_\ell(1)$ is undefined. Both of these issues can be resolved with careful numerical integration scheme, which we will now outline.

First, we turn our attention to calculating $\mathcal{I}_\ell$ while, for the time being at least, ignoring $R = R'$. In order to ensure convergence we must integrate to $x \gg 1/|R/R' - 1|$. To do this, while not taking too large a step around the lower limit of the integral, we make a non-linear change of variables so that $x = u^2$ and include the Jacobian of the transformation, $dx = 2udu$, in the integrand. We then use the trapezium rule, with constant spacing in u, to calculate $\mathcal{I}_\ell$.

We now must turn to the outer integral and the divergence at $R = R'$. If we define $R'/R = y$ then the potential is

$$\mathcal{U}_\ell^n(R) = -4R \int_0^\infty \mathcal{I}_\ell \left(\frac{1}{y}\right) \mathcal{D}_\ell^n(yR) dy. \tag{A.7}$$

We can tame the divergence that now occurs at $y = 1$ by once again changing variables to $y = (v-1)^3 + 1$, so that the area element in the v coordinate system goes to zero at $v = y = 1$. We can then calculate the final integral by the trapezium rule, with a constant spacing in v and Jacobean $dy = 3(v-1)^2 dv$, once again.

In Figure A.1 we plot the Gaussian basis functions calculated via this method. As expected with the normalisation of the density functions, the magnitude of the potential responses are independent of σ_n. Note that in Figure A.1 we plot every fifth basis function.

B

Details of test-particle and N-body integrator

Contents

B.1 Sampling the DF 143
B.2 N-Body Evolution 145

In order to understand the limitation of our linearity assumptions we compare them to N-body simulations in chapter 6. In this appendix we describe our N-body method.

B.1 Sampling the DF

Our first task is to sample from the DF described in section 2.3, we do this in two parts: initially, we sample the positions of the particles, before sampling their generalised momenta. For each particle we must sample four coordinates, $(r_i, \phi_i, v_{r,i}, v_{\phi,i})$ and assign a weight μ. We sample in polar coordinates due to the angular symmetry of the DF. For velocity sampling we work in (v_r, v_ϕ) coordinates, rather than (E, L) as the former, along with the generalised coordinates (r, ϕ), forms a canonical coordinate system, and therefore does not require the inclusion of a Jacobian factor.

Initially, we sample (r_i, ϕ_i). We start by calculating the cumulative radial distribution function,

$$D(< R) = \int_0^R (2\pi r) dr \int d\mathbf{v} F(r, v_r, v_\phi), \tag{B.1}$$

where the distribution function, F, is given in Equation 2.24. Once this is calculated we can sample r_i using the inverse transform sampling method. Due to the angular symmetry, ϕ_i is trivially sampled, $\phi \sim \mathcal{U}(0, 2\pi)$.

We now turn our attention to the velocity sampling, the radial velocity is easily sampled, $v_r \sim \mathcal{N}(0, \sigma_r^2)$, however the angular velocity is harder to sample, due to the non-trivial tapers.

To sample $v_{\phi,i}$, we use a rejection sampling method. At this point in the sampling algorithm we have three out of four phase space coordinates, namely r_i, ϕ_i and $v_{r,i}$, this reduces the DF to one dimension, $F(v_\phi | r_i, \phi_i, v_{r,i})$. For rejection sampling, we assume that this constrained DF is bounded $v_\phi \in [0, v_{\phi,\max}]$ and $F(v_\phi) \in [0, \mathcal{M}]$. The upper cut off on v_ϕ is chosen as $v_{\phi,max} = v_c + 10\sigma_r$, as the probability of being above this cut-off is less than $10^{-3}\%$. We then pick two uniformly distributed values $v_{\phi,i}$ and y_i, distributed according to $v_\phi \sim \mathcal{U}(0, v_{\phi,max})$ and $y \sim \mathcal{U}(0, \mathcal{M})$. We accept $v_{\phi,i}$ if,

$$\begin{aligned} y_i \leq F(v_{\phi,i} | r_i, \phi_i, v_{r,i}) &\implies v_{\phi,i} \text{ is accepted,} \\ y_i > F(v_{\phi,i} | r_i, \phi_i, v_{r,i}) &\implies v_{\phi,i} \text{ is rejected.} \end{aligned} \tag{B.2}$$

We continue to resample until the criteria in Equation B.2 is met. This algorithm is made more efficient by picking $\mathcal{M}$ as close to the maximum of $F(v_\phi)$ as possible. As both $0 \leq T_{\text{inner}}, T_{\text{outer}} \leq 1$, we choose $\mathcal{M}$ as the maximum value of the untapered Mestel DF (Equation 2.20) given the values already sampled.

Once we have sampled the four components for each particle, we transform into Cartesian coordinates - the coordinate system in which our N-body simulation will work - using the relations,

$$x_i = r_i \cos \phi_i, \tag{B.3a}$$

$$y_i = r_i \sin \phi_i, \tag{B.3b}$$

$$v_{x,i} = v_{r,i}\cos(\phi_i) - v_{\phi,i}\sin(\phi_i), \tag{B.3c}$$

$$v_{y,i} = v_{r,i}\sin(\phi_i) + v_{\phi,i}\cos(\phi_i). \tag{B.3d}$$

To each particle we assign a mass $\mu = M_{\rm disk}/N$, where $M_{\rm disk}$ is the mass of the active part of the disk and N are the number of particles in the simulation.

B.2 N-Body Evolution

Once we have sampled the particles for the N-body simulation, we must generate the potential they source, and which they move in. To do this we split the potential into three parts:

1. An axisymmetric contribution $\Phi_0(r)$ that is sourced by the background Mestel disk and is calculated analytically using $\Phi_0(r) = v_c^2 \log(r/r_0)$.

2. The potential sourced by the fluctuations in the disk, $\psi^f(r,\phi)$.

3. The external perturbation that will 'stir' the disk, $\psi^e(r,\phi)$.

To calculate ψ^f we assign the particles to a square mesh of size $N_{\rm grid} \times N_{\rm grid}$, with uniform grid spacing ΔX, using a cloud-in-cell interpolation scheme. This density is then filtered in order to isolate the Fourier harmonic we are looking for, i.e. if we are interested in the $\ell = 1$ response, we calculate the $\ell = 1$ density perturbation. This filtered density is then turned into a potential using the 'doubling up' method (e.g. Magorrian, 2007; Binney & Tremaine, 2008). A slight complication arises when calculating the $\ell = 0$ response as two parts remain after the filtering step , the $\ell = 0$ perturbation due to the density fluctuations and the potential due to the background disk, i.e. a combination of (i) and (ii). This background will be subject to fluctuations that will drive dynamics, and therefore we want to work with the analytic background of a Mestel disk. To subtract this component out, we calculate the potential caused by the particles when they are initially sampled and subtract this from filtered potential, leaving the potential due the density fluctuation. To this we then add the analytic background when doing the leapfrog integrator.

To generate the external perturbation, we initially calculate the coordinates of the perturbation in the chosen basis, i.e. we calculate $A_p(t)$ using Equation 2.9. These coefficients can then be used to generate a grid representing the external perturbation (and can also be fed into Equation 2.16 to calculate the density response via the linear response method set out in section 2.2).

For test particle simulations, we turn off the self-consistent response, in effect setting the active fraction to zero.

Once we have the total acceleration, from the background potential, the self-consistent response and the external perturbation, we use a second order leapfrog integrator to evolve the particles in time.

In order to compare the predictions made by Equation 2.16 and N-body we must calculate the expansion coefficients of the density fluctuations, $B_p(t)$, from the N-body simulation. To do this we run the code twice with the same initial conditions: initially, the 'background' run, in which the external perturbation is turned off and secondly, the 'foreground' run, including the external perturbation. For each foreground and background simulation we construct the foreground and background Klimontovich distribution function, f_f and f_b respectively,

$$f_{\mathrm{f/b}} = \sum_{n=1}^{N} \mu \delta^2(\mathbf{x} - \mathbf{x}_{i,\mathrm{f/b}}) \delta^2(\mathbf{v} - \mathbf{v}_{i,\mathrm{f/b}}), \tag{B.4}$$

where $\mu = M_{\mathrm{disk}}/N$ is the weight of each particle and N is the total number of particles, ensuring that the mass of the disk is the same in the N-body and linear response calculation. The density fluctuation can then be calculated from the marginalised Klimontovich DFs,

$$\delta\rho(\mathbf{x}) = \int d^2\mathbf{v} \left[f_{\mathrm{f}}(\mathbf{x}, \mathbf{v}) - f_{\mathrm{b}}(\mathbf{x}, \mathbf{v}) \right]. \tag{B.5}$$

Combining the density fluctuation along with the bi-orthogonality relation (2.9), we can calculate $B_p(t)$,

$$B_p(t) = -\mathcal{E}^{-1}_{pq} \int d^2\mathbf{x} \left[\psi^q(\mathbf{x}) \right]^* \delta\rho(\mathbf{x}). \tag{B.6}$$

For the N-body simulation we set $R_0 = v_c = G = 1$, as they are scale parameters. For all results presented here we use $R_{max} = 20$, with a uniform time step $\Delta t = 10^{-2}$, that we check for convergence. For our background DF we set $\nu_t = 4$, $\mu_t = 5$, $R_i = 1$, $R_o = 11.4$ and $\xi = 1/2$ inline with chapter 2. For the potential solver we take a maximum radius of $26r_0$ with a box size of $N_{\rm grid} = 120$. When taking the Fourier modes of the density we use 120 radial rings, each with 720 steps. The softening length is taken as $0.18r_0$. These values are taken from Fouvry et al. (2015b) as the same potential solver code is used.

C

Response Matrix Via Direct Integration

[The approach taken to calculate the response matrix follows closely the work of Fouvry et al. (2015b)]

In this section we outline the direct integration method used to calculate the response matrix in Fouvry et al. (2015b), which we use to compare against our own kernel method subsection 5.2.1.

In order to make use of already calculated component parts, we follow the same approach as that taken for calculating the evolution kernel. To do this we convert the integral over action space, (J_1, J_2) to one over apocentre and pericentre coordinates, (R_+, R_-), in Equation 5.4. This simplifies the integration domain to a triangle, as by definition $R_- \leq R_+$. We first convert into energy-angular momentum coordinates (E, L). To do this the we note that $\Omega_1 d^2\mathbf{J} = dEdL$, reducing Equation 5.4 to

$$\hat{\mathcal{M}}_{pq}(\omega) = (2\pi)^2 \delta^{lq}_{lp} \sum_{m_1} \int dEdL \frac{1}{\Omega_1} \frac{\mathbf{m} \cdot \partial F_0/\partial \mathbf{J}}{\omega - \mathbf{m} \cdot \mathbf{\Omega}} \mathcal{W}^{m_1}_{\ell^p \ell^p n^p}(\mathbf{J}) \mathcal{W}^{m_1}_{\ell^p \ell^p n^q}(\mathbf{J}), \tag{C.1}$$

where we have used the Kronecker delta in Equation 2.38 to note that $m_2 = \ell_q = \ell_p$ for non-zero elements of the response matrix. We drop the complex conjugate on $\mathcal{W}$ as it is a real function. The next step is to transform from $(E, L) \to (R_+, R_-)$ which will allow our previously calculated expressions for $\mathcal{W}_{\mathbf{m}}$ to be used. We label the resulting integral as

$$\hat{\mathcal{M}}_{pq}(\omega) = \delta^{lq}_{lp} \sum_{m_1} \int dR_+ dR_- \frac{g^{\ell^p n^p n^q}_{m_1}(R_+, R_-)}{h^{\omega}_{m_1 \ell^p}(R_+, R_-)}. \tag{C.2}$$

We define the numerator of the integrand, $g_{m_1}^{\ell^p n^p n^q}(R_+, R_-)$, as

$$g_{m_1}^{\ell^p n^p n^q}(R_+, R_-) = (2\pi)^2 \left| \frac{\partial(E, L)}{\partial(R_+, R_-)} \right| \frac{1}{\Omega_1} \mathbf{m} \cdot \frac{\partial F_0}{\partial \mathbf{J}} \times \mathcal{W}_{\ell^p \ell^p n^p}^{m_1}(R_+, R_-) \mathcal{W}_{\ell^p \ell^p n^q}^{m_1}(R_+, R_-), \quad \text{(C.3)}$$

where the Jacobian and $1/\Omega_1$ factor come from the transformation from $(J_1, J_2) \rightarrow (E, L)$ and $(E, L) \rightarrow (R_+, R_-)$ respectively. In order to efficiently calculate the response matrix, we precompute the portion of the expression that does not depend on the basis function's radial index, n^p or n^q. We define the numerator of Equation C.2 as

$$h_{m_1 \ell^p}^{\omega}(R_+, R_-) = \omega - \mathbf{m} \cdot \mathbf{\Omega}. \quad \text{(C.4)}$$

In order to find the modes of our disc we will need to evaluate $\hat{\mathcal{M}}(\omega)$ for different ω, therefore to expedite our calculation we precompute $\mathbf{m} \cdot \mathbf{\Omega}$.

We now must evaluate the integral over (R_+, R_-), to do this we split the integration domain into square patches, such that the i$^{\text{th}}$ integration domain is $R_+ \in \left[R_+^i - \Delta R_{grid}, R_+^i + \Delta R_{grid}\right]$ and $R_- \in \left[R_-^i - \Delta R_{grid}, R_-^i + \Delta R_{grid}\right]$ where ΔR_{grid} is the grid spacing on which we calculated $\mathcal{W}$. For brevity in the expression that follow, we simplify ΔR_{grid} to just Δ. In each of the integration patches we can approximate h and g via a first order Taylor expansion such that

$$g(R_+^i + x_+, R_-^i + x_-) \approx a_g^i + b_g^i x_- + c_g^i x_+, \quad \text{(C.5)}$$

where $g(R_+^i, R_-^i) = a_g^i$ and b_g^i and c_g^i are respectively the values of the derivatives of g with respect to R_+ and R_- evaluated at (R_+^i, R_-^i). With an equivalent expressions for h, with the subscript g replaced by h. We calculate the derivatives using second order finite differences. This allow us to write the integral over the i$^{\text{th}}$ patch as

$$\int\int_i dR_+ dR_- \frac{g(R_+, R_-)}{h(R_+, R_-)} \approx \int_{-\Delta}^{\Delta}\int_{-\Delta}^{\Delta} dx_+ dx_- \frac{a_g^i + b_g^i x_- + c_g^i x_+}{a_h^i + b_h^i x_- + c_h^i x_+ + \mathrm{i}\eta} = \aleph(a_g^i, b_g^i, c_g^i, a_h^i, b_h^i, c_h^i, \eta), \quad \text{(C.6)}$$

where to force $a_h^i \in \mathbb{R}$ we have not included the complex part of $\omega = \omega_0 + i\eta$ in a_h^i. By definition the rest of the expansion coefficents are real functions. Making use of this definition, we can approximate the response matrix as

$$\hat{\mathcal{M}}_{pq}(\omega) = \delta_{\ell p}^{\ell q} \sum_{m_1} \sum_i \aleph(a_g^i, b_g^i, c_g^i, a_h^i, b_h^i, c_h^i, \eta). \tag{C.7}$$

We now must turn out attention to the calculation of $\aleph$. However there are two issues to note: first, how to handle the integration patchwork, especially around the edge of the tapestry; second, the sum over m_1 must be performed judiciously to ensure convergence. Both of these complications are discussed further once we calculate $\aleph$.

To calculate $\aleph$ we first introduce the non-dimensional integral

$$\aleph_D(b, c, e, f, \eta) = \int_{\frac{1}{2}}^{\frac{1}{2}} \int_{\frac{1}{2}}^{\frac{1}{2}} dxdy \frac{1 + bx + cy}{1 + ex + fy + i\eta}. \tag{C.8}$$

If we define new integration variables $x = x_-/(2\Delta)$ and $y = x_+/(2\Delta)$ we can express

$$\aleph(a_g^i, b_g^i, c_g^i, a_h^i, b_h^i, c_h^i, \eta) = \frac{a_g^i}{a_h^i} (2\Delta)^2 \aleph_D \left[\frac{2b_g^i\Delta}{a_g^i}, \frac{2c_g^i\Delta}{a_g^i}, \frac{2b_h^i\Delta}{a_g^i}, \frac{2c_h^i\Delta}{a_h^i}, \frac{\eta}{a_h^i} \right]. \tag{C.9}$$

To evaluate $\aleph_D$ if we can pick a function, $G(x, y)$, such that

$$\frac{\partial G(x, y)}{\partial x \partial y} = \frac{1 + bx + cy}{1 + ex + fy + i\eta}, \tag{C.10}$$

then $\aleph_D$ can be immediately calculated as

$$\aleph_D = G\left[\frac{1}{2}, \frac{1}{2}\right] - G\left[\frac{1}{2}, -\frac{1}{2}\right] - G\left[-\frac{1}{2}, \frac{1}{2}\right] + G\left[-\frac{1}{2}, -\frac{1}{2}\right]. \tag{C.11}$$

Our unique choice for $G(x, y)$ is,

$$\begin{aligned} G[x, y] =& \frac{1}{4e^2 f^2} \log\left[e^2x^2 + 2e(fxy + x)f^2y^2 + 2fy + \eta^2 + 1\right] \\ & \times \left\{bf\left(e^2x^2 - (fy + \mathrm{i}\eta + 1)^2\right) + 2ef(ex + \mathrm{i}\eta + 1) - ce(ex\mathrm{i}\eta + 1)^2\right\} \\ & + \frac{\mathrm{i}}{2e^2f^2}\left\{\frac{\pi}{2} - \arctan\left[\frac{ex + fy + 1}{\eta}\right]\right\} \\ & \times \left\{bf\left(e^2x^2 - (fy + \mathrm{i}\eta + 1)^2\right) + 2ef(ex + \mathrm{i}\eta + 1) - ce(ex\mathrm{i}\eta + 1)^2\right\} \\ & + \frac{y}{4e^2f}\Bigg\{f(-4e + b(2ex + fy + 2\mathrm{i}\eta + 2)) \\ & + ce(2ex - fy + 2\mathrm{i}\eta + 2) + 2ef(cy + 2)\log[ex + fy + \mathrm{i}\eta + 1]\Bigg\}. \end{aligned} \tag{C.12}$$

As $e, f, \eta \in \mathbb{R}$ and $\eta \neq 0$ we need not worry about the branch cuts for the complex log and arctan functions in Equation C.12.

Now that we have calculated $\aleph$, we must sum over all the patches of the tapestry to obtain our expression for $\hat{\mathcal{M}}$. For the interior of the tapestry this is a straight forward exercise, however the edge involves integrating over orbits for which pericentre is at the origin. We choose to ignore these orbits. By testing for stability in the spacing of the uniform grid, we can show that the error this approximation introduces can be arbitrarily reduced by decreasing the cell size of the grid.

When calculating the response matrix we use the Kalnajs basis function to calculate the response matrix, which we limit to the first 40 terms. The basis functions are expanded on a uniform grid of 250 points, giving a spacing $\Delta = 0.06$ as the scale length of the Kalnajs basis functions are set at $R_{\mathrm{Ka}} = 15$. We sum over $-5 \leq m_1 \leq 5$. These numbers ensure the convergence of the response matrix to three significant figures. As we will ultimately want to find the roots of $|\det\left[\hat{\mathcal{M}} - I\right]|$, it is imperative to ensure convergence in this delicate calculation. In subsection 5.2.3 we compare different calculation of the response matrix against those given in Evans & Read (1998).

Bibliography

Adibekyan V. Z., et al., 2013, A&A, 554, A44

Al Kazwini H., et al., 2022, A&A, 658, A50

Antoja T., et al., 2018, Nature, 561, 360

Arnold V. I., 1987, Mathematical methods of classical mechanics. Springer, New York

Athanassoula E., 2012, MNRAS, 426, L46

Athanassoula E., Machado R. E. G., Rodionov S. A., 2013, MNRAS, 429, 1949

Banik U., van den Bosch F. C., 2021, ApJ, 912, 43

Benjamin R. A., et al., 2005, ApJ, 630, L149

Bensby T., Zenn A. R., Oey M. S., Feltzing S., 2007, ApJ, 663, L13

Bertin G., 2014, Global Spiral Modes, 2 edn. Cambridge University Press, p. 229–246, doi:10.1017/CBO9780511731990.019

Bertin G., Lin C. C., Lowe S. A., Thurstans R. P., 1989, ApJ, 338, 104

Binney J., 2016, MNRAS, 462, 2792

Binney J., 2018, MNRAS, 474, 2706

Binney J., 2020, MNRAS, 496, 767

Binney J., Tremaine S., 2008, Galactic Dynamics. Princeton university press

Binney J., Vasiliev E., 2022, MNRAS

Binney J., Gerhard O. E., Stark A. A., Bally J., Uchida K. I., 1991, MNRAS, 252, 210

Binney J., Gerhard O., Spergel D., 1997, MNRAS, 288, 365

Blitz L., Spergel D. N., 1991, ApJ, 379, 631

Chavanis P., 2012, Physica A: Statistical Mechanics and its Applications, 391, 3680

Chiba R., Schönrich R., 2022, MNRAS, 513, 768

Chiba R., Friske J., Schönrich R., 2021, MNRAS, 500, 4710

Colombi S., Sousbie T., Peirani S., Plum G., Suto Y., 2015, MNRAS, 450, 3724

Colpi M., Pallavicini A., 1998, ApJ, 502, 150

De Rijcke S., Fouvry J. B., Pichon C., 2019a, MNRAS, 484, 3198

De Rijcke S., Fouvry J. B., Dehnen W., 2019b, MNRAS, 485, 150

Dehnen W., 2001, MNRAS, 324, 273

Dehnen W., McLaughlin D. E., Sachania J., 2006, MNRAS, 369, 1688

Dobbs S., Baba J., 2014, Publications of the Astronomical Society of Australia, 31, 35

Dobbs C. L., Theis C., Pringle J. E., Bate M. R., 2010, MNRAS, 403, 625

Eilers A. C., Hogg D. W., Rix H. W., Frankel N., Hunt J. A. S., Fouvry J. B., Buck T., 2020, ApJ, 900, 186

Evans N. W., Read J. C. A., 1998, MNRAS, 300, 106

Fouvry J. B., Pichon C., Prunet S., 2015a, MNRAS, 449, 1967

Fouvry J. B., Pichon C., Magorrian J., Chavanis P. H., 2015b, A&A, 584, A129

Fridman A. M., Polyachenko V. L., 2012, Physics of Gravitating Systems I: Equilibrium and Stability. Springer, New York

Fuhrmann K., 1998, A&A, 338, 161

Gaia Collaboration 2018, A&A, 616

Gilmore G., Reid N., 1983, MNRAS, 202, 1025

Hamilton C., 2023, arXiv preprint arXiv:2302.06602

Heyvaerts J., 2010, MNRAS, 407, 355

Hohl F., 1971, ApJ, 168, 343

Holmberg E., 1941, ApJ, 94, 385

Hunter C., Toomre A., 1969, ApJ, 155, 747

Jackson J. D., 1999, Classical electrodynamics, 3 edn. Wiley

Jalali M. A., Hunter C., 2005, ApJ, 630, 804

Jeans J. H., 1915, MNRAS, 76, 70

Julian W. H., Toomre A., 1966, ApJ, 146, 810

Jurić M., et al., 2008, ApJ, 673, 864

Kaasalainen M., 1994, MNRAS, 268, 1041

Kalnajs A. J., 1965, Ph. D. Thesis

Kalnajs A. J., 1976, ApJ, 205, 745

Kalnajs A. J., 1977, ApJ, 212, 637

Kent S. M., 1987, The Astronomical Journal, 93, 816

Kh S. R., Bailer-Jones C. A. L., Hogg D. W., Schultheis M., 2018, A&A, 618, A168

Khoperskov A. V., Just A., Korchagin V. I., Jalali M. A., 2007, A&A, 473, 31

King I. R., 1966, The Astronomical Journal, 71, 64

Lau J. Y., Binney J., 2021, MNRAS, 507, 2241

Leeuwin F., Combes F., Binney J., 1993, MNRAS, 262, 1013

Levine E. S., Blitz L., Heiles C., 2006, Science, 312, 1773

Lichtenberg A., Lieberman M., 2013, Regular and chaotic dynamics. Springer, New York

Lin C. C., Shu F. H., 1966, Proceedings of the National Academy of Sciences of the United States of America, 55, 229

Lintott C., et al., 2011, MNRAS, 410, 166

Lynden-Bell D., 1967, Monthly Notices of the Royal Astronomical Society, Vol. 136, p. 101, 136, 101

Lynden-Bell D., Kalnajs A. J., 1972, MNRAS, 157, 1

Magorrian J., 2007, MNRAS, 381, 1663

Magorrian J., 2021, MNRAS, 507, 4840

Magorrian J., 2023, to be submitted

Mangeney A., Califano F., Cavazzoni C., Travnicek P., 2002, Journal of Computational Physics, 179, 495

Mark J. W. K., 1974, ApJ, 193, 539

Mark J. W. K., 1977, ApJ, 212, 645

Mayer L., 2022, Journal of Physics G: Nuclear and Particle Physics, 49, 063001

McGill C., Binney J., 1990, MNRAS, 244, 634

Mestel L., 1963, MNRAS, 126, 553

Monari G., Famaey B., Siebert A., 2016, MNRAS, 457, 2569

Monari G., Famaey B., Fouvry J. B., Binney J., 2017, MNRAS, 471, 4314

Morgan W. W., Whitford A. E., Code A. D., 1953, ApJ, 118, 318

Murali C., 1999, ApJ, 519, 580

Nakada Y., Degucji S., Hashimoto O., Izumiura H., Onaka T., Sekiguchi K., Yamamura I., 1991, Nature, 353, 140

Navarro J. F., Frenk C. S., White S. D. M., 1996, in Symposium-international astronomical union. pp 255–258

Nietzsche F., 2005, Thus spoke Zarathustra: A book for everyone and nobody. Oxford University Press, Oxford

Oort J. H., Kerr F. J., Westerhout G., 1958, MNRAS, 118, 379

Pasha I. I., 2004a, arXiv e-prints

Pasha I. I., 2004b, arXiv

Pichon C., Cannon R. C., 1997, MNRAS, 291, 616

Polyachenko E. V., Just A., 2015, MNRAS, 446, 1203

Posti L., Helmi A., 2019, A&A, 621, A56

Pouliasis E., Di Matteo P., Haywood M., 2017, A&A, 598, A66

Quillen A. C., Nolting E., Minchev I., De Silva G., Chiappini C., 2018, MNRAS, 475, 4450

Reid M. J., Brunthaler A., 2004, APJ, 616, 872

Reid M. J., et al., 2014, ApJ, 783, 130

Reid M. J., et al., 2019, ApJ, 885, 131

Rein H., Tremaine S., 2011, MNRAS, 415, 3168

Rozier S., Famaey B., Siebert A., Monari G., Pichon C., Ibata R., 2022, ApJ, 933, 113

Samland M., Gerhard O. E., 2003, A&A, 399, 961

Schwarz M. P., 1981, ApJ, 247, 77

Sellwood J. A., 2006, ApJ, 637, 567

Sellwood J. A., 2010, MNRAS, 409, 145

Sellwood J. A., 2012, ApJ, 751, 44

Sellwood J. A., Carlberg R. G., 1984, ApJ, 282, 61

Sellwood J. A., Carlberg R. G., 2014, ApJ, 785, 137

Sellwood J., Carlberg R. G., 2019, MNRAS, 489, 116

Sellwood J. A., Kahn F. D., 1991, MNRAS, 250, 278

Sellwood J. A., Masters K. L., 2022, Annual Review of Astronomy and Astrophysics, 60, 73

Sellwood J. A., Trick W. H., Carlberg R. G., Coronado J., Rix H. W., 2019, MNRAS, 484, 3154

Siegel M. H., Majewski S. R., Reid I. N., Thompson I. B., 2002, ApJ, 578, 151

Sormani M., Binney J., Magorrian J., 2015, MNRAS, 454, 1818

Springel V., 2005, MNRAS, 364, 1105

Sridhar S., 2019, ApJ, 884, 3

Sridhar S., Tourna J., 1996, MNRAS, 279, 1263

The GRAVITY Collaboration 2019, A&A, 625, L10

Toomre A., 1964, ApJ, 139, 1217

Toomre A., 1969, ApJ, 158, 899

Toomre A., 1977, A&ARv, 15, 437

Toomre A., 1981, in Structure and evolution of normal Galaxies. pp 111–136

Toomre A., Toomre J., 1972, ApJ, 178, 623

Tremaine S., Weinberg M. D., 1984, MNRAS, 209, 729

Van de Hulst H. C., Muller C. A., Oort J. H., 1954, Bulletin of the Astronomical Institutes of the Netherlands, 12, 117

Varghese A., Ibata R., Lewis G. F., 2011, MNRAS, 417, 198

Vasiliev E., 2019, MNRAS, 482, 1525

Vauterin P., Dejonghe H., 1996, A&A, 313, 465

Wang L., Spurzem R., Aarseth S., Nitadori K., Berczik P., Kouwenhoven M. B. N., Naab T., 2015, MNRAS, 450, 4070

Weiland J. L., et al., 1994, ApJ, 425, L81

Weinberg M., 1989, MNRAS, 239, 549

Weinberg M. D., 1991, ApJ, 368, 66

Weinberg M. D., 1992, ApJ, 384, 81

Whitham G. B., 1974, Linear and nonlinear waves. Bantam, New York

Winnicott D. W., 1991, Playing and Reality, 2 edn. Routledge, Abingdon

Yoshii Y., 1982, Publications of the Astronomical Society of Japan, 34, 365

de Vaucouleurs G., 1964, in Symposium-International Astronomical Union. pp 195–199