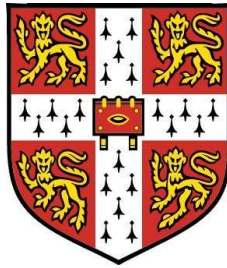


Dynamical instabilities in disc-planet interactions



Min-Kai Lin
St. John's College
University of Cambridge

This dissertation is submitted for the degree of

Doctor of Philosophy

September 2011

Dedicated to my parents, Yen-Chu Lu and Ing-Young Lin.

Acknowledgements

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated.

The contents of Chapter 2, 3 and 4 have respectively been published as Lin & Papaloizou (2010, 2011a,b):

- *Type III migration in a low viscosity disc*, Lin, M.-K., Papaloizou, J.C.B, 2010, MNRAS, 405, 1473
- *The effect of self-gravity on vortex instabilities in disc-planet interactions*, Lin, M.-K., Papaloizou, J.C.B, 2011, MNRAS, 415, 1426
- *Edge modes in self-gravitating disc-planet interactions*, Lin, M.-K., Papaloizou, J.C.B, 2011, MNRAS, 415, 1445

Specifically, the stabilisation of vortex modes (§3.2.4) was proved by J. Papaloizou. The analytical interpretation of edge modes (§4.3.6, §4.4.1—4.4.3) was given by J. Papaloizou. All semi-analytical modelling, linear calculations, hydrodynamic simulations and their results analysis have been performed by myself.

I thank my supervisor Professor John Papaloizou for support, guidance and inspiration. I thank C. Baruteau for his version of the `FARGO` code and the ASIAA CFD-MHD group in Taiwan for providing a copy of their `ANTARES` code for reference. I thank members of the AFD group in DAMTP for three fruitful years. I thank the Isaac Newton Trust for a studentship and St John's College, Cambridge for a Benefactor's Scholarship. This work was also supported by an Overseas Research award.

Abstract

Protoplanetary discs can be dynamically unstable due to structure induced by an embedded giant planet. In this thesis, I discuss the stability of such systems and explore the consequence of instability on planetary migration.

I present semi-analytical models to understand the formation of the unstable structure induced by a Saturn mass planet, which leads to vortex formation. I then investigate the effect of such vortices on the migration of a Saturn-mass planet using hydrodynamic simulations. I explain the resulting non-monotonic behaviour in the framework of type III planetary migration.

I then examine the role of disc self-gravity on the vortex instabilities. It can be shown that self-gravity has a stabilising effect. Linear numerical calculations confirms this. When applied to disc-planet systems, modes with small azimuthal wavelengths are preferred with increasing disc self-gravity. This is in agreement the observation that more vortices develop in simulations with increasing disc mass. Vortices in more massive discs also resist merging. I show that this is because inclusion of self-gravity sets a minimal vortex separation preventing their coalescence, which would readily occur without self-gravity.

I show that in sufficiently massive discs vortex modes are suppressed. Instead, global spiral instabilities develop. They are interpreted as disturbances associated with the planet-induced structure, which interacts with the wider disc leading to instability. I carry out linear calculations to confirm this physical picture. Results from nonlinear hydrodynamic simulations are also in agreement with linear theory. I give examples of the effect of these global modes on planetary migration, which can be outwards, contrasting to standard inwards migration in more typical disc models.

I also present the first three-dimensional computer simulations examining planetary gap stability. I confirm that the results discussed above, obtained from two-dimensional disc approximations, persist in three-dimensional discs.

Contents

1	Introduction	1
1.1	Disc-planet interactions	2
1.2	Instabilities in structured astrophysical discs	4
1.3	Previous studies and relation to present work	5
1.4	Modelling disc-planet systems	6
2	Vortex-induced type III migration	13
2.1	Disc-planet setup	13
2.2	Ring structures in disc-planet interactions	15
2.3	A semi-analytic model for co-orbital vortensity generation by shocks	19
2.4	Dynamical stability of vortensity rings	29
2.5	Simulations of fast migration driven by vortex-planet interaction	46
2.6	The inviscid disc and parameter studies	56
2.7	Additional simulations	68
2.8	Summary and discussion	69
3	Vortex modes in self-gravitating discs	73
3.1	Disc-planet model	74
3.2	Linear modes in discs with structured surface density	75
3.3	Linear calculations	83
3.4	Nonlinear hydrodynamic simulations	96
3.5	Simulations of co-orbital Kida vortices	111
3.6	Implications on vortex-induced migration	119
3.7	Summary and discussion	121
4	Edge modes in self-gravitating discs	125
4.1	Disc model	126

CONTENTS

4.2	Properties of edge modes	127
4.3	Analytical discussion of edge modes	133
4.4	Description of the edge mode instability	141
4.5	Linear calculations	149
4.6	Nonlinear hydrodynamic simulations	165
4.7	Applications to disc-planet interactions	179
4.8	Summary and discussion	184
5	Planetary migration in the presence of edge modes	187
5.1	Simulation setup	187
5.2	Disc models and their stability	189
5.3	Migration and disc structure	191
5.4	A fiducial case in detail	195
5.5	Summary and discussion	197
6	Three-dimensional simulations	201
6.1	Three-dimensional models	201
6.2	Adapting ZEUS-MP	203
6.3	Gap stability in 3D	206
6.4	Summary and discussion	216
7	Conclusions and future work	219
7.1	Main results	219
7.2	Directions for future work	221
	Appendices	223
A	Upper limit on the horseshoe width	225
B	Linear model for flow perturbed by a planet	229
C	Shocks induced by a planet and the vorticity jump across it	233
D	Approximate equations in linear calculations	239
E	Artificial vortices in an accretion disc	241
F	Energy densities in linear theory of edge modes	243

Chapter 1

Introduction

The interaction between planets and their environments —protoplanetary discs—can potentially account for the structure of the Solar system as well as extra-Solar planetary systems. The number of detected exoplanets currently stand at 603 (September 2011). They are found with a variety of orbital properties. Among these are the ‘hot Jupiters’ — Jovian mass planets orbiting close to their host star (Marcy et al., 2005; Udry & Santos, 2007). The first exoplanet discovered around a Solar-type star, 51 Pegasi B, has a period of only 4 days (Mayor & Queloz, 1995). It is difficult to explain their formation in situ with either the core accretion or gravitational instability theories of giant planet formation (see, e.g. Lissauer & Stevenson, 2007; Durisen et al., 2007; D’Angelo et al., 2010, for recent reviews).

One possibility is that hot Jupiters formed further out, where conditions are more favourable, then through gravitational interaction with the gaseous protoplanetary disc migrated inwards (Lin & Papaloizou, 1986). Understanding disc-planet interactions for a range of disc and planet properties is desirable, because its application is not restricted to planet formation. Analogous interactions occur for stars in black hole accretion discs, which has recently received attention and results from planetary migration applied (Baruteau et al., 2011a; Kocsis et al., 2011; McKernan et al., 2011).

Embedded objects can affect stability properties of the disc, and this can back-react on the migration of said object. Thus, a complete theory of planetary migration should also account for possible planet-induced instabilities. The studies presented in this thesis consider instabilities associated with radially structured protoplanetary discs due to a planet and their effects on migration.

1. INTRODUCTION

1.1 Disc-planet interactions

Tidal interaction between the planet and disc fluid leads to angular momentum exchange between them, and a change in the planet's orbital radius from the star. This is planetary migration. It is classically distinguished between type I, type II and more recently, type III (see, e.g. Papaloizou et al., 2007; Masset, 2008; Paardekooper & Nelson, 2009, for recent reviews).

1.1.1 Type I migration

The response of a protoplanetary disc to an embedded low-mass planet can be treated with linear perturbation analysis (e.g. Goldreich & Tremaine, 1979, 1980; Lin & Papaloizou, 1993; Artymowicz, 1993; Ward, 1997). Disc-planet torques are exerted at Lindblad and co-rotation resonances. In standard disc models, e.g. the minimum Solar mass nebula (Weidenschilling, 1977; Hayashi, 1981), the corotation torque is negligible and the total Lindblad torque is negative, and yield inward migration timescales too small compared to disc lifetimes for the survival of protoplanetary cores. Thus, type I migration remains an active area of study.

Recent developments in type I migration include three-dimensional linear calculations (Tanaka et al., 2002), non-isothermal effects (Paardekooper & Mellema, 2006, 2008; Paardekooper et al., 2010b, 2011), disc self-gravity (Pierens & Huré, 2005; Baruteau & Masset, 2008) and interaction with magnetic turbulence (Nelson & Papaloizou, 2004; Baruteau et al., 2011b; Uribe et al., 2011). However, low mass planets do not perturb the disc significantly to render it unstable, so type I migration will not be the most relevant mechanism for disc-planet properties considered later.

1.1.2 Type II migration

Massive planets can clear an annular gap in the disc. This requires the disc to be perturbed nonlinearly, leading to shocks, which amounts to requiring the planet's Hill radius to exceed the disc thickness. The planet also needs to exert a torque on the disc, which tends to push material away from its orbital radius, that is sufficiently large to overcome viscous torques trying to close the gap. Gap-opening is discussed in Lin & Papaloizou (1993), Bryden et al. (1999), and the two criteria above have been combined by Crida et al. (2006).

In type II migration, the planet remains in the gap evolving inwards with disc accretion. Migration therefore proceeds on disc viscous timescales (Lin & Papaloizou, 1986).

The planetary masses to be considered are gap-opening, but standard type II migration is still not relevant because for the conditions presented later, migration is strongly influenced by gap instabilities operating on much shorter, dynamical timescales.

1.1.3 Type III migration

Type III migration is relatively new. It was first described by Masset & Papaloizou (2003) and further discussed in Artymowicz (2004a) and Papaloizou (2005). Recently, a series of detailed numerical studies were performed by Pepliński et al. (2008a,b,c). This is a rapid form of migration with timescales less than 100's of orbits, compared to at least a few thousand for type I and type II. Type III migration is most applicable to Saturn-mass planets opening partial gaps in massive discs.

Type III migration is driven by flow of disc material across the planetary gap, due to its migration in the first place. It is self-sustaining and can be directed inwards or outwards. However, it requires an initial kick. A simplified description of this mechanism, following Papaloizou et al. (2007), is outlined below.

The flow topology near a (partial) gap is shown in Fig. 1.1. If the planet circulates at fixed orbital radius $r = r_p$ then the flow is divided into the horseshoe region (red) where material trapped in libration, and the circulating region where fluid completes circular orbits (blue). If the planet migrates with rate $\dot{r}_p < 0$ then material can cross from the inner disc (blue) to the outer disc by executing one horseshoe turn behind the planet. Fluid elements change orbital radius $r_p - x_s \rightarrow r_p + x_s$, gaining angular momentum and exerts a negative torque on the planet:

$$\Gamma_3 = 2\pi r_p^2 \dot{r}_p \Sigma_e \Omega(r_p) x_s, \quad (1.1)$$

where Σ_e is the density at the upstream separatrix, $\Omega(r)$ is the disc angular speed and x_s is half the horseshoe width (see Fig. 1.1). The migrating mass includes the planet M_p , fluid gravitationally bound to the planet M_r and fluid trapped on horseshoe orbits $M_h = 4\pi \Sigma_g r_p x_s$. Assuming circular Keplerian orbits, the migration rate is then

$$\dot{r}_p = \frac{2\Gamma_L}{\Omega(r_p) r_p \underbrace{(M_p + M_r - \delta m)}_{M'_p}} \quad (1.2)$$

where Γ_L is the total Lindblad torque and

$$\delta m = 4\pi \Sigma_e r_p x_s - M_h = 4\pi r_p x_s (\Sigma_e - \Sigma_g) \quad (1.3)$$

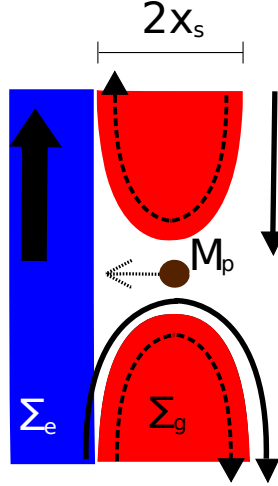


Figure 1.1: *Illustration of inwards type III migration. Fluid crossing from the inner disc to outer disc executes one horse-shoe turn behind the planet. The fluid gains angular momentum which exerts a negative torque on the planet, encouraging further orbital decay.*

is called the co-orbital mass deficit (a more precise definition is used in Chapter 2). δm essentially compares the mass supplying the torque to the mass being moved, so migration speed increases with δm .

Instabilities associated with the flow around the coorbital region are expected to modify the standard type III picture above. For the disc-planet systems to be considered, the planetary gap becomes unstable. Hence, type III migration is the most relevant form of disc-planet interaction. Instabilities contributing to Σ_e (Σ_g) will favour (disfavour) the mechanism. Fundamentally, the mechanism is angular momentum exchange due to direct gravitational scattering of disc material by the planet. Hence, instabilities that supply material for the planet to scatter may be interpreted as a modified type III migration. Explicit examples of these are described in subsequent Chapters.

1.2 Instabilities in structured astrophysical discs

Instabilities in astrophysical discs have important applications. The magneto-rotational instability (MRI, Balbus & Hawley, 1991) and gravitational instability (GI, Gammie, 2001) are robust paths to turbulent transport of angular momentum in accretion discs. The baroclinic instability (BI, Klahr & Bodenheimer, 2003; Petersen et al., 2007a,b) has also been invoked, but BI is probably more relevant to planetesimal formation. See

1.3 Previous studies and relation to present work

Armitage (2010) for a recent review of these processes in protoplanetary discs.

Instability may also be associated with steep surface density or vortensity gradients¹. This is commonly called the disc Rossby wave or vortex instability (Lovelace et al., 1999; Li et al., 2000), as it results from unstable interaction between Rossby waves across a vortensity extremum. The mechanism is similar to instability in tori (Papaloizou & Pringle, 1984, 1985, 1987). They lead to vortex formation in the non-linear regime (Hawley, 1987; Li et al., 2001). Recently, linear calculations have been extended to magnetic discs (Yu & Li, 2009; Fu & Lai, 2011) and 3D nonlinear simulations have begun (Meheut et al., 2010).

The original analysis was applied to non-self-gravitating discs. Analogous instabilities exist in structured self-gravitating particle discs (Lovelace & Hohlfield, 1978; Sellwood & Kahn, 1991) and gaseous discs (Papaloizou & Lin, 1989; Papaloizou & Savonije, 1991). These discs support global modes in addition to the localised vortices.

1.3 Previous studies and relation to present work

The stability of planetary gaps has been studied with fix-orbit simulations (Koller et al., 2003; Li et al., 2005; de Val-Borro et al., 2007). The consequence of instability on migration was noted in simulations with very low disc viscosity, when migration becomes non-monotonic (Ou et al., 2007; Li et al., 2009; Yu et al., 2010). In Chapter 2, the effect of vortex instabilities on migration is considered in detail for Saturn mass planets (previous studies consider smaller masses), and it is shown that results can be understood in the framework of type III migration.

Most works on disc-planet interactions ignore disc self-gravity. Simulations by Li et al. (2009) and Yu et al. (2010) did include self-gravity, but its effect was not explored. Lyra et al. (2009) also simulated the vortex instability in a self-gravitating disc, but for artificial internal edges. Chapter 3 extends the linear theory for the vortex instability to account for disc self-gravity, and extensive simulations performed to examine the effect of varying the strength of self-gravity. Lyra et al. noticed that more vortices form and persist longer in calculations with self-gravity switched on. The results in Chapter 3 explain this observation.

Global instabilities have been shown to exist in self-gravitating discs with prescribed gap profiles (Meschiari & Laughlin, 2008). Chapter 4 considers the stability of gaps self-consistently opened by a planet in self-gravitating discs. Analytic discussion, originally given by Sellwood & Kahn (1991) for particle discs, is extended to gaseous discs.

¹The term vortensity is used for the ratio of vorticity to surface density.

1. INTRODUCTION

Chapter 4—5 also presents the first simulations of the effect of such global modes on planetary migration.

Planetary gap stability has been studied in two-dimensional discs only (although Meheut et al. (2010) have simulated the vortex instability for artificial density bumps in three-dimensional discs). Chapter 6 presents the first three-dimensional simulations of planetary gap stability in non-self-gravitating and self-gravitating discs, as well as their effects on migration. Chapter 6 confirms 2D results in Chapters 3—5.

1.4 Modelling disc-planet systems

The studies presented in this thesis are based on direct simulations. The general numerical setup and governing equations are presented here. Specific disc models and initial conditions are described in each subsequent Chapter.

The system is a fluid disc of mass M_d with an embedded planet of mass M_p , both rotating about a central star of mass M_* . A schematic representation is shown in Fig. 1.2, in which H denotes the pressure scale-height and R is the cylindrical distance from the star. The disc is governed by the Navier-Stokes equations in a non-rotating frame in spherical polar co-ordinates $\mathbf{r} = (r, \theta, \phi)$ centred on the central star:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0, \quad (1.4)$$

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} = -\frac{1}{\rho} \nabla p + \frac{1}{\rho} \nabla \cdot \mathbf{T} - \nabla \Phi_{\text{eff}} \quad (1.5)$$

where ρ is the mass density, $\mathbf{u} = (u_r, u_\theta, u_\phi)$ is the velocity field, p is the pressure, \mathbf{T} is the viscous stress tensor and Φ_{eff} is an effective potential. The azimuthal velocity is also written $u_\phi = R\Omega$ where Ω is the angular velocity. For thin and low mass discs, the angular speed is well approximated by the Keplerian profile:

$$\Omega \simeq \Omega_k(R) \equiv \sqrt{\frac{GM_*}{R^3}}, \quad (1.6)$$

where G is the gravitational constant.

1.4.1 Units

Units such that $G = M_* = 1$ are adopted in numerical calculations, and time will be quoted in initial orbital periods of the planet, $P_0 = 2\pi/\Omega_k(r_p(0))$, where $r_p(t)$ is the planet's orbital radius.

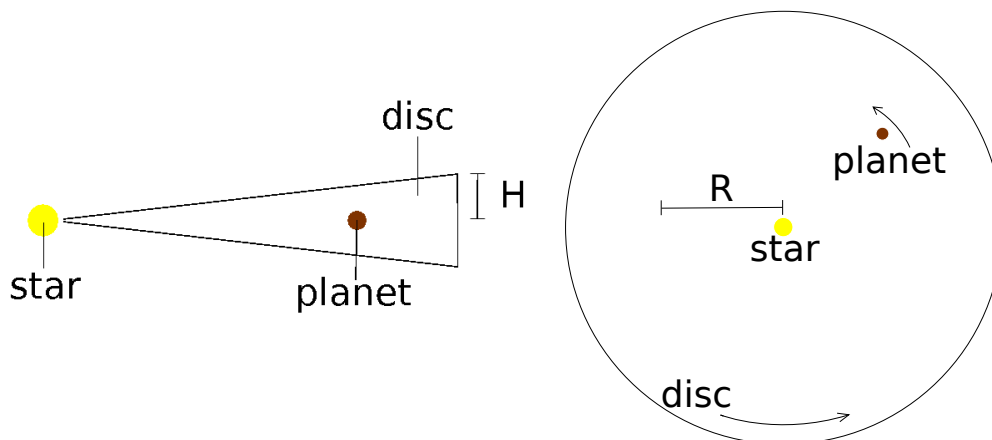


Figure 1.2: Schematic illustration of a protoplanetary disc with an embedded planet. Left: edge-on view. Right: face-on view.

1.4.2 Equations of state

Two equations of state (EOS) are adopted: *barotropic* and *locally isothermal*. These close the system of equations above without the need for an energy equation to describe the thermodynamics. Neglecting the energy equation prohibits more realistic modelling but does allow tractable analysis and faster computational simulations.

For a barotropic disc, the pressure is a function of density only:

$$p = p(\rho), \quad c_s \equiv \sqrt{\frac{dP}{d\rho}}, \quad (1.7)$$

where c_s is the sound-speed. Eq. 1.7 is used in analytical discussions, where general results may be derived without explicitly specifying the functional form of $p(\rho)$.

In a locally isothermal disc the sound-speed is a specified function of space:

$$p = c_s^2 \rho, \quad c_s = hR\Omega_k(R), \quad (1.8)$$

where $h = H/R$ is the disc aspect-ratio. Eq. 1.8 results from vertical hydrostatic balance between pressure forces and stellar gravitational force, assuming $c_s = c_s(R)$ and that the vertical extent of the disc is small compared to the local radius.

Eq. 1.8 with constant h is commonly used in disc-planet studies, and is adopted in the numerical studies later. The temperature profile of the disc is fixed in space by Eq. 1.8, even when a planet is present. It is possible to modify Eq. 1.8 to account for expected heating as disc material falls into the planet potential (see §1.4.5).

1. INTRODUCTION

1.4.3 Gravitational potentials

The effective potential is:

$$\Phi_{\text{eff}} = \Phi_* + \Phi_p + \Phi + \Phi_i, \quad (1.9)$$

where

$$\Phi_* = -\frac{GM_*}{r} \quad (1.10)$$

is the stellar potential and

$$\Phi_p = -\frac{GM_p}{\sqrt{|\mathbf{r} - \mathbf{r}_p|^2 + \epsilon_p^2}} \quad (1.11)$$

is the *softened* planet potential, where \mathbf{r}_p is the vector position of the planet and ϵ_p is the softening length.

The gravitational potential Φ arising from the disc material is given via the Poisson equation

$$\nabla^2 \Phi = 4\pi G\rho. \quad (1.12)$$

It is uncommon to include Φ in Φ_{eff} for protoplanetary discs because typically $M_d \ll M_*$. Models neglecting Φ in the hydrodynamic equations are called *non-self-gravitating* whereas models including Φ are called *self-gravitating*. Solving Eq. 1.12 usually adds substantial cost to numerical simulations.

Φ_i is the indirect potential due to the disc and the planet, which arises from the non-inertial reference frame:

$$\Phi_i(\mathbf{r}) = \int \frac{G\rho(\mathbf{r}')}{r'^3} \mathbf{r} \cdot \mathbf{r}' d^3\mathbf{r}' + \frac{GM_p}{|\mathbf{r}_p|^3} \mathbf{r} \cdot \mathbf{r}_p. \quad (1.13)$$

The indirect potential accounts for the acceleration of the co-ordinate origin relative to the inertial frame. This is an unimportant term for the dynamics of concern.

1.4.4 Viscosity

The viscous stress tensor is given by

$$\mathbf{T} = \mu \left[\nabla \mathbf{u} + (\nabla \mathbf{u})^T - \frac{2}{3} (\nabla \cdot \mathbf{u}) \mathbf{1} \right], \quad (1.14)$$

where superscript T denotes transpose. In Eq. 1.14, $\mu = \rho\nu$ is the *shear viscosity* and ν is the *kinematic viscosity*. Note that zero bulk viscosity has been assumed. Explicit expressions of \mathbf{T} in different co-ordinate systems may be found in Tassoul (1978).

Eq. 1.14 represents the effects of turbulent angular momentum transport in accretion discs. For a rotationally supported disc, angular momentum must be transported outwards if disc material is to be transported inwards and accreted onto the star. It is often easier to use Eq. 1.14 and prescribe ν to circumvent detailed modelling of turbulence. For example, if turbulence results from the MRI then one would need to consider the magnetohydrodynamic equations.

The disc models presented in subsequent Chapters adopt constant kinematic viscosity, so that ν is a parameter independent of space or time. This includes the *inviscid disc*, for which $\nu = 0$. The present work is concerned with shear instabilities that lead to large-scale structure, rather than fully fledged turbulence. A constant ν model is an adequate approach to describe the level of viscosity (or turbulence) and its effect on the instabilities of interest.

Although not used in the present work, it is worth mentioning the more frequently used α -disc models, in which one assumes

$$\nu = \alpha c_s H, \quad (1.15)$$

where $\alpha \lesssim 1$ is a dimensionless parameter (Shakura & Sunyaev, 1973) characterising turbulence of unspecified origin. Due to its popularity, it has become common practice to quote α as a non-dimensional measure of turbulence strength in an accretion disc. If a particular physical process is responsible for the turbulence, one may accordingly define an associated α and compare to α 's due to other turbulence sources.

1.4.5 Motion of the planet

The embedded planet feels the gravitational force from the protoplanetary disc, in addition to that from the star. The motion of the planet through the disc is therefore

1. INTRODUCTION

given by Newton's second law:

$$\frac{d^2 \mathbf{r}_p}{dt^2} = -\nabla(\Phi + \Phi_i + \Phi_*). \quad (1.16)$$

The contribution due to the disc is given by

$$\left(\frac{d^2 \mathbf{r}_p}{dt^2}\right)_{\text{disc}} = - \int \frac{G\rho(\mathbf{r})(\mathbf{r}_p - \mathbf{r})}{(|\mathbf{r}_p - \mathbf{r}|^2 + \epsilon_p^2)^{3/2}} d^3 \mathbf{r}, \quad (1.17)$$

where the distance between the planet and a fluid element has been softened to prevent a singularity in numerical computations. Disc-planet forces arising near the planet are subject to numerical errors because of the diverging potential and finite resolution. To overcome this issue one can apply a Gaussian envelope to disc-planet forces within the Hill radius $r_h = (M_p/3M_*)^{1/3}r_p$ of the planet. This procedure is called tapering.

In some of the simulations a modified locally isothermal equation of state is adopted, with sound-speed given by

$$c_s = \frac{hRh_p d_p}{[(hR)^{7/2} + (h_p d_p)^{7/2}]^{2/7}} \sqrt{\Omega_k^2 + \Omega_{\text{kp}}^2}, \quad (1.18)$$

where $\Omega_{\text{kp}}^2 = GM_p/d_p^3$ and $d_p = \sqrt{|\mathbf{r} - \mathbf{r}_p|^2 + \epsilon_p^2}$ is the softened distance to the planet. Increasing the parameter h_p increases the temperature near \mathbf{r}_p , thereby decreasing the mass built up near the planet and hence numerical errors from this region. This EOS is taken from Pepliński et al. (2008a), in which h_p has the physical meaning of the aspect-ratio of the circumplanetary disc. However, the use of Eq. 1.18 here is entirely numerical.

1.4.6 Two-dimensional approximation

Protoplanetary discs are thin because $h \ll 1$. A good approximation is the *razor-thin disc*. The fluid equations (Eq. 1.4) are integrated over the vertical extent of the disc, assuming no vertical motions and the fluid radial and azimuthal velocities are independent of the cylindrical z . The planet is restricted to move in the midplane ($z = 0$).

The co-ordinate system becomes (r, ϕ) cylindrical polars by setting $\theta = \pi/2$. The three-dimensional density ρ is replaced by surface density Σ , \mathbf{u} becomes the two-dimensional velocity (u_r, u_ϕ) and p is re-interpreted as the vertically integrated pressure. In the integral expressions above, the expression for the mass element $\rho d^3 \mathbf{r}$ is replaced

by $\Sigma d^2\mathbf{r}$.

The 2D hydrodynamic equations are

$$\begin{aligned} \frac{\partial \Sigma}{\partial t} + \nabla \cdot (\mathbf{u}\Sigma) &= 0, \\ \Sigma \left(\frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} \right) &= -\nabla p - \Sigma \nabla \Phi_{\text{eff}} + \mathbf{f}, \end{aligned} \quad (1.19)$$

where \mathbf{f} is the 2D viscous force. All quantities in Eq. 1.19 are evaluated at $z = 0$. The kinematic viscosity is now related to the shear viscosity by $\mu = \nu\Sigma$. Explicit expressions for \mathbf{f} are given in Masset (2002).

The Poisson equation for a 2D fluid is

$$\nabla^2 \Phi = 4\pi G \Sigma \delta(z) \quad (1.20)$$

where $\delta(z)$ is the Dirac delta function. Note that a 2D disc still has a 3D potential. Eq. 1.20 can be used for analytical discussions, but it cannot be used in numerical computations. For 2D numerical work, the midplane disc potential is given by

$$\Phi = - \int \frac{G \Sigma(r', \phi') r' dr' d\phi'}{\sqrt{r^2 + r'^2 - 2rr' \cos(\phi - \phi') + \epsilon_g^2}}, \quad (1.21)$$

where ϵ_g is a softening length. In a 3D disc, material is spread over the vertical extent, which dilutes the disc and planet gravitational forces compared to the case where all the disc material is confined in the plane. Thus in 2D, softening accounts for the vertical extent of the disc, in addition to preventing numerical divergence.

1.4.7 Stability parameters

The *vortensity profile* and *Toomre Q parameter* are of fundamental importance to disc stability. They arise naturally in linear stability analysis of 2D discs. Vortensity concerns shear instabilities and Toomre Q concerns gravitational instability.

1.4.8 Vortensity

The vortensity η is

$$\eta \equiv \frac{\nabla \times \mathbf{u}}{\rho} = \frac{\omega}{\rho}, \quad (1.22)$$

1. INTRODUCTION

where $\boldsymbol{\omega}$ is the absolute vorticity, defined from the velocity in a non-rotating frame. In 2D models, ρ is replaced by Σ and there is only one component of vortensity, $\boldsymbol{\eta}_{2D} = \eta \hat{z}$ where

$$\eta = \frac{1}{r\Sigma} \left[\frac{\partial}{\partial r} (ru_\phi) - \frac{\partial u_r}{\partial \phi} \right]. \quad (1.23)$$

Furthermore, if the 2D disc is axisymmetric ($\partial/\partial\phi = 0$) or has zero radial velocity, then η can be written as

$$\eta = \frac{\kappa^2}{2\Omega\Sigma}, \quad (1.24)$$

where

$$\kappa^2 = \frac{1}{r^3} \frac{\partial}{\partial r} (r^4 \Omega^2) \quad (1.25)$$

is the square of the epicycle frequency. κ is the oscillation frequency of a fluid element perturbed about a circular orbit. The existence of vortensity extrema allow shear instabilities (e.g. Lovelace et al., 1999).

1.4.9 Toomre parameter

The Toomre parameter is

$$Q = \frac{c_s \kappa}{\pi G \Sigma}, \quad (1.26)$$

and was originally derived in a linear analysis of the gravitational stability of razor-thin discs (Toomre, 1964). Hence, the surface density appears in Eq. 1.26 rather than ρ .

The Toomre Q is often used to describe the strength of disc self-gravity. The razor-thin disc is gravitationally unstable to axisymmetric, local perturbations when $Q < 1$, although general gravitational instabilities may appear at somewhat larger values (e.g. Papaloizou & Lin, 1989; Papaloizou & Savonije, 1991).

Chapter 2

Vortex-induced type III migration

A sufficiently massive planet in a protoplanetary disc will induce spiral shocks extending close to the planet's orbital radius. Vortensity generation across shock tips results in thin high vortensity rings outlining the gap. However, vortensity rings are unstable if the disc viscosity is sufficiently small (de Val-Borro et al., 2007) and lead to vortex formation.

This Chapter considers the effect of such vortices on planetary migration and is organised as follows. §2.1 describes the disc-planet models and parameters. The formation of vortensity rings is studied with numerical simulations and a semi-analytical model in §2.2—§2.3. The linear stability problem is discussed in §2.4. Nonlinear hydrodynamic simulations of giant planet migration as a function of viscosity are presented in §2.5, high-lighting the effect of vortices at low viscosity. The inviscid case is examined in detail in §2.6, including simulations exploring the effect of disc mass, temperature and planetary mass. Additional simulations concerning numerical issues are described in §2.7. §2.8 concludes this Chapter.

2.1 Disc-planet setup

The protoplanetary disc is modelled as razor-thin and non-self-gravitating. The appropriate hydrodynamic equations are thus Eq. 1.19 with the disc potential Φ neglected. The general setup is described here and problem-dependent parameters are described in the following sections. The initial orbital radius of the planet is $r_p(t = 0) = 2$.

The disc has initially uniform surface density $\Sigma = \Sigma_0 \times 10^{-4}$. Throughout sim-

2. VORTEX-INDUCED TYPE III MIGRATION

ulations, a uniform kinematic viscosity $\nu = \nu_0 \times 10^{-5}$ is imposed, with ν_0 being a dimensionless constant. The locally isothermal equation of state is adopted (§1.4.2) with constant aspect-ratio h . Discs with $\Sigma_0 = 1-9$, $\nu_0 = 0-1$ and $h = 0.03-0.06$ are considered. Planetary masses of $q \equiv M_p/M_* = 10^{-4}-10^{-3}$ are used. If $M_* = 1M_\odot$, then $q = 3 \times 10^{-4}$, 10^{-3} corresponds to Saturn and Jupiter, respectively. The planet position in the disc plane is denoted (r_p, ϕ_p) and its potential is softened with a softening length $\epsilon_p = 0.6hr_p$.

2.1.1 Vortensity conservation

The vortensity η is of fundamental importance to the issue to gap stability. For a 2D disc the vortensity is simply

$$\eta = \frac{\omega}{\Sigma} \equiv \frac{\omega_r + 2\Omega_p}{\Sigma},$$

where $\omega = \hat{z} \cdot \nabla \times \mathbf{u}$ is the z component of the absolute vorticity and ω_r is the vertical component of the relative vorticity seen in a rotating frame with angular speed Ω_p .

It is well known that in inviscid barotropic flows without shocks, the vortensity $\eta = \omega/\Sigma$ is conserved for a fluid particle ($D\eta/Dt = 0$). When a locally isothermal equation of state with spatially varying sound-speed as is used here is adopted, vortensity is no longer strictly conserved. However, the adopted sound-speed profile varies on a global scale so that when phenomena are considered on a local scale, vortensity is conserved to a good approximation in the absence of shocks and viscosity. Thus, analysing the vortensity distribution for inviscid discs allows one to track fluid elements as their orbits are perturbed by the planet.

2.1.2 Numerical method for nonlinear simulations

The `FARGO` code is used to numerically evolve the hydrodynamics equations describing the disc in response to the planet potential (Masset, 2000a,b). The code and its extended versions are publicly available ¹. `FARGO` uses a finite-difference scheme with van Leer upwind advection, similar to the popular `ZEUS` code (Stone & Norman, 1992a), but `FARGO` employs a modified algorithm for azimuthal transport that allows for large time steps. The 2D disc is divided into $N_r \times N_\phi$ zones. The grid is uniformly spaced in radius and azimuth. Wave damping boundary conditions are imposed at disc boundaries (de Val-Borro, 2006) and 2π -periodic boundary condition imposed in azimuth. Since vortex formation occur near the planet, the phenomenon is unaffected by distant

¹<http://fargo.in2p3.fr/>

2.2 Ring structures in disc-planet interactions

radial boundary conditions.

A fifth order Runge-Kutta method was used to integrate the planet's equation of motion when migration is considered. Simulations ran with halved time-step show very similar results to the corresponding standard set-up. The RK5 integrator is sufficiently accurate to capture the vortex-planet interaction, which can be regarded as a two-body problem.

2.2 Ring structures in disc-planet interactions

2.2.1 Vortensity generation by shocks

Shocks are induced by a sufficiently massive planet and vortensity conservation is broken in the presence of shocks. A high-resolution numerical disc-planet simulation demonstrates this below. The disc is thin and inviscid ($h = 0.05$, $\nu = 0$), occupying $r \in [1, 3]$ with $\Sigma_0 = 1$. The planet has mass $M_p = 2.8 \times 10^{-4} M_*$, its potential is introduced at $t = 0$ and switched on to its full value over $5P_0$. The planet is held on a fixed circular Keplerian orbit. The computational domain is divided into $N_r \times N_\phi = 1024 \times 3072$ zones, corresponding to radial and azimuthal resolution of $\Delta r \simeq 0.02r_h$ and $r\Delta\phi \simeq 0.05r_h$.

Fig.2.1 shows the vortensity field at $t = 7.07P_0$ close to the planet. Vortensity is generated/destroyed as material passes through the two spiral shocks. For the outer shock, vortensity generation occurs for fluid elements executing a horseshoe turn ($|r - r_p| \lesssim r_h$) while vortensity is reduced for fluid that passes by the planet, but the change is smaller in magnitude in the latter case. The situation is similar for the inner shock, but some post-shock material with increased vortensity continues to pass by the planet. Note that a pre-shock fluid element that would be on a horseshoe trajectory, may in fact pass by the planet after crossing the shock. It is clear that vortensity rings originate from passage through shock fronts interior to the co-orbital region that would correspond to the horseshoe region for free particle motion.

The streams of high vortensity eventually move around the whole orbit outlining the entire co-orbital region. Fig. 2.1 shows that they are generated along a small part of the shock front of length $\sim r_h$. This results in thin rings with a similar radial width. The fact that they originate from horseshoe material can enhance the contrast as post-shock inner disc horseshoe material is mapped from $r - x$ to $r + x$ to become adjacent to post-shock outer disc material passing by the planet, and the two trajectories experience opposite vortensity jumps.

2. VORTEX-INDUCED TYPE III MIGRATION

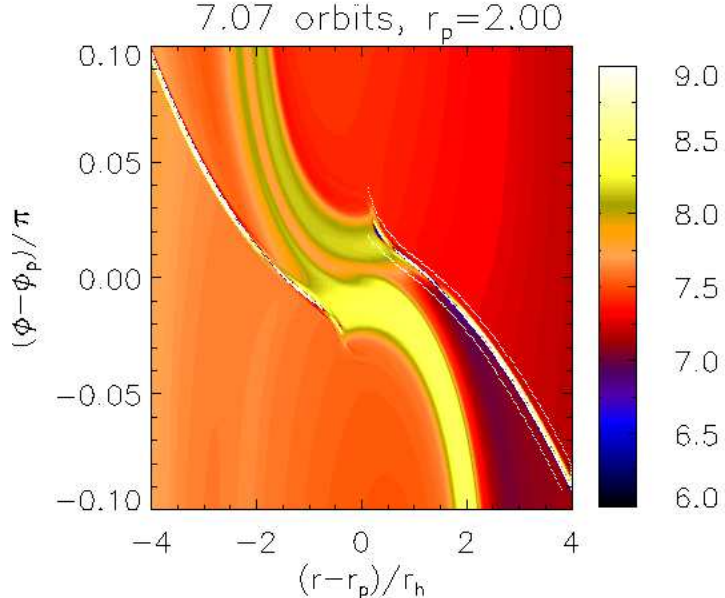


Figure 2.1: *Vortensity generation and destruction across shocks induced by a Saturn-mass planet in an inviscid disc. The plot shows $\ln \eta = \ln(\omega/\Sigma)$. Regions of increased vortensity are clearly visible as half horseshoes above and below the planet. The increase occurs as material passes through parts of the shock fronts that extend into the co-orbital region. Thick white lines indicate the shock front, whereas thin white lines above (below) the outer shock indicate where pre-shock (post-shock) quantities are measured.*

2.2.2 Long term evolution

Fig. 2.2 shows the long-term evolution of average co-orbital properties from a corresponding lower resolution run ($N_r \times N_\phi = 256 \times 768$). The co-orbital region is taken to be the annulus $[r_p - x_s, r_p + x_s]$, with $x_s = 2.5r_h$ as is typically measured from hydrodynamic simulations for intermediate or massive planets (Artymowicz, 2004b; Paardekooper & Papaloizou, 2009). In Appendix A, it is shown that in the pressureless limit, $x_s \lesssim 2.3r_h$, comparable to the value adopted above.

Vorticity generation occurs within $t \lesssim 25P_0$, after which it remains approximately steady. It is important to note that subsequent vortensity increases is in narrow rings and fluctuations are due to instabilities associated with the rings. The average surface density falls as the planet opens a gap. In fact, gap formation is a requirement for consistency with vortensity rings. Fig. 2.2 reflects modification of co-orbital properties on dynamical timescales due to shocks, so migration mechanisms that depend on co-orbital structure will be affected.

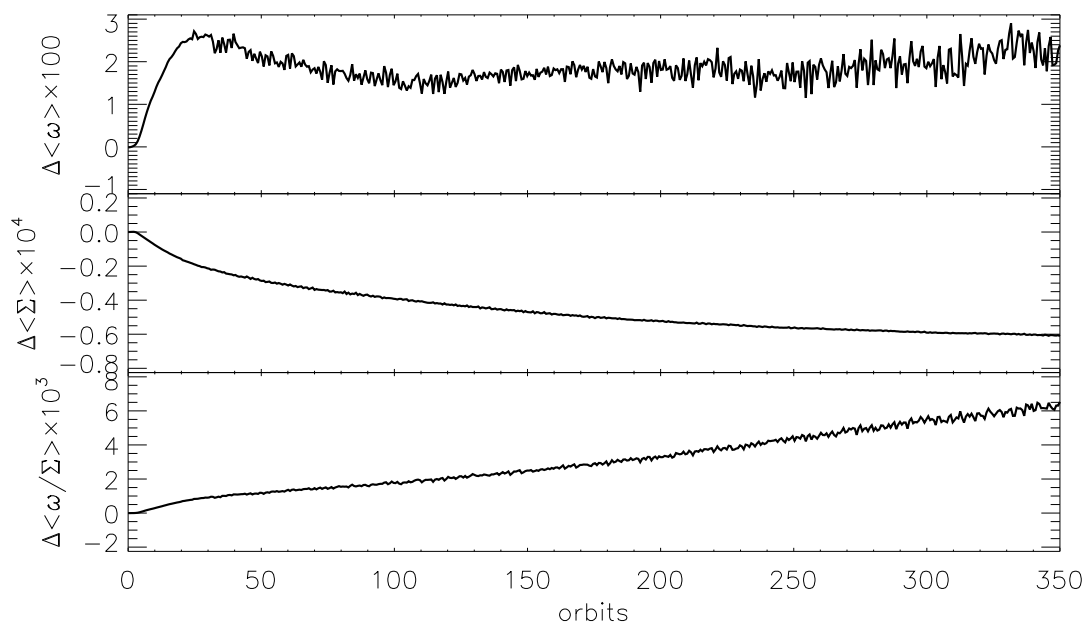


Figure 2.2: Long term evolution of co-orbital properties for a Saturn-mass planet held on fixed orbit. Vorticity, density and vortensity perturbations are expressed relative to their values at $t = 0$. Angle brackets denote an average over the co-orbital region defined by the annulus $r = [r_p - 2.5r_h, r_p + 2.5x_s]$.

2.2.3 Location of vortensity generation for different planet masses

Vortensity ring formation has been observed in disc-planet simulations by Koller et al. (2003) and Li et al. (2005), but there the rings are generated by shocks *outside* the co-orbital region whereas here rings are inside it. This is because the authors used a smaller planet mass. To illustrate this, simulations with $M_p/M_* = 10^{-4}$, 2.8×10^{-4} and 10^{-3} are compared in Fig. 2.3, which show azimuthally averaged vortensity perturbations induced after $t = 14.14P_0$.

The intermediate and high mass cases are qualitatively similar, having $\Delta(\omega/\Sigma)$ maxima at $r - r_p \sim \pm 2r_h$ and minima at $r - r_p \sim 3r_h$. The magnitude of $\Delta(\omega/\Sigma)$ increases with M_p because higher masses induce stronger shocks with higher Mach number. As the half-width of the horseshoe region is $x_s = 2.5r_h$ for such masses, vortensity rings are co-orbital features.

The low mass $M_p = 10^{-4}M_*$ case has much smaller $|\Delta(\omega/\Sigma)|$. There is a vortensity maximum at $r - r_p = -3r_h$ but nothing of similar magnitude at $r - r_p > 0$. Paardekooper & Papaloizou (2009) found the co-orbital half-width, in the limit of zero softening for

2. VORTEX-INDUCED TYPE III MIGRATION

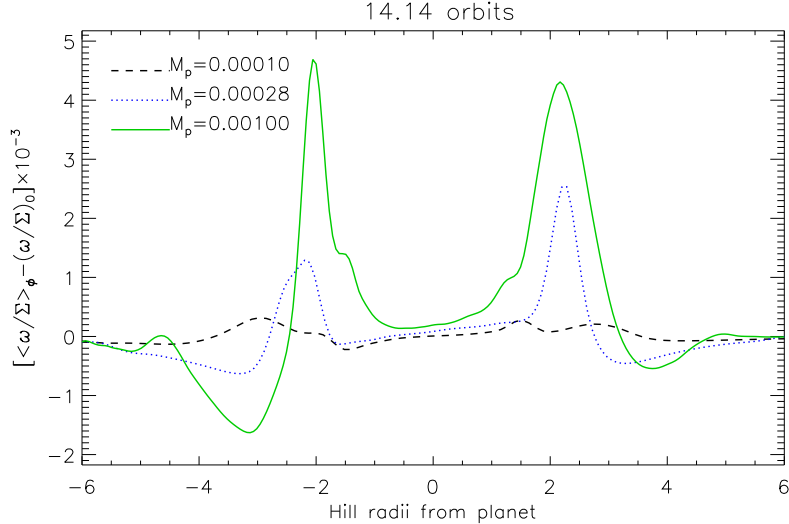


Figure 2.3: Azimuthally averaged vortensity perturbation for different planet masses M_p (in units of the central mass). The resolution for these runs were $N_r \times N_\phi = 256 \times 768$.

low mass planets, to be:

$$x_s = 1.68(q/h)^{1/2}r_p.$$

For $q \equiv M_p/M_* = 10^{-4}$, $h = 0.05$, $x_s = 2.33r_h$. Non-zero softening gives a smaller x_s . Hence, vortensity rings for low mass planets occur outside their co-orbital region as is confirmed and they are very much weaker. Note that in Li et al. (2009) significantly longer timescales are simulated.

In the standard type III migration described by Masset & Papaloizou (2003), the migrating system is the planet plus the co-orbital material. If vortensity rings are co-orbital features, they should migrate with the planet. On the other hand, if they reside just outside the horseshoe region, the migrating system must work against these rings to migrate. Because type III migration is driven by the contrast between co-orbital and circulating material, whether vortensity rings count as the former or the latter can affect type III migration. Furthermore, the fact that vortensity rings lie closer to the planet (in units of Hill radius) as M_p increases imply that instabilities associated with the rings will interact more readily with more massive planets.

2.3 A semi-analytic model for co-orbital vortensity generation by shocks

A semi-analytic model describing shock-generation of vortensity is presented in order to provide an understanding of the physical processes involved. More specifically, the outer spiral shock in Fig. 2.1 is modelled. Three pieces of information are required: the pre-shock flow field, the shock front location and the vortensity change undergone by material as it passes through the shock.

2.3.1 Flow field

The pre-shock flow field is calculated in the shearing sheet approximation (e.g. Paardekooper & Papaloizou, 2009). A local Cartesian co-ordinate system (x, y) is adopted, which co-rotates with the planet at angular speed Ω_p . The co-ordinates x, y correspond to radial and azimuthal directions, respectively. Velocities in the shearing sheet are thus relative to the planet. The unperturbed velocity field in the shearing sheet is assumed Keplerian $\mathbf{u} = (0, -3\Omega_p x/2)$.

In the pre-shock flow, pressure forces are assumed negligible compared to the planet potential. This ballistic approximation is appropriate for a slowly varying supersonic flow as is expected to be appropriate for the pre-shock fluid. Pressure should also be insignificant because giant planet masses are being considered. Fluid elements then behave like test particles moving under the potential:

$$\Phi_{\text{tot}} = -\frac{GM_p}{\sqrt{x^2 + y^2 + \epsilon_p^2}} - \frac{3}{2}\Omega_p^2 x^2. \quad (2.1)$$

The particle dynamic equations are those listed in Appendix A, except here the planet potential is softened. Indirect potentials are neglected in this treatment for simplicity, and viscosity ignored for consistency (Fig. 2.1 is an inviscid simulation). In this section, the equations of motion are integrated numerically. In Appendix B, a linearised version of this calculation is presented.

A fluid particle's trajectory in a steady state flow is written as $x = x(y)$ and the corresponding velocity field as $\mathbf{u} = \mathbf{u}(y)$. Noting that on a particle trajectory $\frac{D}{Dt} = u_y \frac{d}{dy}$, a particle trajectory can be described by the following system of three

2. VORTEX-INDUCED TYPE III MIGRATION

simultaneous first order differential equations:

$$\frac{du_y^2}{dy} = -4\Omega_p u_x - \frac{2GM_p y}{(x^2 + y^2 + \epsilon_p^2)^{3/2}} \equiv \mathcal{Q}, \quad (2.2)$$

$$\frac{d}{dy} (u_x u_y^2) = u_y \left[3\Omega_p^2 x - \frac{GM_p x}{(x^2 + y^2 + \epsilon_p^2)^{3/2}} \right] + u_x \mathcal{Q} + 2\Omega_p u_y^2, \quad (2.3)$$

$$\frac{d}{dy} (x u_y^2) = x \mathcal{Q} + u_x u_y. \quad (2.4)$$

Eq.2.2—2.4 are equivalent to Eq. A.1, but this particular form is chosen so that the RHS of Eq. 2.2—2.4 remain finite at a horseshoe U-turn, where $u_y = 0$. The explicit expressions for du_y/dy involve divisions by u_y , which would be numerically inconvenient close to the U-turn. The state vector $\mathbf{U}(y) = [u_y^2, u_x u_y^2, x u_y^2]$ is solved for a particular particle in $x > 0$, with the boundary condition

$$(u_y, u_x, x) \rightarrow (-3\Omega_p x_0/2, 0, x_0) \text{ as } y \rightarrow \infty, \quad (2.5)$$

where $x = x_0$ is the particle's unperturbed path (or impact parameter). A simple fourth order Runge-Kutta integrator was used. The totality of paths obtained by considering a continuous range of x_0 then constitutes the flow field. Having obtained the velocity field, vortensity conservation is used to obtain the surface density,

$$\frac{\omega}{\Sigma} = \frac{\omega_0}{\Sigma_i}, \quad (2.6)$$

where the unperturbed absolute vorticity in the shearing sheet is $\omega_0 = \Omega_p/2$ and Σ_i is the surface density at the initial particle position. Since the unperturbed surface density is uniform, all particles have the same vortensity at the beginning of their trajectory. When a particle reaches position (x, y) , the surface density there is given by

$$\Sigma(x, y) = \frac{2\Sigma_i}{\Omega_p} [\omega_r(x, y) + 2\Omega_p]. \quad (2.7)$$

Numerically, the relative vorticity is calculated by relating it to the circulation through

$$\omega_r \simeq \frac{1}{\Delta A} \oint \mathbf{u} \cdot d\mathbf{l} \quad (2.8)$$

where the integration is taken over a closed loop about the point of interest and ΔA is the enclosed area. This avoids errors due to numerical differentiation on the uneven grid in (x, y) generated by solving the above system.

2.3 A semi-analytic model for co-orbital vortensity generation by shocks

2.3.2 Location of the shock front

A shock forms because the planet presents an obstacle to the flow. Where the relative flow is subsonic, the presence of this obstacle can be communicated to the fluid via sound waves. In supersonic regions, the fluid is unaware of the planet via sound waves (but the planet's gravity is felt). An estimate the location of the boundary separating these two regions can be made by specifying an appropriate characteristic curve or ray defining a sound wave. This is a natural location for shocks. Applying this idea to Keplerian flow, Papaloizou et al. (2004) obtained a good match between the predicted theoretical shock front and the wakes associated with a low mass planet (which perturbs the Keplerian flow by a negligible amount). For general velocity field \mathbf{u} the characteristic curves satisfy the equation

$$\frac{dy_s}{dx} = \frac{\hat{u}_y^2 - 1}{\hat{u}_x \hat{u}_y - \sqrt{\hat{u}_x^2 + \hat{u}_y^2 - 1}}, \quad (2.9)$$

where $\hat{u}_i = u_i/c_s$. A brief derivation of Eq. 2.9 is given in Appendix C.1. The sign of the square root has been chosen so that the curves have negative slope in the domain $x > 0$ with $u_y < 0$ (as required by the outer shock in Fig. 2.1). The fluid flows from $y > y_s$ (super-sonic) to $y < y_s$ (sub-sonic). Fluid located at $y = y_s$ begins to know about the planet through pressure waves (Papaloizou et al., 2004).

The solution to Eq. 2.9 for Keplerian flow is

$$\hat{y} = -\frac{\hat{x}}{2}\sqrt{\hat{x}^2 - 1} + \frac{1}{2}\ln\left(\sqrt{\hat{x}^2 - 1} + \hat{x}\right), \quad (2.10)$$

where $\hat{x} = x/x_i$, $\hat{y} = y/x_i$ with x_i being the starting point of the curve (Papaloizou et al., 2004). In Keplerian flow, the rays defining the shock fronts originate from the sonic points at $x = x_i = \pm 2H/3$, $y = 0$.

In a general flow, sonic points ($|\mathbf{u}| = c_s$) at which the rays may start, lie on curves and can occur for $x < 2H/3$. This is possible, for example, if the flow is perturbed by a large planet mass and fluid particles move supersonically across a U-turn as they are accelerated by the planet. In this case, shocks can extend close to the planet's orbital radius. The starting sonic point for solving Eq. 2.9 that is eventually adopted has $x = 0$ (the lowest dotted curve in Fig. 2.4, see §2.3.4).

2. VORTEX-INDUCED TYPE III MIGRATION

2.3.3 Vorticity and vortensity changes across a shock

The jump in absolute vorticity $[\omega]$ is readily obtained by resolving the fluid motion parallel and perpendicular to the shock front (eg. Kevlahan, 1997). As the energy equation is neglected, shocks are locally isothermal and the expression for $[\omega]$ here differs from those of Kevlahan (1997), accordingly a brief derivation of $[\omega]$ is presented in Appendix C.2. The result for a steady shock is

$$[\omega] = -\frac{(M^2 - 1)^2}{M^2} \frac{\partial u_\perp}{\partial S} + (M^2 - 1)\omega - \left(\frac{M^2 - 1}{u_\perp}\right) \frac{\partial c_s^2}{\partial S}, \quad (2.11)$$

where u_\perp is the pre-shock velocity component perpendicular to the shock front, $M = u_\perp/c_s$ is perpendicular Mach number and ω is the pre-shock absolute vorticity. $\partial/\partial S$ is the derivative along the shock (increasing S is taken to be moving away from the planet, see Fig. C.1). It is important to note that for Eq. 2.11 to hold, the direction of increasing S , the direction of positive u_\perp , and the vertical direction should form a right handed triad.

The vortensity jump $[\omega/\Sigma]$ follows immediately from Eq. 2.11 as

$$\left[\frac{\omega}{\Sigma}\right] = -\frac{(M^2 - 1)^2}{\Sigma M^4} \frac{\partial u_\perp}{\partial S} - \left(\frac{M^2 - 1}{\Sigma M^2 u_\perp}\right) \frac{\partial c_s^2}{\partial S}, \quad (2.12)$$

which reduces to the expression derived by Li et al. (2005) if $c_s = \text{constant}$ (and a sign change due to different convention). The sign of $[\omega/\Sigma]$ depends mainly on the gradient of u_\perp (or M) along the shock. For the outer shock induced by the planet $u_\perp < 0$, so the width of the increased vortensity rings produced is determined by the length along the shock where $|M|$ is increasing. Note that $[\omega/\Sigma]$ does not depend explicitly on the pre-shock vortensity, unlike the absolute vorticity jump.

The last term on the RHS of Eq. 2.11–2.12 is produced by the local isothermal equation of state. For the $c_s \propto r^{-1/2}$ profile adopted in this study, the sound-speed varies slowly in the region of vortensity generation, which occurs locally. Hence, the contribution from the $\partial c_s^2/\partial S$ term to the vorticity/vortensity jump is not important here (as checked numerically). However, Eq. 2.11–2.12 are valid for any sound-speed profile. Should a c_s profile varying on a local scale be adopted, then $\partial c_s^2/\partial S$ could be important.

2.3 A semi-analytic model for co-orbital vortensity generation by shocks

2.3.4 Comparing the semi-analytic model to numerical simulations

The three subsections above can be combined and compared to results of hydrodynamic simulations. First, Fig. 2.4 illustrates the particle paths that constitute the flow field together with the shock fronts obtained by assuming coincidence with the characteristic curves that are obtained from the semi-analytic model. A polynomial fit to the simulation shock front is also shown. Particle paths cross for $x > r_h$, $y < -2r_h$ so that the neglect of pressure becomes invalid. Accordingly the solution to Eq. 2.4 should not be trusted in this region. Fortunately, vortensity generation occurs within r_h from the planet, where pre-shock particle paths do not cross.

In Fig. 2.4, the estimated shock location is qualitatively good and tends to the Keplerian solution further away. The important feature is that the shock can extend close to $x = 0$, across horseshoe orbits. If the flow were purely Keplerian there could be no significant vortensity generation close to $x = 0$ because the flow becomes sub-sonic and no shock occurs there. Only circulating fluid can be shocked in that case. This implies that for low mass planets where the flow is nearly Keplerian, shock-generation of vortensity cannot occur close to the planet's orbital radius. Shock-generation of vortensity inside the co-orbital region is only possible for sufficiently massive planets that induce large enough non-Keplerian velocities.

The key quantity determining vortensity generation is the perpendicular Mach number. Fig. 2.5 compares M from simulation and model. Although the semi-analytic model gives a shock Mach number that is somewhat smaller than found from the simulation, all curves have $|M|$ increasing from $x = 0 \rightarrow 1.3r_h$ which is the important domain for vortensity generation. Thus Eq. 2.12 implies vortensity generation in this region for all cases.

Fig. 2.6(a) illustrates various combinations of semi-analytic modelling and simulation data that have been used to estimate the vortensity jump across the shock. The qualitative similarities between the various curves confirms that vortensity generation occurs within co-orbital material about one Hill radius away from the planet. It is shocked as it undergoes horseshoe turns.

Assuming material is mapped from $x \rightarrow -x$ as it switches to the inner leg of its horseshoe orbit, the outer spiral shock will produce a vortensity ring peaked at $r - r_p \sim -0.5r_h$ of width $O(r_h)$ ($= O(H)$). Similarly the inner shock is expected to produce a vortensity ring peaked at $r - r_p \sim 0.5r_h$. As a fluid element moves away from the U-turn region, $|r - r_p|$ increases, but it remains on a horseshoe orbit. Thus, thin vortensity rings are natural features of the co-orbital region for such planet masses.

2. VORTEX-INDUCED TYPE III MIGRATION

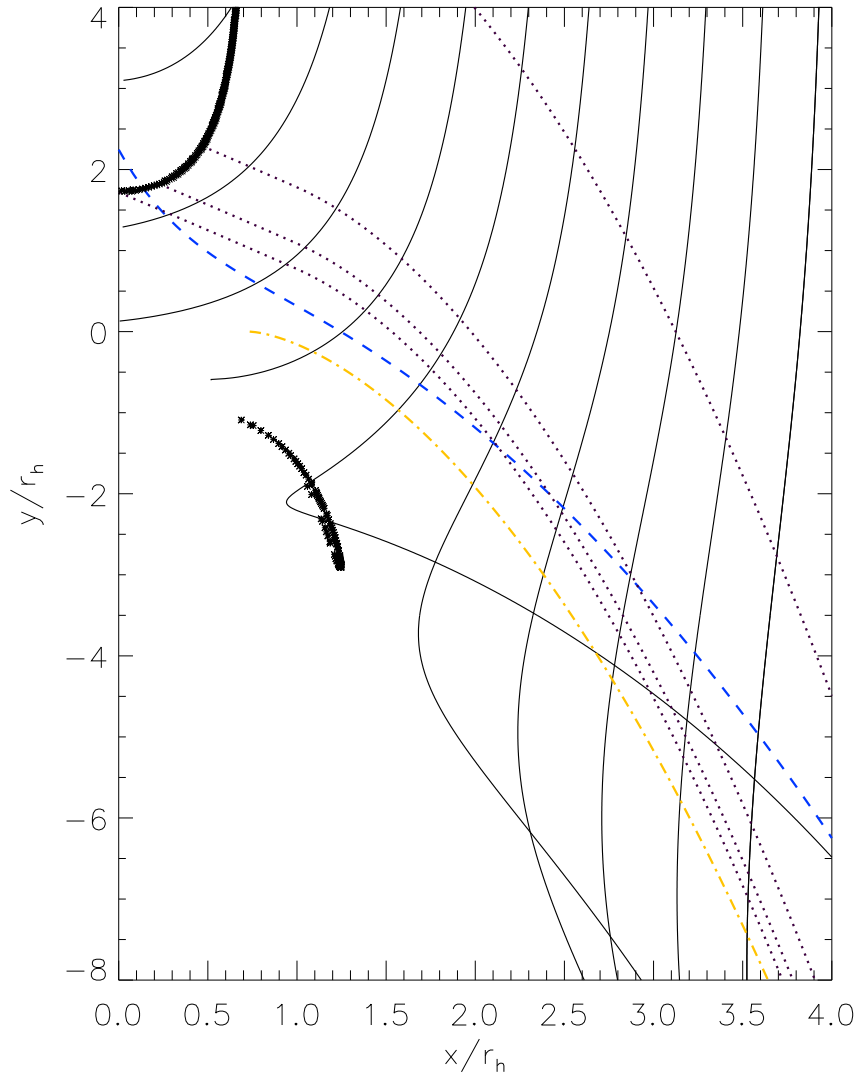


Figure 2.4: *Solid lines: particle paths from the reduced zero-pressure momentum equations (Eq. 2.2 - 2.4); thick lines: curves composed of sonic points on which $|\mathbf{u}| = c_s$; dotted lines: theoretical shock fronts; dash-dot line: solution to Eq. 2.9 for Keplerian flow; dashed line: polynomial fit to simulation shock front. The actual shock front begins at a sonic point around $x = 0.2r_h$.*

The model has also been tested for $M_p = 2 \times 10^{-4} M_*$ in Fig. 2.6(b). There is still a good qualitative match between the simulation and the model; even though lowering M_p makes the zero-pressure approximation, adopted to determine the semi-analytic flow field, less good. Decreasing M_p shifts vortensity generation away from the planet

2.3 A semi-analytic model for co-orbital vortensity generation by shocks

in the semi-analytic model, as is also observed in the hydrodynamic simulation (Fig. 2.3). In this case, there is no vortensity generation in $r - r_p < 0.5r_h$ but vortensity rings are still co-orbital (with peaks at $\sim 0.7r_h$).

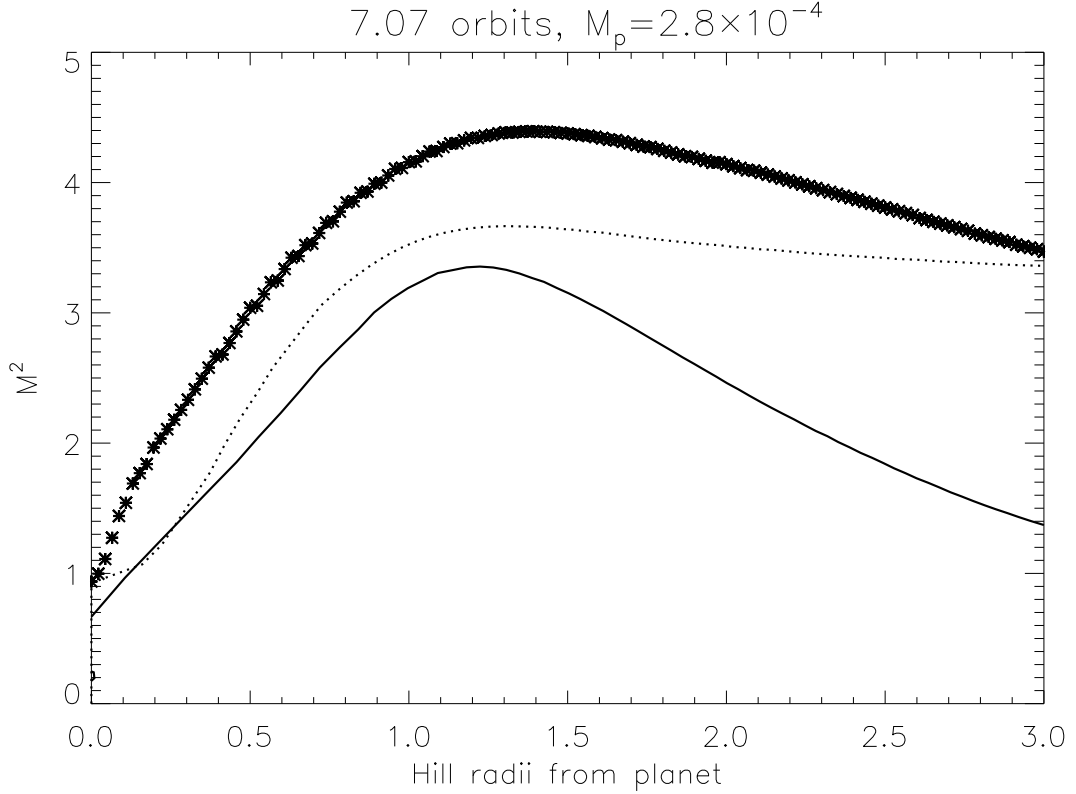


Figure 2.5: Square of the perpendicular Mach number along the outer spiral shock illustrated in Fig. 2.1. Asterisks: simulation data; dotted line: M^2 obtained from the simulation shock front with the semi-analytic flow; solid line: M^2 obtained from the semi-analytic shock front and flow. The planet mass is in units of M_* .

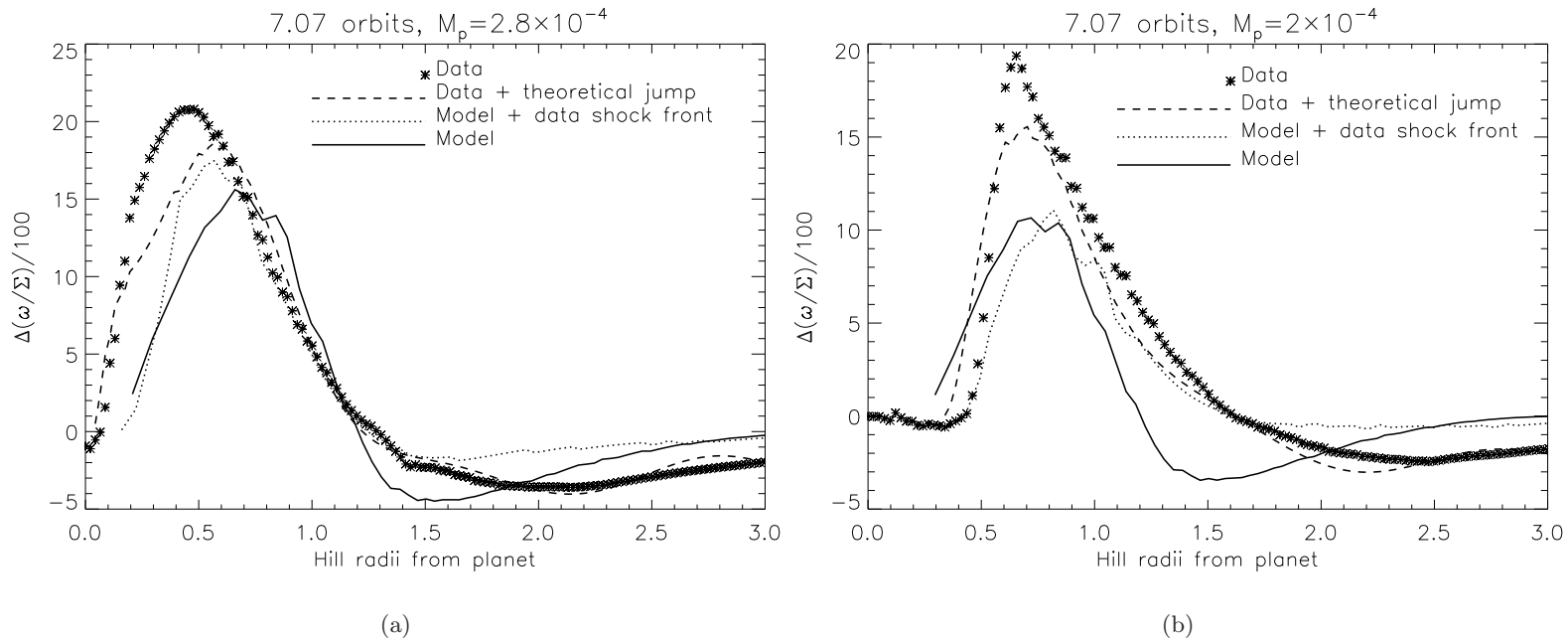


Figure 2.6: Semi-analytic and actual vortensity jumps across a shock in the co-orbital region. Asterisks are measured from the simulations. The dashed lines were obtained by using pre-shock simulation data coupled with the jump condition specified by Eq.2.12. The dotted lines were obtained using the semi-analytic flow field together with the location of the shock front obtained from the simulation. The solid curves correspond to the semi-analytic model for both the flow field and shock front. The planet mass is in units of M_* .

2.3 A semi-analytic model for co-orbital vortensity generation by shocks

2.3.5 A recipe for vortensity rings

The modelling above can be used to construct the azimuthally averaged vortensity profile of the shock-modified co-orbital region of a giant planet. However, this calculation is only illustrative due to the highly nonlinear nature of shocks and flows near a giant planet, plus the various assumptions already made in estimating the pre-shock flow and shock-front. It is usually more practical to run nonlinear disc-planet simulations to obtain accurate vortensity profiles.

To make progress, further assumptions are made:

1. Vortensity of a fluid element is only changed across the shock and conserved elsewhere (see §2.1.1).
2. Symmetric spiral shocks. The jump in vortensity across the shock, $[\omega/\Sigma]$, for $x < 0$ (inner spiral shock) is the same as for the outer spiral shock. This is consistent with the shearing sheet.
3. The shearing sheet model above is used to obtain $[\omega/\Sigma]$. The vortensity jump is then applied to the unperturbed vortensity profile of the global model. Assuming Keplerian rotation, this is $\eta_0 = r^{-3/2}/(2\Sigma)$ with Σ being uniform.

In Fig. 2.6(a)—2.6(b), the horizontal axis corresponds to the x -coordinate of the shock front. A fluid element which shocks at a particular x originated from a location $(x_0, y > 0)$. Fluid particles with $x_0 < x_s \equiv 2.5r_h$ are assumed to be on horseshoe orbits and the shock does not change the nature of such trajectories. Horseshoe particles are then mapped to $-x_0$ well after executing the horseshoe turn. The inner vortensity ring is then attributed to the outer spiral shock, and similarly the outer ring results from the inner spiral shock (see Fig. 2.1). Particles with $x_0 > x_s$ are assumed to circulate after the shock.

Both librating and circulating particles can be re-shocked so the vortensity peaks and troughs are expected to increase in amplitude to some extent. Eventually after a few returns the system approaches a steady state (Fig. 2.2), implying $\partial u_{\perp}/\partial S \rightarrow 0$ (recall from Eq. 2.12 that the vortensity jump is mainly attributed to this term). However, the shock front would have changed since what was post-shock becomes pre-shock, as material returns to the shock. This detail shall be neglected and only the initial ring formation as a result of the first shocking will be considered below.

The post-shock vortensity of a particle originating from $x_0 > 0, y > 0$ and which

2. VORTEX-INDUCED TYPE III MIGRATION

shocks at x is

$$\left(\frac{\omega}{\Sigma}\right)_x = (\eta_0)_{x_0} + \left[\frac{\omega}{\Sigma}\right]_x. \quad (2.13)$$

This new vortensity value is conserved following the particle after it passes through the shock. If the origin of the particle satisfied $x_0 < x_s$ it will be located at $(-x_0, |y/r_h| \gg 1)$ well after it executes the horseshoe turn. If $x_0 > x_s$, it continues to circulate and will be located at $(x_0, y < 0)$. Thus,

$$\left(\frac{\omega}{\Sigma}\right)_x = \begin{cases} (\omega/\Sigma)_{-x_0} & x_0 < x_s \\ (\omega/\Sigma)_{x_0} & x_0 > x_s. \end{cases} \quad (2.14)$$

A similar argument is applied to the inner shock. The particle position is then transformed to the global cylindrical co-ordinates using $r_0 = x_0 + r_p$.

Fig. 2.7 compares this procedure to simulation results. The snapshot is chosen at the point when the rings first occupy the complete azimuth. The central region correspond to particles not crossing the shock so their vortensity is unmodified. However, strictly speaking the vortensity distribution still changes because particles switch between $r_p \pm x_0$ conserving their vortensity and the unperturbed vortensity distribution is non-uniform in the global disc. This is a minor effect that is expected to lead to a uniform distribution due to phase mixing. To deal with this, the vortensity everywhere in this region is simply replaced by the unperturbed vortensity at the planet's location $r_p^{-3/2}/(2\Sigma)$.

The predicted profile in Fig. 2.7 leads to the expectation of sharp vortensity rings adjacent to the planet, with troughs exterior to them. In comparison to simulation data, although the semi-analytic model reproduce the essential features (e.g. vortensity peaks and troughs with the correct relative positions), the predicted distribution is slightly shifted towards the planet (by $0.2r_h$). The model does not reproduce the troughs very well due to inadequate modelling of the pre-shock flow and the shock position away from the co-orbital region.

The modelling shows that it is possible to predict the vortensity ring profile induced by a giant planet, from first principles. When combined with centrifugal balance, the corresponding equilibrium surface density profile is a gap (see §2.4.1). This is one method to calculate gap profiles induced by giant planets in low viscosity discs.

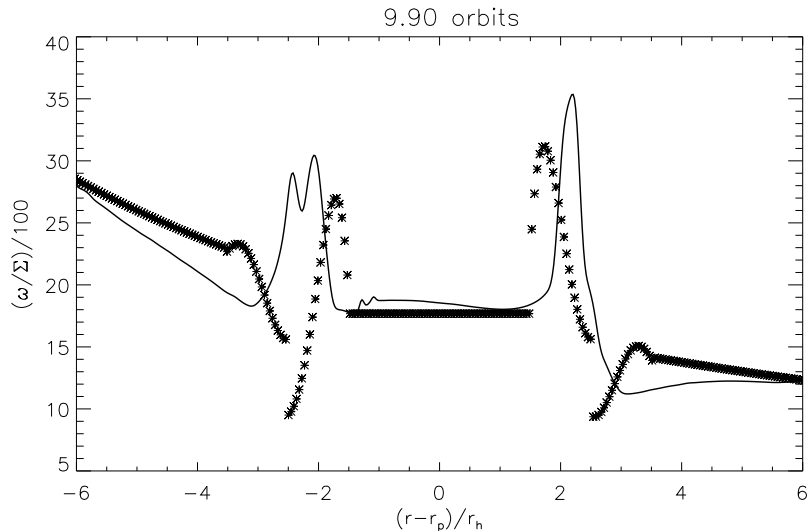


Figure 2.7: Vortensity rings constructed using the modelling from §2.3.1—2.3.3 (asterisks). The solid line is the azimuthally averaged vortensity profile taken from a hydrodynamic simulation. The double peak in the solid line, near the inner vortensity ring, shows that fluid elements do not return to exactly the same radius after being shocked. This is because the snapshot is taken early on in the simulation, during gap-opening by the planet, where material is pushed away from the planet.

2.4 Dynamical stability of vortensity rings

Now that the origin of vortensity rings is established and understood, one can proceed to a linear stability analysis on the shock-modified protoplanetary disc models described above. This is an important issue as instability can lead to their breaking up into mobile non-axisymmetric structures, or vortices, which can affect the migration of the planet significantly.

The analysis applies to a two-dimensional disc and locally isothermal perturbations, for consistency with hydrodynamic simulations presented above and later on. The governing equations are Eq. 1.19 without viscous and potential terms (Euler equations). The planet potential is assumed to contribute to the background gap profile only.

2.4.1 Basic background state

In order to perform a linear analysis, one needs to define an appropriate background equilibrium structure to perturb. The basic state should be axisymmetric and time independent with no radial velocity ($\partial/\partial\phi = \partial/\partial t = u_r = 0$). Suppose the vortensity profile $\eta(r)$ is known (e.g. via shock modelling as above). Recall the angular velocity

2. VORTEX-INDUCED TYPE III MIGRATION

is related to the vortensity by

$$\eta(r) = \frac{1}{r\Sigma} \frac{d}{dr} (r^2\Omega),$$

and the radial momentum equation gives

$$r\Omega^2 = \frac{1}{\Sigma} \frac{dp}{dr} + \frac{GM_*}{r^2}. \quad (2.15)$$

Note that the planet potential is absent in Eq. 2.15. The only role of the planet is to generate vortensity rings, after which the fluid is assumed to be in centrifugal balance with stellar gravity and pressure gradients. This is precisely the assumption to be validated. Other source terms in the momentum equation, e.g. viscosity, are neglected for simplicity.

Using these together with the locally isothermal equation of state, $p = h^2 GM_* \Sigma / r$, a single equation for Σ can be derived:

$$h^2 \frac{d^2 \ln \Sigma}{dr^2} = \frac{h^2 - 1}{r^2} + \frac{2\Sigma\eta}{\sqrt{GM_*}} \left(\frac{1 - h^2}{r} + h^2 \frac{d \ln \Sigma}{dr} \right)^{1/2} - \frac{2h^2}{r} \frac{d \ln \Sigma}{dr}. \quad (2.16)$$

Eq. 2.16 is solved to obtain Σ , with $\eta(r)$ taken as an azimuthal average from the fiducial simulation described in §2.2.1, at a time at which vortensity rings have developed. These structures are essentially axisymmetric apart from in the close neighbourhood of the planet. They are illustrated in Fig. 2.8. The boundary condition is $\Sigma = 10^{-4}$ (the unperturbed uniform density value) at $r = 1.1$ and $r = 3$. These conditions are consistent with the fact that shock-modification of the surface density profile only occurs near the planet ($r_p = 2$).

A comparison between surface density and angular velocity profiles obtained by solving Eq. 2.16 with those obtained as azimuthal averages from the corresponding simulation is made in Fig. 2.9(a) and Fig. 2.9(b). The agreement is generally very good indicating that the adoption of the basic axisymmetric state defined by the simulation vortensity profile, together with centrifugal balance (Eq. 2.15), for stability analysis should be a valid procedure. This validates neglecting the planet potential, even though it was responsible for ring generation.

Fig. 2.9(a) shows that vortensity rings reside in the horseshoe region just inside the

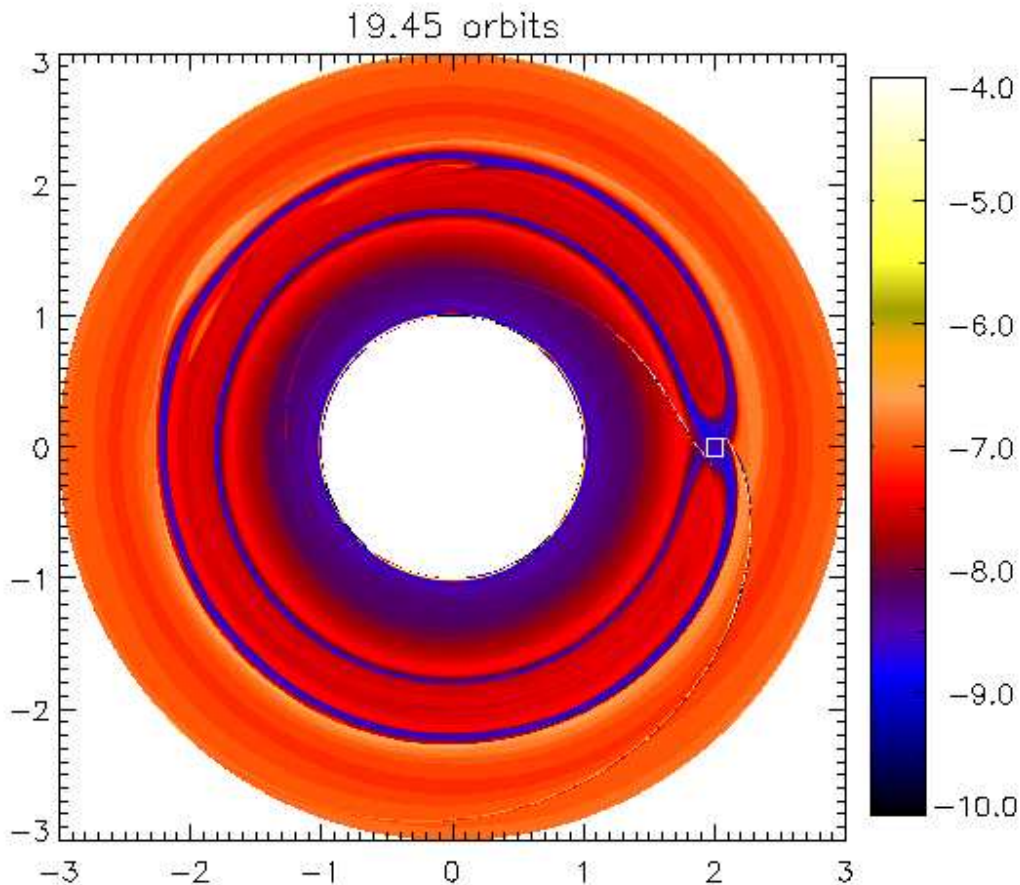


Figure 2.8: Vortensity rings in the co-orbital region. $\ln(\Sigma/\omega)$ is plotted, so the rings correspond to small values (blue). The centre of the small square marks the planet position. This snapshot is taken from a high-resolution ($N_r \times N_\phi = 1024 \times 3072$) inviscid simulation.

gap. At a surface density extrema Eq. 2.16 implies that

$$h^2 \frac{d^2 \ln \Sigma}{dr^2} = \frac{h^2 - 1}{r^2} + 2\Sigma\eta \sqrt{\frac{1 - h^2}{GM_* r}}.$$

Since $h \ll 1$, if η is sufficiently large then $d^2 \Sigma / dr^2 > 0$. Hence vortensity maxima nearly coincides with surface density minima. In Fig. 2.9(a), local vortensity minima and maxima are separated by $O(H)$ ($\simeq 0.1$ in dimensionless units). As vortensity rings originate from spiral shocks, Eq. 2.16 demonstrates the link between shocks and gap formation. Given the double-ringed vortensity profile $\eta(r)$, which may be estimated by

2. VORTEX-INDUCED TYPE III MIGRATION

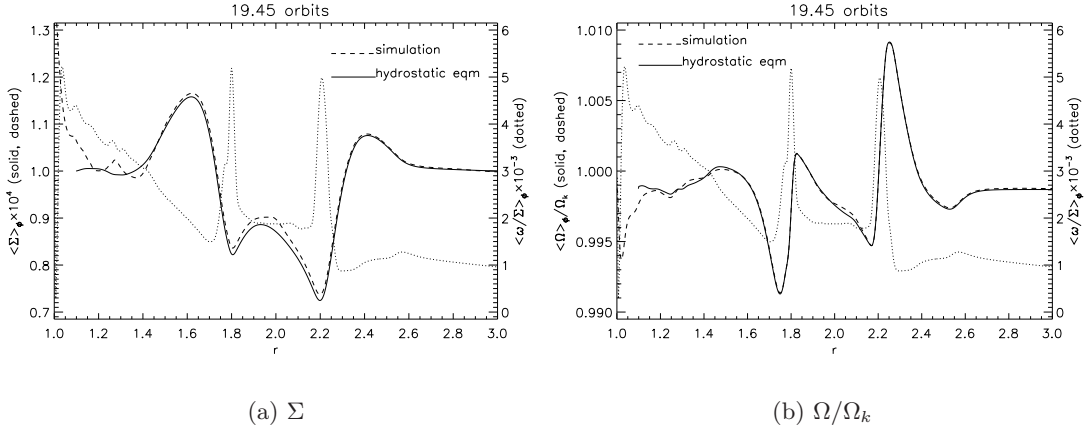


Figure 2.9: A comparison of the axisymmetric surface density and angular velocity (scaled by Keplerian speed) profiles obtained by solving Eq. 2.16 (solid lines), with the same quantities obtained by azimuthally averaging simulation data (dashed lines). The azimuthally averaged vortensity profile from simulation data, used as input in solving Eq. 2.16, is also shown (dotted). The co-orbital region of the planet is $r = [1.77, 2.22]$.

modelling shocks, one can solve Eq. 2.16 for the axisymmetric surface density profile $\Sigma(r)$. One then finds that in order for the rings to be in hydrostatic equilibrium, a gap in the surface density must be present around the planet's orbital radius, which is in between the vortensity peaks. Sufficiently massive planets induce strong shocks, implying larger vortensity maxima, and hence deeper surface density minima or gaps.

2.4.2 Linearised equations

Having established the ring-basic state, a linear analysis can now be performed to determine its stability. The hydrodynamic variables are written as

$$\begin{pmatrix} \Sigma \\ u_r \\ u_\phi \end{pmatrix} = \begin{pmatrix} \Sigma \\ 0 \\ u_\phi \end{pmatrix} + \begin{pmatrix} \delta \Sigma(r) \\ \delta u_r(r) \\ \delta u_\phi(r) \end{pmatrix} \times \exp i(\sigma t + m\phi),$$

where the first term on the RHS corresponds to the basic background state, δ denote small Eulerian perturbations, σ is a complex frequency and m , the azimuthal mode

2.4 Dynamical stability of vortensity rings

number, is a positive integer. The linearised equation of motion gives

$$\delta u_r = -\frac{c_s^2}{\kappa^2 - \bar{\sigma}^2} \left(i\bar{\sigma} \frac{dW}{dr} + \frac{2im\Omega W}{r} \right) \quad (2.17)$$

$$\delta u_\phi = \frac{c_s^2}{\kappa^2 - \bar{\sigma}^2} \left(\Sigma \eta \frac{dW}{dr} + \frac{m\bar{\sigma}}{r} W \right) \quad (2.18)$$

where $W \equiv \delta\Sigma/\Sigma$ is the relative surface density perturbation, recall $\kappa^2 = 2\Omega\Sigma\eta$ is the epicycle frequency expressed in terms of the vortensity η , and $\bar{\sigma} \equiv \sigma + m\Omega(r) = \sigma_R + m\Omega(r) + i\gamma$ is the Doppler-shifted frequency with σ_R and γ being real. In this notation, $\gamma < 0$ corresponds to an exponential growth of perturbations and therefore instability. Substituting Eq. 2.17–2.18 into the linearised continuity equation

$$i\bar{\sigma}W = -\frac{1}{r\Sigma} \frac{d}{dr} (r\Sigma\delta u_r) - \frac{im}{r} \delta u_\phi, \quad (2.19)$$

yields a governing equation for W of the form

$$\frac{d}{dr} \left(\frac{rc_s^2\Sigma}{\kappa^2 - \bar{\sigma}^2} \frac{dW}{dr} \right) + \left\{ \frac{m}{\bar{\sigma}} \frac{d}{dr} \left[\frac{c_s^2\kappa^2}{\eta(\kappa^2 - \bar{\sigma}^2)} \right] - r\Sigma - \frac{c_s^2 m^2 \Sigma}{r(\kappa^2 - \bar{\sigma}^2)} \right\} W = 0. \quad (2.20)$$

Specifically, for the locally isothermal equation of state, this is

$$\frac{d}{dr} \left(\frac{\Sigma}{\kappa^2 - \bar{\sigma}^2} \frac{dW}{dr} \right) + \left\{ \frac{m}{\bar{\sigma}} \frac{d}{dr} \left[\frac{\kappa^2}{r\eta(\kappa^2 - \bar{\sigma}^2)} \right] - \frac{r\Sigma}{GM_* h^2} - \frac{m^2 \Sigma}{r^2(\kappa^2 - \bar{\sigma}^2)} \right\} W = 0. \quad (2.21)$$

Eq. 2.21 is an eigenvalue problem for the complex eigenvalue σ . The co-rotation radius r_c is where $\sigma_R + m\Omega(r_c) = 0$. Co-rotation resonance occurs at $\bar{\sigma} = 0$ which requires $\gamma = 0$ for it to be on the real axis. Then for the equation to remain regular, the gradient of the terms in square brackets must vanish at co-rotation. This results in the requirement that the gradient of ηr vanish there. Because the sound speed varies with radius, this is slightly different from the condition that the gradient of η should vanish which applies in the barotropic or strictly isothermal case (Papaloizou & Pringle, 1984, 1985; Papaloizou & Lin, 1989). However, because η varies rapidly in the region of interest and the modes of interest are locally confined in radius, this difference is of no essential consequence. Lindblad resonances occur when $\kappa^2 - \bar{\sigma}^2 = 0$, but as is well known, and can be seen from formulating a governing equation for δu_r rather than W , these do not result in a singularity.

2. VORTEX-INDUCED TYPE III MIGRATION

2.4.3 Simplification of the governing ODE

It is useful to simplify Eq. 2.21 to gain further insight. To do this, consider modes localised around the co-rotation circle such that the condition $\kappa^2 \gg |\bar{\sigma}^2|$ is satisfied. Beyond this region the mode amplitude is presumed to be exponentially small. Now, the ratio of the second to last to last term in Eq. 2.20 is

$$\frac{r^2 \kappa^2}{m^2 c_s^2} \sim \frac{1}{m^2 h^2}.$$

For a thin disc $h \ll 1$, so for $m = O(1)$ this ratio is large and the last term in Eq. 2.21 can be neglected. This is also motivated by the fact that only low m modes are observed in simulations. Doing this and replacing $\kappa^2 - \bar{\sigma}^2$ by κ^2 , Eq. 2.20 reduces to the simplified form

$$\frac{d}{dr} \left(\frac{r c_s^2 \Sigma}{\kappa^2} \frac{dW}{dr} \right) + \left\{ \frac{m}{\bar{\sigma}} \frac{d}{dr} \left[\frac{c_s^2}{\eta} \right] - r \Sigma \right\} W = 0, \quad (2.22)$$

which is valid for any fixed c_s profile.

Localised modes described by the approximations above have been called ‘co-rotation modes’ (Lovelace et al., 1999). Suppose the basic state has a localised vortensity extremum at r_c . Then far away from r_c , vortensity gradients are small and the $-r\Sigma W$ term in Eq. 2.22 dominates, implying exponentially decaying solutions. Conversely, near r_c vortensity gradients are large and the term proportional to $1/\bar{\sigma}$ dominates. Evaluating Eq. 2.22 for a neutral mode ($\gamma = 0$) at r_c implies the balance

$$\left[\left(\frac{r c_s^2 \Sigma}{\kappa^2} \right)' \frac{dW}{dr} \right]_{r_c} + \left[\left(\frac{r c_s^2 \Sigma}{\kappa^2} \right) \frac{d^2 W}{dr^2} \right]_{r_c} + \left[\frac{1}{\Omega'} \left(\frac{c_s^2}{\eta} \right)'' \right]_{r_c} W(r_c) \sim 0,$$

where $'$ denotes d/dr , and $\bar{\sigma}' = m\Omega'$ has been used. Assuming the mode is symmetric about r_c so that $W'(r_c) = 0$, one deduces

$$\frac{W''}{W} \Big|_{r_c} \sim \left[\frac{-2\Omega\eta}{r c_s^2 \Omega'} \left(\frac{c_s^2}{\eta} \right)'' \right]_{r_c}, \quad (2.23)$$

where $\kappa^2 = 2\Omega\Sigma\eta$ has been used. Eq. 2.23 says that the degree of localisation, measured by $W''(r_c)$, is proportional to $(c_s^2/\eta)''|_{r_c}$, so the more sharply peaked the vortensity extremum is, the more localised the mode will be.

2.4.4 Necessity of extrema

Vortensity extrema must exist for unstable modes. Multiplying Eq. 2.22 by W^* and integrating between $[r_1, r_2]$ assuming, consistent with a sharp exponential decay, that $W = 0$ or $dW/dr = 0$ at these boundaries, yields

$$\int_{r_1}^{r_2} \frac{m}{\bar{\sigma}} \left(\frac{c_s^2}{\eta} \right)' |W|^2 dr = \int_{r_1}^{r_2} r \Sigma |W|^2 dr + \int_{r_1}^{r_2} \frac{r \Sigma c_s^2}{\kappa^2} |W'|^2 dr \quad (2.24)$$

Since the RHS is real, the imaginary part of the LHS must vanish. For general complex σ this implies that

$$\gamma \int_{r_1}^{r_2} \frac{m}{(\sigma_R + m\Omega)^2 + \gamma^2} \left(\frac{c_s^2}{\eta} \right)' |W|^2 dr = 0. \quad (2.25)$$

Thus for a growing mode ($\gamma \neq 0$) to exist, one requires $(c_s^2/\eta)' = 0$ at some r in $[r_1, r_2]$. For the locally isothermal equation of state, c_s^2 varies on a scale $O(r)$, but η varies on a scale $H \ll r$, thus given that the range of relative variation of the vortensity η is of order unity, one infers that η needs to have stationary points in order for there to be unstable modes. Referring back to Fig. 2.9(a) it is clear that the basic state being considered satisfies the necessary criterion for instability.

2.4.5 Limit on the growth rate

The approximation for co-rotational modes, $\kappa^2 \gg |\bar{\sigma}^2|$, implicitly assumes the growth rate $|\gamma|/\Omega \ll 1$, since $\kappa \sim \Omega$. A more quantitative estimate of the upper limit to the growth rate can be used to provide bounds on initial guess solutions when solving the eigenvalue problem numerically. Let

$$W = g\bar{\sigma}, \quad (2.26)$$

Eq. 2.22 then gives an alternative governing equation:

$$\begin{aligned} \left(\frac{r \Sigma c_s^2 \bar{\sigma}^2}{\kappa^2} g' \right)' &= \bar{\sigma} \left[r \Sigma \bar{\sigma} - m \left(\frac{c_s^2}{\eta} + \frac{r \Sigma c_s^2 \Omega'}{\kappa^2} \right)' \right] g \\ &= \bar{\sigma} \left[r \Sigma \bar{\sigma} - m \left(\frac{c_s^2 \Sigma}{2\Omega} \right)' \right] g, \end{aligned} \quad (2.27)$$

2. VORTEX-INDUCED TYPE III MIGRATION

where the second line is simplified from the first line by using $\kappa^2 = (1/r^3)d(r^4\Omega^2)/dr = 2\Omega\Sigma\eta$. Multiplying Eq. 2.27 by g^* , integrating and neglecting boundary terms gives

$$-\int_{r_1}^{r_2} \frac{r\Sigma c_s^2 \bar{\sigma}^2}{\kappa^2} |g'|^2 dr = \int_{r_1}^{r_2} \bar{\sigma} \left[r\Sigma\bar{\sigma} - m \left(\frac{c_s^2 \Sigma}{2\Omega} \right)' \right] |g|^2 dr. \quad (2.28)$$

The real part of Eq. 2.28 is

$$\gamma^2 \int_{r_1}^{r_2} r\Sigma \left(\frac{c_s^2}{\kappa^2} |g'|^2 + |g|^2 \right) dr = \int_{r_1}^{r_2} \left[r\Sigma\bar{\sigma}_R^2 \left(\frac{c_s^2}{\kappa^2} |g'|^2 + |g|^2 \right) - m\bar{\sigma}_R \left(\frac{c_s^2 \Sigma}{2\Omega} \right)' |g|^2 \right] dr, \quad (2.29)$$

where $\bar{\sigma}_R = \sigma_R + m\Omega$. The imaginary part of Eq. 2.28 implies (assuming $\gamma \neq 0$),

$$\int_{r_1}^{r_2} \left[2r\Sigma\bar{\sigma}_R \left(\frac{c_s^2}{\kappa^2} |g'|^2 + |g|^2 \right) - m \left(\frac{c_s^2 \Sigma}{2\Omega} \right)' |g|^2 \right] dr = 0. \quad (2.30)$$

If the second term of the integrand in Eq. 2.30 is negligible (justified below), Eq. 2.30 reduces to

$$\int_{r_1}^{r_2} \left[2r\Sigma(\sigma_R + m\Omega) \left(\frac{c_s^2}{\kappa^2} |g'|^2 + |g|^2 \right) \right] dr = 0, \quad (2.31)$$

which implies that $\sigma_R + m\Omega = 0$ somewhere in $[r_1, r_2]$, since the other terms are non-negative. This confirms the co-rotation point r_c , such that $\sigma_R = -m\Omega(r_c)$, is inside the domain.

Eq. 2.29 and Eq. 2.30 can be combined to give:

$$\int_{r_1}^{r_2} r\Sigma(\gamma^2 + \sigma_R^2) \left(\frac{c_s^2}{\kappa^2} |g'|^2 + |g|^2 \right) dr = \int_{r_1}^{r_2} m^2\Omega^2 r\Sigma \left(\frac{c_s^2}{\kappa^2} |g'|^2 + |g|^2 \right) dr - \int_{r_1}^{r_2} m^2\Omega \left(\frac{c_s^2 \Sigma}{2\Omega} \right)' |g|^2 dr. \quad (2.32)$$

Note that in Eq. 2.32, the LHS and the first term on RHS are non-negative. The second integral on RHS, involving $(c_s^2 \Sigma / 2\Omega)'$, may be positive or negative. Far away from the gap edge the disc is smoothly varying and gradients are small. Near the gap edges, the disc varies rapidly over local scale-heights. One can expect the integral involving $(c_s^2 \Sigma / 2\Omega)'$ makes a small contribution in Eq. 2.32, due to small gradients and internal cancellations when the integral is performed. One can also compare coefficients of $|g|^2$

2.4 Dynamical stability of vortensity rings

in the integrands on RHS of Eq. 2.32:

$$\tau \equiv \left| \frac{m^2 \Omega^2 r \Sigma}{m^2 \Omega \left(\frac{c_s^2 \Sigma}{2\Omega} \right)'} \right|. \quad (2.33)$$

If the disc varies on a length-scale $L \sim O(r)$ then $\tau = O(h^{-2})$, whereas if $L \sim O(H)$ then $\tau = O(h^{-1})$. In either case, the ratio is large for thin discs ($h \ll 1$). Inserting the actual disc profiles obtained from numerical simulation gives $\tau \sim 90$. The last term on RHS of Eq. 2.32 can be safely neglected. This also justifies Eq. 2.31.

Next, consider the inequality

$$0 \geq m^2 \int_{r_1}^{r_2} (\Omega - \Omega_+) (\Omega - \Omega_-) r \Sigma \left(\frac{c_s^2}{\kappa^2} |g'|^2 + |g|^2 \right) dr, \quad (2.34)$$

where Ω_+ , Ω_- are the maximum and minimum angular speeds in the range of integration. Using Eq. 2.32 together with Eq. 2.30, and neglecting the integrals involving $(c_s^2 \Sigma / 2\Omega)'$ in both equations, one can show that

$$\gamma^2 + \left[\sigma_R + \frac{1}{2} m (\Omega_+ + \Omega_-) \right]^2 \leq m^2 \left(\frac{\Omega_+ - \Omega_-}{2} \right)^2. \quad (2.35)$$

The complex eigenfrequency σ is contained inside a circle centred on the real axis at $-m(\Omega_+ + \Omega_-)/2$ with radius $m(\Omega_+ - \Omega_-)/2$. Furthermore, if the mode is localised about r_c with characteristic width $\Delta > 0$, so the range of integration may be taken as $r_1 = r_c - \Delta$, $r_2 = r_c + \Delta$, and the angular speed is monotonically decreasing then

$$\Omega_{\pm} \simeq \Omega(r_c) \mp \Omega'(r_c) \Delta.$$

The term in square brackets in Eq. 2.35 is then $\sigma_R + m\Omega(r_c)$, which is zero by the definition of r_c . This means the growth rate is limited to

$$\gamma^2 \leq (m\Omega'(r_c)\Delta)^2. \quad (2.36)$$

The growth of co-rotational modes is limited by local shear. Strong shear may enable higher growth rates. Inserting Keplerian shear and $\Delta = H(r_c) = hr_c$ in Eq. 2.36 gives

$$\frac{|\gamma|}{\Omega(r_c)} \leq \frac{3}{2} hm. \quad (2.37)$$

2. VORTEX-INDUCED TYPE III MIGRATION

For $h \ll 1$ and $m = O(1)$, growth rates are at most a fraction of the local angular speed. The growth timescale is $|\gamma|^{-1}$. When measured in Keplerian orbital periods at r_p , the growth time is

$$\frac{|\gamma|^{-1}}{P_0} \geq \frac{1}{3\pi hm} \left(\frac{r_p}{r_c}\right)^{-3/2},$$

where Keplerian rotation has been used. Vortex formation will be shown to be associated with vortensity minimum at gap edges. For the reference case later (Fig. 2.9(a)), the vortensity minimum close to the inner gap edge is located at $r_c = 1.7$. The planet is at $r_p = 2$ and the aspect-ratio is $h = 0.05$. Inserting these for the $m = 3$ mode in Eq. 2.37 yields a growth time $|\gamma|^{-1} \gtrsim 0.6P_0$, so the instability can grow on dynamical timescales.

2.4.6 Numerical solution of the eigenvalue problem

The eigenvalue problem Eq. 2.21 is solved using a shooting method that employs an adaptive Runge-Kutta integrator and a multi-dimensional Newton method (Press et al., 1992).

The shooting method works as follows, beginning with a trial value for σ . The previous analysis (and nonlinear simulations described later) hints that a good guess would be $\sigma_R = -m\Omega(r_0)$, where r_0 is a vortensity minimum. The imaginary part of σ should be a small fraction of $\Omega(r_0)$ and negative for unstable modes. The governing equation is then evolved as an initial value problem, from the inner boundary (with imposed boundary conditions) to the outer boundary. At the outer boundary, one compares the evolved solution to the required outer boundary condition. The eigenvalue is accordingly adjusted, using the Newton method, and the initial value problem solved again to obtain a new solution with outer boundary values closer to the desired conditions. The process is repeated until convergence.

For low $m (\leq 3)$, unstable modes mainly comprise an evanescent disturbance near co-rotation (vortensity minimum) and the simple boundary condition $W = 0$ applied at the inner boundary $r/r_p = 0.55$ and the outer boundary $r/r_p = 1.5$ produces good results. As m increases, the Lindblad resonances (where $\kappa^2 = \bar{\sigma}^2$) approach $r = r_0$ and a significant portion of the mode is wave-like requiring the application of outgoing radiation boundary conditions. These are determined using the WKBJ approximation (see eg. Korycansky & Papaloizou (1995)). Recognising the governing equation (Eq.

2.21) as

$$\frac{d}{dr} \left(A \frac{dW}{dr} \right) = -BW,$$

the WKBJ solution is given by

$$W \propto \frac{1}{(AB)^{1/4}} \exp \pm i \int \sqrt{\frac{B}{A}} dr.$$

However, high m , wave-like modes are in fact irrelevant because these are not the fastest growing modes.

2.4.7 Fiducial case

Some example solutions are presented to illustrate the instability of gap edges. Hydrodynamic simulations indicate the ultimate dominance of small m values. One class of mode is associated with the inner vortensity ring while another is associated with the outer ring.

As a typical example of the behaviour that is found for low m , each type of eigenmode for $h = 0.05$ and $m = 3$ is shown in Fig. 2.10. The background vortensity profile is also shown. The instabilities ($\gamma < 0$) are associated with the vortensity minima at the inner and outer gap edge, as has also been observed by Li et al. (2005) in simulations. The modes are evanescent around co-rotation and the vortensity peaks behave like walls through which the instability scarcely penetrates. The mode decays away from $r = r_0$. For $m = 3$, Lindblad resonances occur at $r_L/r_0 = 0.76, 1.21$ from which waves travelling away from co-rotation are emitted. However, the oscillatory amplitude is at most $\simeq 20\%$ of that at $r = r_0$. Hence for low- m the dominant effect of the instability will be due to perturbations near co-rotation. Increasing m brings r_L even closer to r_0 , waves then propagate through the planetary gap. The growth timescale of the inner mode with $m = 3$ is $\sim 14P_0$. The outer mode has a growth rate that is about three times faster. Since the instability grows on dynamical timescales, non-linear interaction of vortices are expected to occur within few tens of orbits, and to affect planet migration if migration were also on similar or longer timescales.

After the onset of linear instability and the formation of several vortices, it has been observed that non-linear effects cause them to eventually merge into a single vortex (de Val-Borro et al., 2007) which interacts with the planet. In the fiducial simulation to be presented, vortex-induced rapid migration begins within $55P_0$ which is compatible

2. VORTEX-INDUCED TYPE III MIGRATION

with the characteristic growth times found from linear theory.

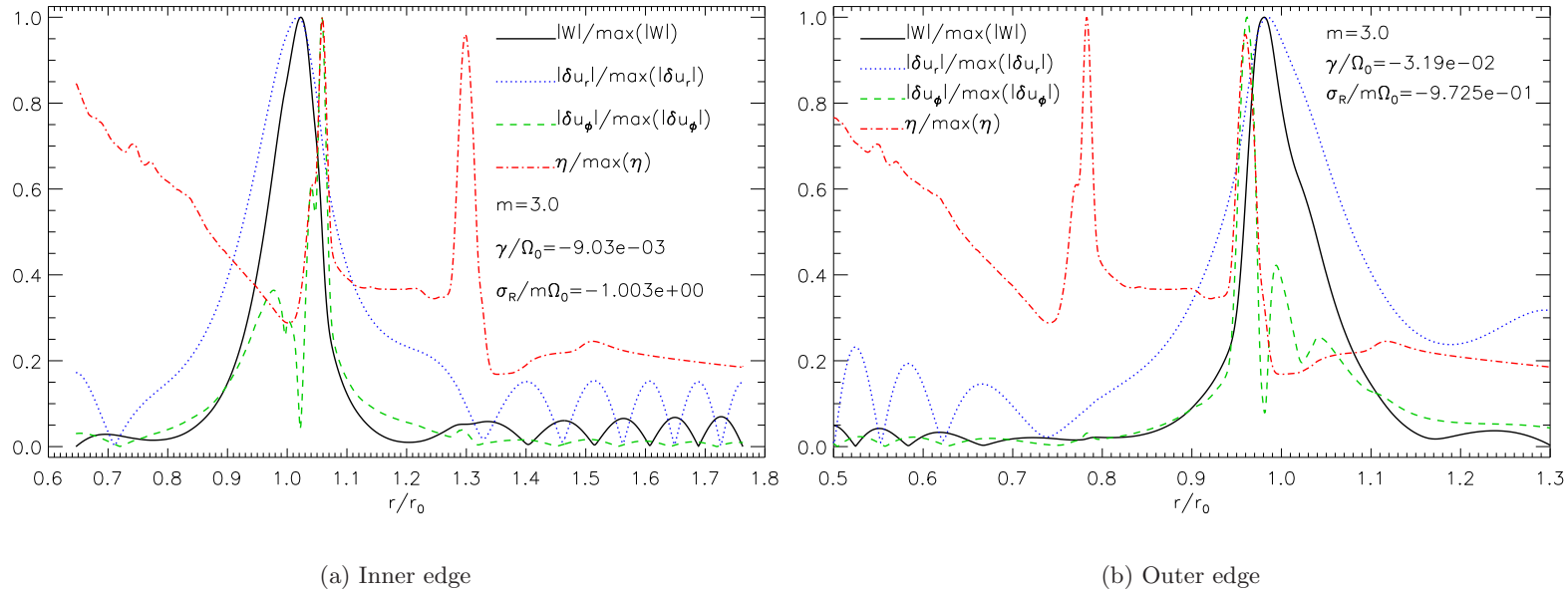


Figure 2.10: $m = 3$ eigenmodes obtained with the boundary condition $W = 0$. The perturbed quantities are plotted as $|W|$ (solid, black), $|\delta u_r|$ (dotted, blue), $|\delta u_\phi|$ (dashed, green) and scaled by their maximum values in $r = [1.1, 3.0]$. The background vortensity profile is also shown (dash-dot, red). $r_0 = 1.7, 2.3$ correspond to the inner and outer vortensity minima, respectively.

2. VORTEX-INDUCED TYPE III MIGRATION

Fig. 2.11 shows the eigenfunctions for the $m = 7$ mode of the outer ring. The equivalent mode was not found for the inner edge because high- m are quenched. Radiative boundary conditions were adopted in this case. Although the WKB condition is the appropriate physical boundary condition, its application here is uncertain because the boundaries cannot be considered ‘far’ from the gap edge. However, solutions are actually not sensitive to boundary conditions as reported by de Val-Borro et al. (2007). Note the two spikes in δu_r and δu_ϕ at $r/r_0 \simeq 0.90, 1.09$ which correspond to Lindblad resonances. These are not singularities as can be seen from $W(r)$ which is smooth there; other eigenfunctions were calculated from the numerical solution for W and thus may be subject to numerical errors. Increasing m increases the amplitude in the wave-like regions of the mode, but the growth rate is smaller than for $m = 3$. As the instability operates on dynamical timescales, low- m modes will dominate over high- m modes, particularly through non-linear evolution and interaction of the former.

2.4.8 Dependence on h

The aspect ratio h , equivalent to disc temperature, can affect stability properties. Lowering h increases shock strength and therefore the magnitude of vortensity jump, but does not affect ring locations and widths. The basic state formed with lower h should be more unstable because gradients are larger in magnitude. h is also explicit in Eq. 2.21, appearing as $-r\Sigma W/(GM_*h^2)$. Recalling from §2.4.3 that this term is responsible for exponential decay, then decreasing h should increase localisation. So lowering h should favour development of co-rotational modes.

In the discussion below, $h = 0.05$ will be abbreviated as h_5 and similarly for other values. The linear problem was solved for h_3 – h_6 . In each case, a disc-planet simulation was performed to generate the basic state. Table 2.1 compare growth rates of $m = 1$ – 3 modes with $W = 0$ boundary condition. Localised modes as those in Fig. 2.10 were found, except for the inner edge for h_6 where no unstable modes were found. This means that only sufficiently extreme profiles support co-rotational modes. As h is lowered, the inner edge becomes more unstable with $|\gamma_{\text{in}}|(h_3) \simeq (1.5 \text{ --} 2) \times |\gamma_{\text{in}}|(h_5)$.

Generally, lowering h is destabilising. The exception to this trend is the outer edge of h_3 , being more stable than h_4 and h_5 . This is an artifact from azimuthally averaging the simulation data to generate the background profile. Hydrodynamic simulations for h_3 indicate, that vortensity rings actually become unstable *before* reaching the full azimuth. By the time the rings reach the full 2π , vortices already appear adjacent to vortensity rings. Radial variations are smoothed during an azimuthal average because

2.4 Dynamical stability of vortensity rings

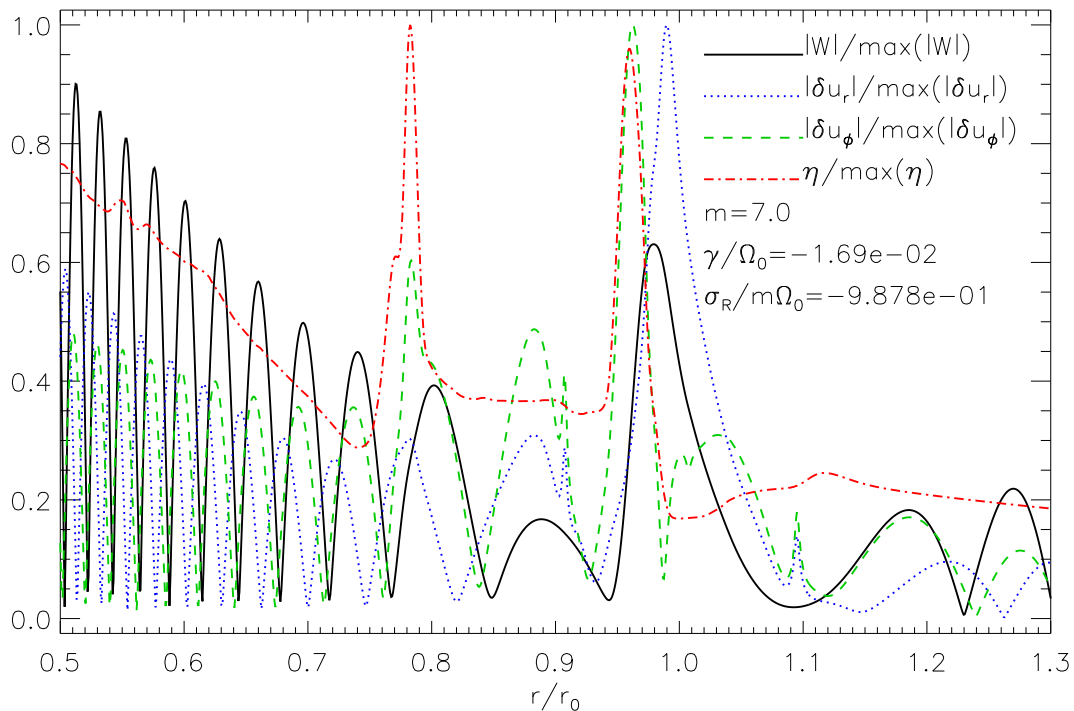


Figure 2.11: $m = 7$ eigenmodes associated with the outer edge, obtained with WKB boundary condition. The perturbed quantities are plotted as $|W|$ (solid, black), $|\delta u_r|$ (dotted, blue), $|\delta u_\phi|$ (dashed, green) and scaled by their maximum values in $r = [1.1, 3.0]$. The background vortensity profile is also shown (dash-dot, red). $r_0 = 2.3$ is the radius of the outer vortensity minima.

Table 2.1: Growth rates of linear modes scaled by local angular speeds for $m = 1-3$ for the inner (subscript ‘i’) outer (subscript ‘o’) gap edges.

h	$m = 1$		$m = 2$		$m = 3$	
	$-10^3 \gamma_i / \Omega_i$	$-10^3 \gamma_o / \Omega_o$	$-10^3 \gamma_i / \Omega_i$	$-10^3 \gamma_o / \Omega_o$	$-10^3 \gamma_i / \Omega_i$	$-10^3 \gamma_o / \Omega_o$
0.03	8.405	8.466	15.72	16.14	20.82	22.33
0.04	7.804	13.51	14.21	25.62	18.20	35.01
0.05	5.586	13.42	9.298	24.79	9.090	31.95
0.06	0.00287	4.304	0.00437	6.526	0.00233	5.741

2. VORTEX-INDUCED TYPE III MIGRATION

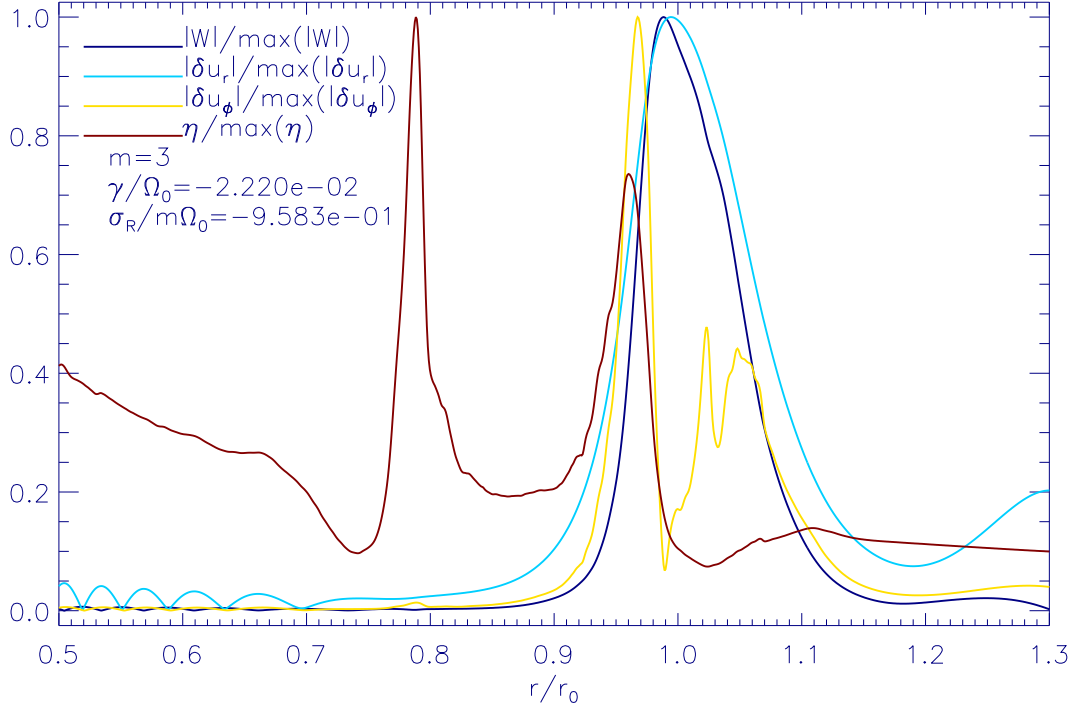


Figure 2.12: Eigenmodes for $m = 3$, $h = 0.03$ associated with the outer vortensity ring. The horizontal axis is scaled by the radius of local vortensity minimum. The boundary condition $W = 0$ was applied in this case.

vortex/ring are separated by a curve $r = r(\phi)$.

Fig. 2.12 show solutions and the background vortensity profile for h_3 . As explained above, the outer vortensity ring in h_3 is apparently less sharp compared to h_5 (Fig. 2.10). The outer edge is *apparently* more stable. In this case, the azimuthally averaged vortensity profile is not an appropriate representation of the outer edge, having already become non-axisymmetric. Nevertheless, the solution in Fig. 2.12 indicate more localisation compared to h_5 (Fig. 2.10, where the amplitude of the wave region is relatively larger than h_3). This is expected from the increased importance of the exponential decay term governing the linear response (Eq. 2.21).

Table 2.1 also show the outer edge is typically more unstable, consistent with simulations that vortices first form near the outer ring. There is increasing difference in growth rates between inner and outer minima as h increases. This may be because as h increase, there is less vortensity jump (in magnitude) across weaker shocks. Hence, the background vortensity variation (before the planet is introduced) is more pronounced, so the inner/outer edge asymmetry is enhanced. Conversely, for small

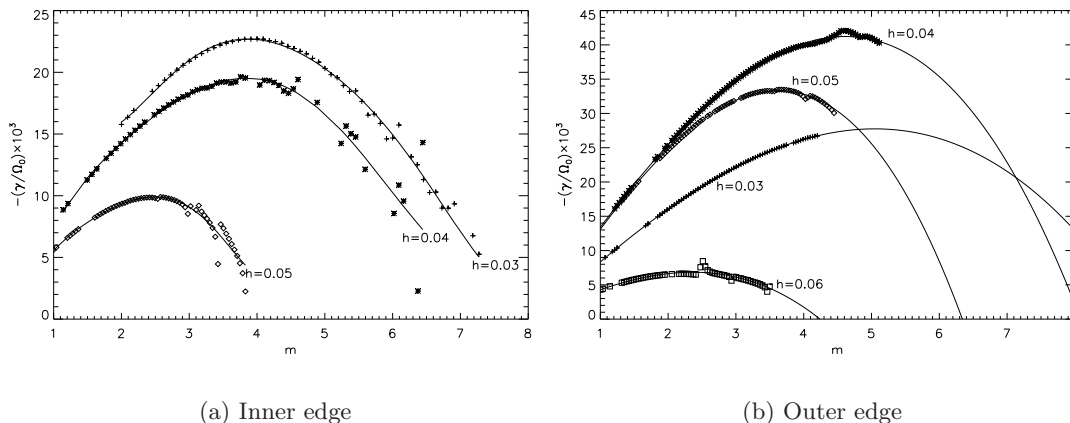


Figure 2.13: Growth rates as a function of azimuthal wave number m and aspect-ratio h . Ω_0 is background angular speed at corresponding vortensity minima. The $h = 0.03$ case for the outer edge is spurious because the disc has already become unstable when the profile is taken. For the inner edge, generally being more stable, modes were not found for $h = 0.06$.

h strong shocks are produced, the final vortensity profile is dominated by vortensity jumps across shocks, and less affected by the background profile.

2.4.9 Growth rates and m

Hydrodynamic simulations will show that low m co-rotational (vortex) modes are most relevant to co-orbital disc-planet interactions. For completeness though, stability as a function of m is briefly discussed. Table 2.1 show $|\gamma|$ typically increase with m . However, rates cannot increase indefinitely. Waves can be stabilising by carrying energy away from co-rotation. The co-rotational disturbance itself has positive energy provided co-rotation is at a vortensity minimum (see Chapter 3). Consequently $|\gamma|$ eventually decrease as m increases. One expects $|\gamma|$ to peak as a function of m (Li et al., 2000; de Val-Borro et al., 2007), and a cut-off at some maximum m .

The linear problem, Eq. 2.21, is solved subject to WKBJ boundary conditions and non-integer m , to map out dependence of γ on azimuthal wave number. Only integer values of m make physical sense, but this is not required to solve Eq. 2.21. Results are shown in Fig. 2.13 with polynomial fits and extrapolation.

Inner edge growth rates in the standard disc h_5 peaks at $m = 2-3$ and extrapolation suggest a cut-off for $m \geq 5$. Decreasing to h_4 , $|\gamma|$ peaks around $m = 4$ with growth rate roughly twice the peak of h_5 . Higher- m modes are enabled as temperature is lowered.

2. VORTEX-INDUCED TYPE III MIGRATION

Lowering h generally destabilise the system while increasing localisation. A similar behaviour was found for the outer edge (except for the spurious h_3).

Boundary effects became apparent as higher m were considered. This is reflected in the sporadic growth rates beyond the most unstable mode (for the inner edge) and the inability to find satisfactory solutions for the outer edge. This is because the boundaries were not very far from gap edges. In these cases, extrapolation to higher m was performed, assuming there is some cut-off at high m . Fortunately, these very high m modes are not the most unstable, nor do they show up in nonlinear simulations.

de Val-Borro et al. (2007) performed a similar analysis for gaps opened by a Jovian-mass planet in a h_5 disc, and obtained growth rates almost an order of magnitude larger than here. de Val-Borro et al. also found peak growth rates around $m = 4$ – 6 . For disc-planet interactions, increasing planetary mass has the same effect as lowering h . Results here are then consistent with that of de Val-Borro et al., that lowering h destabilises the system and shifts the modes to higher m .

2.5 Simulations of fast migration driven by vortex-planet interaction

Disc-planet simulations are now examined, where the planet is free to migrate, so that $r_p = r_p(t)$. Vortices that form near the gap edge move through the co-orbital region and cause torques to be exerted on the planet. The interaction between these vortices and the planet can be interpreted as non-monotonic type III migration. In the original description of type III migration by Masset & Papaloizou (2003), the type III torque increases with the co-orbital mass deficit δm defined as:

$$\delta m = 8\pi r_p B_p \left[x_s w(-x_s) - \int_{-x_s}^0 w(x) dx \right] \quad (2.38)$$

where $w = \Sigma/\omega = \eta^{-1}$, $B_p = \frac{1}{2r} \partial_r (r^2 \Omega)$ is the Oort constant evaluated at the planet radius r_p , and the co-ordinate $x = r - r_p$. Hence the inverse vortensity Σ/ω will be central to the discussion.

Simulations described below have computational domain $r = [0.4, 4.0]$ with resolution $N_r \times N_\phi = 768 \times 2304$ uniformly spaced in both directions. The initial surface density scale is chosen to be $\Sigma_0 = 7.0$, corresponding to a few times the value appropriate to the minimum mass Solar nebula in order to achieve smooth rapid migration when a typical viscosity $\nu_0 = 1$ is used (Masset & Papaloizou, 2003). The initial radial velocity of the disc is set to be $u_r = -3\nu/2r$ as expected for a steady accretion

2.5 Simulations of fast migration driven by vortex-planet interaction

disc (Lynden-Bell & Pringle, 1974), while the initial azimuthal velocity is slightly sub-Keplerian to achieve centrifugal balance with pressure and stellar gravity. The planet is introduced in Keplerian circular orbit. For most of these simulations the full planet potential is applied from $t = 0$, but similar results were obtained when the potential is switched on over $5P_0$. In those cases, vortices were also observed to form. Switching on the planet potential over several orbits does not weaken the instability, and vortex-planet interactions still occur.

Type III migration is numerically challenging due to its potential dependence on flow near the planet, one issue being the numerical resolution. D’Angelo et al. (2005) reported the suppression of type III migration in high resolution simulations. The main migration feature discussed below is brief phases of rapid migration due to vortex-planet interaction, which does not depend on conditions very close to the planet. Lower resolution simulations with $N_r \times N_\phi = 192 \times 576$, 256×768 both show such behaviour, thus the higher resolution described below is sufficient to study this interaction.

2.5.1 Dependence of the migration rate on viscosity

The previous sections showed that ring structures formed by a Saturn-mass planet can be linearly unstable. Other fixed-orbit simulations of disc-planet interactions show that a dimensionless viscosity of order 10^{-5} suppresses instability and subsequent vortex formation (e.g. de Val-Borro et al., 2007, who used a Jupiter-mass planet). As a consequence studies using viscous discs typically yield smooth migration curves. This motivates a study of type III migration as a function of viscosity. In this section the planet mass is fixed to be $M_p = 2.8 \times 10^{-4} M_*$.

Fig. 2.14 shows the orbital semi-major axis $a(t)$ for viscosities $\nu_0 = 0$ to $\nu_0 = 1$. As the orbit is very nearly circular, $a(t)$ is always close to the instantaneous orbital radius $r_p(t)$. A case with $\nu_0 = 10^{-3}$ was considered, which showed almost identical $a(t)$ curve to $\nu_0 = 0$. The numerical viscosity is thus between $10^{-8} \lesssim \nu \lesssim 10^{-6}$ (from test simulations, see also Fig. 2.14). It is expected to be much smaller than the typically adopted physical viscosity of $\nu = 10^{-5}$ in disc-planet simulations.

With the standard viscosity $\nu_0 = 1$, a is halved in less than $100P_0$, implying classic type III migration (Papaloizou et al., 2006). Comparing different ν_0 , $a(t)$ is indistinguishable for $0 < t \lesssim 15P_0$, since viscous timescales are much longer than the orbital timescale. At $t = 20P_0$, $|\dot{a}|$ increases with ν_0 . It has been shown that in the limit $\nu \rightarrow 0$, the horse-shoe drag on a planet on fixed orbit is $\propto \nu$ (Balmforth et al., 2001; Masset, 2002). However, in this case $\dot{a} \neq 0$ so there is a much larger rate-dependent

2. VORTEX-INDUCED TYPE III MIGRATION

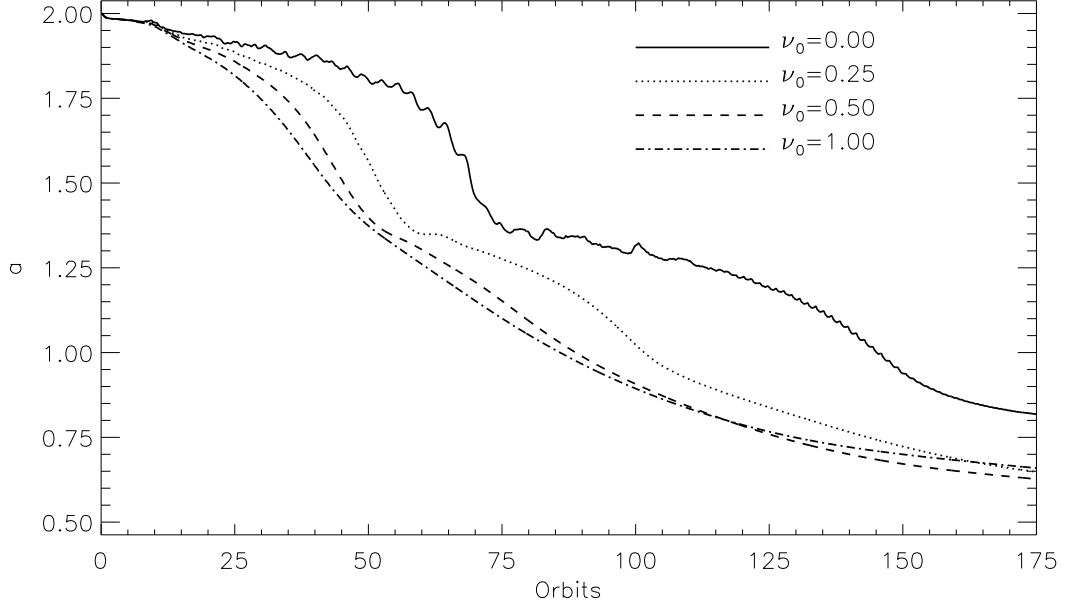


Figure 2.14: *Migration as a function of viscosity. Migration for $\nu_0 = 1$ indicates classic type III migration. As viscosity is lowered, migration becomes non-smooth. The inviscid case has two distinct migration timescales: the initial slower migration followed by a sudden drop in a .*

torque responsible for type III migration, for which explicit dependence on viscosity has not been demonstrated analytically.

Migration initially accelerates inwards ($\dot{a}, \ddot{a} < 0$) and subsequently slows down at $r \sim 1.4$ (independent of ν_0). For $\nu_0 = 1.0, 0.5$ migration proceeds smoothly, decelerating towards the end of the simulation at which point a has decreased by a factor of ~ 2.7 . Migration curves for $\nu_0 = 1.0$ and $\nu_0 = 0.5$ are quantitatively similar. Lowering ν_0 further enhances the deceleration at $r \sim 1.4$ until in the inviscid limit the migration stalls before eventually restarting.

Despite differences in detail, the overall extent of the orbital decay in all of these cases is similar. This is expected in the model of type III migration where the torque is due to circulating fluid material switching from $r_p - x_s \rightarrow r_p + x_s$. In this model the extent of the orbital decay should not depend on the nature of the flow across r_p , but only on the amount of disc material participating in the interaction, or equivalently the disc mass and this does not depend on ν . On the other hand, the flow may not be a smooth function of time with migration proceeding through a series of fast and slow episodes as observed in Fig. 2.14.

2.5 Simulations of fast migration driven by vortex-planet interaction

The interpretation above is only valid if migration proceeds via the type III mechanism. As indicated by Eq. 2.38, the torque depends on the contrast between the co-orbital region and the flow just outside. If this contrast is small, e.g. due to viscous diffusion when large ν is employed, then the type III mechanism cannot operate at all.

2.5.2 Stalling of type III migration

Classic type III migration is fast because it is self-sustaining (Masset & Papaloizou, 2003). The issue discussed here is what inhibits the growth of $|\dot{a}|$ seen in Fig. 2.14? Descriptions of type III migration usually assume that the libration time at $r_p - x_s$ is much less than the time to migrate across the co-orbital region (Masset & Papaloizou, 2003). This implies that

$$\chi \equiv \frac{|\dot{a}|\pi a}{|A_p|x_s^2} \ll 1, \quad (2.39)$$

where $A_p = 1/2(\partial\Omega/\partial r)$ at $r_p(t)$ ($\simeq a(t)$). Papaloizou et al. (2006) present a similar critical rate with the same dependence on Ω , a , x_s . If Eq. 2.39 holds, co-orbital material is trapped in libration on horseshoe orbits and migrates with the planet. When $\chi \gtrsim 1$ the horseshoe region shrinks to a tadpole, and material is trapped in libration about the L4 and L5 Lagrange points (as observed by Pepliński et al., 2008b). This can tend to remove the co-orbital mass deficit δm ¹ which reduces the migration torque. Comparing χ for cases shown in Fig. 2.15, it is clear that migration with $\chi \ll 1$ does not always hold, with $\max(\chi) \sim 0.6$ being comparable for different ν_0 .

By following the evolution of a passive scalar in Fig. 2.16, it can be seen that horseshoe material indeed no longer migrates with the planet when $|\dot{a}|$ is large. This occurs for all ν but only the low viscosity cases exhibit stalling. Hence, while horseshoe material is lost due to fast migration, it is not responsible for stopping it. By examining the inviscid case later, the stopping of migration will be shown to be due to the flow of a vortex across the co-orbital region, where some of it becomes trapped in libration.

2.5.3 The connection between vortensity and fast migration

The difference between the value of the inverse vortensity Σ/ω evaluated in the co-orbital region and the value associated with material that passes from one side of the co-orbital region to the other, defines the co-orbital mass deficit δm in Masset &

¹The process of shrinking from a partial gap that extends nearly the whole azimuth to one with a smaller azimuthal extent, can be regarded as gap filling.

2. VORTEX-INDUCED TYPE III MIGRATION

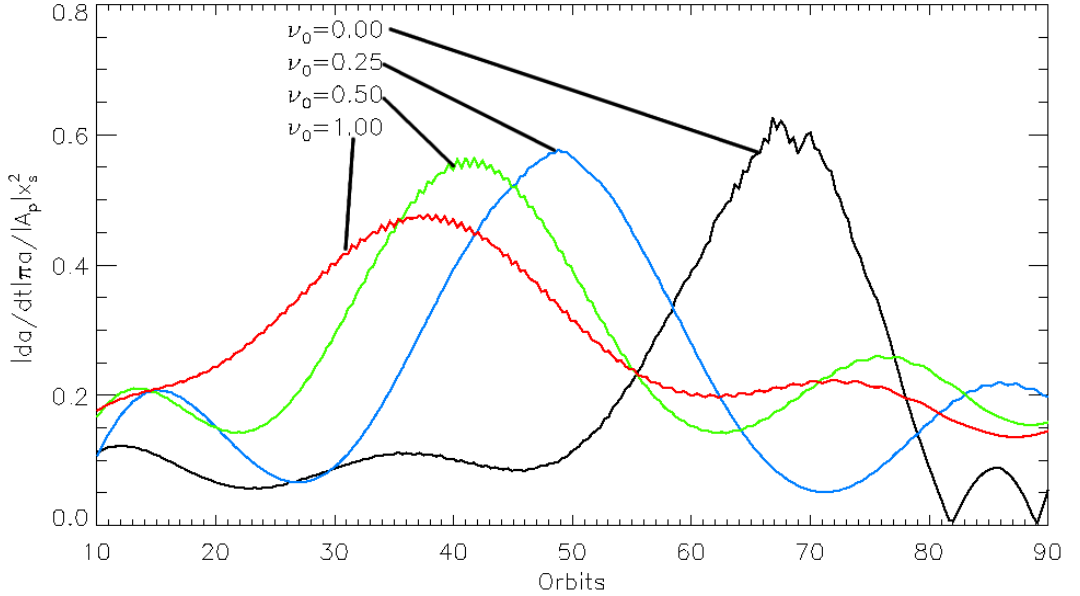


Figure 2.15: Evolution of χ , the ratio of libration-to-migration timescale as a function of viscosity. Classic, smooth type III migration can be analytically described in the $\chi \ll 1$ regime. Libration time is measured at $r_p - r = x_s$ and migration timescale is that across x_s , with $x_s = 2.5r_h$.

Papaloizou (2003). It is often assumed that the vorticity is slowly varying so that the difference in the values of inverse vortensity reduces, to within a scaling factor, simply to the difference in the values of the surface density (Eq. 1.3).

Although Masset & Papaloizou (2003) assumed steady, slow migration in the low viscosity limit, which clearly does not hold when the disc is unstable, it is nevertheless useful to examine the evolution of Σ/ω in relation to the migration of the planet. Fig. 2.17 shows the azimuthally averaged Σ/ω -perturbation following planet migration. Introducing the planet modifies the co-orbital structure on orbital timescales. Vortensity rings develop at $r - r_p \simeq \pm 2r_h$ within $t \lesssim 10P_0$ (Fig. 2.17(a)), as modelled previously. Note that vortensity rings were not accounted in Masset & Papaloizou’s model.

Increasing ν reduces the rings’ amplitude, but their locations are unaffected. Taking the length-scale of interest as $l = r_h \simeq 0.1$, for $\nu_0 = 1$ the viscous timescale is $t_d = l^2/\nu \simeq 56P_0$. Hence at $t \lesssim 10$ viscous diffusion is not significant even locally. Thus ring-formation is not sensitive to the value of ν within the range concerned. Note the correspondence between the similarity of the Σ/ω profiles and similarity in $a(t)$ for different ν in the initial phase. That is, the co-orbital disc structure strongly affects

2.5 Simulations of fast migration driven by vortex-planet interaction

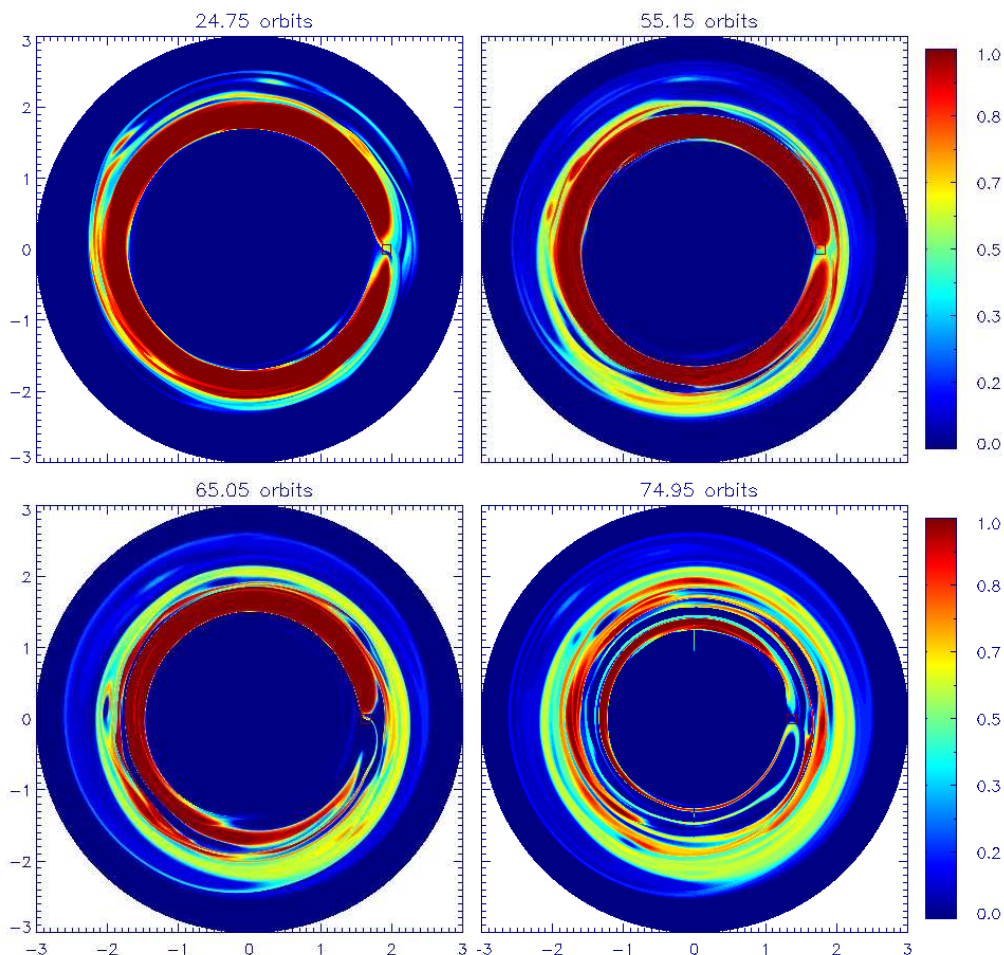


Figure 2.16: *Loss of horseshoe (co-orbital) material during fast migration. Snapshots are taken from the inviscid case ($\nu_0 = 0$). The colour bar shows the surface density of a tracer fluid (on a linear scale), initially placed in the co-orbital region of the planet.*

migration (Masset & Papaloizou, 2003). Dependence on the value of the viscosity is seen beyond $t = 50P_0$ (Fig. 2.17(b)), producing much smoother (and similar) profiles for $\nu_0 = 0.5$ and $\nu = 1.0$.

In Fig. 2.17(b) only the inviscid case is still in slow migration, and only this disc retains the inner ring (low Σ/ω). This suggests that the inner vortensity ring inhibits inward migration. In terms of δm , for $\nu \neq 0$ the planet resides in a gap (co-orbital Σ/ω is less than that at the inner separatrix, or $\delta m > 0$) whereas in the inviscid case $\delta m \sim 0$.

Consider the $\nu = 0$ case. The outer vortensity ring has widened to $\sim 2r_h$ (c.f.

2. VORTEX-INDUCED TYPE III MIGRATION

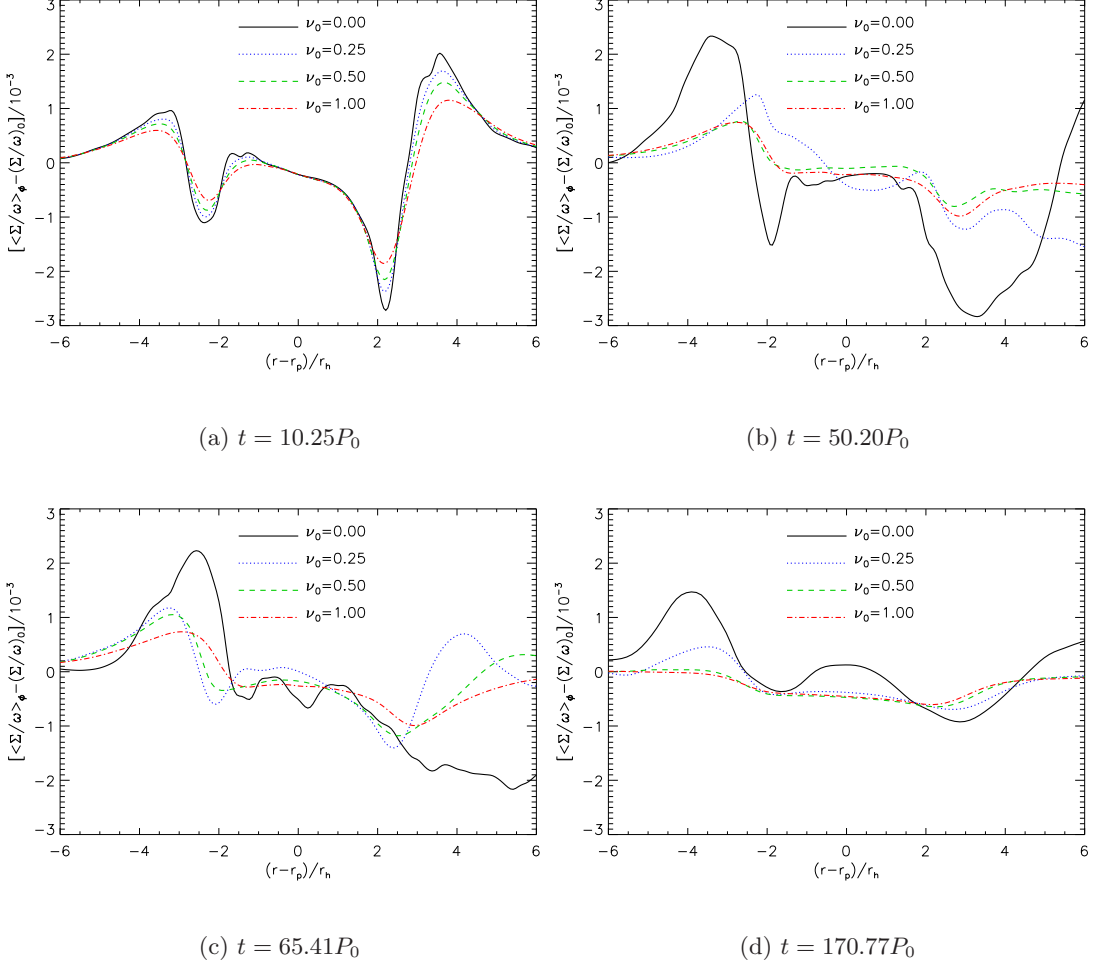


Figure 2.17: Azimuthally averaged Σ/ω -perturbation for different viscosities for the case $\Sigma_0 = 7$ and $M_p = 2.8 \times 10^{-4} M_*$. The co-orbital disc structure in terms of Σ/ω strongly correlates to migration behaviour.

Fig. 2.17(a)). It is centred at $3r_h$ so that co-orbital dynamics may not account for it. However, inward migration implies a flow of material across r_p from the interior region. The increased region of low Σ/ω exterior to the planet may be due to this flow. Notice the high- Σ/ω ring at $+4r_h$ in Fig. 2.17(a) is no longer present in Fig. 2.17(b) because this ring is not co-orbital and therefore does not migrate with the planet.

At $t = 65P_0$ (Fig. 2.17(c)) the $\nu_0 = 0.5$ and $\nu = 1.0$ cases continue smooth migration (Fig. 2.14) with qualitatively unchanged profiles. For $\nu_0 = 0.25$, characteristic vortensity double-rings re-develop after a stalling event at $t \sim 60P_0$ (Fig. 2.14). The

2.5 Simulations of fast migration driven by vortex-planet interaction

peaks and troughs of Σ/ω recover forms that are close to those in the initial phase (Fig. 2.17(a)). At this time the inviscid case is in rapid migration, and there is a large Σ/ω just interior to the planet, i.e. a positive co-orbital mass deficit.

Fig. 2.17(d) shows the final Σ/ω -perturbation profiles. Viscous cases are again in slow migration, and have much smoother profiles. This can be due to diffusion (since this is the late stage of evolution) and/or migration across the background, in both situations there is little disc material carried by the planet. The inviscid case is also in slow migration but retains the double-ring structure. This indicates two possible co-orbital configurations which slow down type III migration.

Fig. 2.17 shows a connection between vortensity evolution and the state of migration. A correlation between migration (fast, slow) and co-orbital structure is apparent. The effective action of viscosity appears to be through its modification of the disc structure, rather than associated viscous torques acting on the co-orbital region.

2.5.4 Evolution of the co-orbital region

Fig. 2.18 illustrates the evolution of the following quantities associated with the co-orbital region of the migrating planet

1. The gravitational torque acting on the planet due to fluid with $|r - r_p| \leq 2.5r_h$, excluding fluid within r_h of the planet. This includes co-orbital material and orbit-crossing fluid, the latter being responsible for the type III torque.
2. The co-orbital mass deficit δm is computed from azimuthally averaged, one-dimensional disc profiles. The separatrices are taken to be at $|r - r_p| = 2.5r_h$.
3. M_{tr} , the mass of a passive scalar initially placed such that $|r - r_p| = 2r_h$ ¹. Note that $M_{\text{tr}} = \text{const.}$ if it migrates with the planet.
4. The average density $\langle \Sigma \rangle$ and vortensity $\langle \omega/\Sigma \rangle$ of the region $|r - r_p| < 2.5r_h$.

The evolution of viscous and inviscid cases is qualitatively similar up to the stalling indicated by vertical lines in Fig. 2.18. Just prior to this there is rapid migration associated with a large negative torque, which was found to originate from material crossing the planet's orbit. Co-orbital torques are oscillatory and the period (amplitude) is longer (larger) for $\nu_0 = 0$ than for $\nu_0 = 0.5$. During rapid migration phases, the inviscid torque is twice as negative than in the viscous case. While the torque in a

¹Due to the uncertainty in the horseshoe half-width x_s , the tracer was initially placed well within the region defined by $x_s = 2.5r_h$ to ensure it is co-orbital.

2. VORTEX-INDUCED TYPE III MIGRATION

viscous disc remains negative, torques in an inviscid disc can be positive due to the formation of large-scale vortices. Note that the torque does not originate from within the Hill sphere since it is excluded from the summation.

Migration is slow until sufficient difference builds up between co-orbital and circulating flow at which point there is a sudden flow-through the co-orbital region. At the same time there is significant loss of the original horse-shoe material (M_{tr} decreases by $\sim 80\%$). The flow-through is reflected in $\nu_0 = 0.5$ ($\nu_0 = 0$) by a $\sim 17\%$ (36%) increase in $\langle \Sigma \rangle$ from $t = 25P_0 \rightarrow 50P_0$ ($t = 50P_0 \rightarrow 75P_0$). In the case of zero viscosity, this material is in the form of a high surface density vortex. Notice in Fig. 2.18(b) that there is a repeated episode where there is a fall followed by a rise in $\langle \Sigma \rangle$. However, in the late stages of $\nu_0 = 0.5$, $\langle \Sigma \rangle$ decreases monotonically and the migration is slower.

The co-orbital mass deficit is initially negative¹ as the rings develop. It subsequently increases resulting in the onset of type III migration and is most positive during the following rapid migration phase, with peak values $\delta m \times 10^3 \sim 1, 2$ for $\nu_0 = 0.5, 0$ respectively. This is a few times larger than the planet mass. δm then falls to $+0.5 \times 10^{-3}$ in the viscous case but to $\lesssim 0$ without viscosity. In the inviscid case, material flowing into the co-orbital region removes the co-orbital mass deficit and type III is suppressed; migration experiences a more abrupt stall. δm increases again for $\nu_0 = 0$ while it remains approximately constant for $\nu_0 = 0.5$ and decreases towards the end. Type III migration can re-start in the inviscid disc but such behaviour is not observed for large viscosity. Type III is not operating in the late stages of the viscous case, in contrast to the inviscid evolution where fast type III migration is recurrent, faster than in the cases with applied viscosity, and is associated with large values of δm .

The above discussion shows that the magnitude of the applied viscosity is significant in determining the character of the migration. This is because the form of the flow through the co-orbital region is sensitive to the choice of viscosity. In particular, for sufficiently high viscosity, this flow is smooth and there is less disruption of the co-orbital region.

¹Ring structure in the vicinity of the separatrix means that the sign of δm is sensitive to the adopted value of x_s . Here, δm is regarded as a representation of gap depth, so $x_s = 2.5r_h$ is fixed.

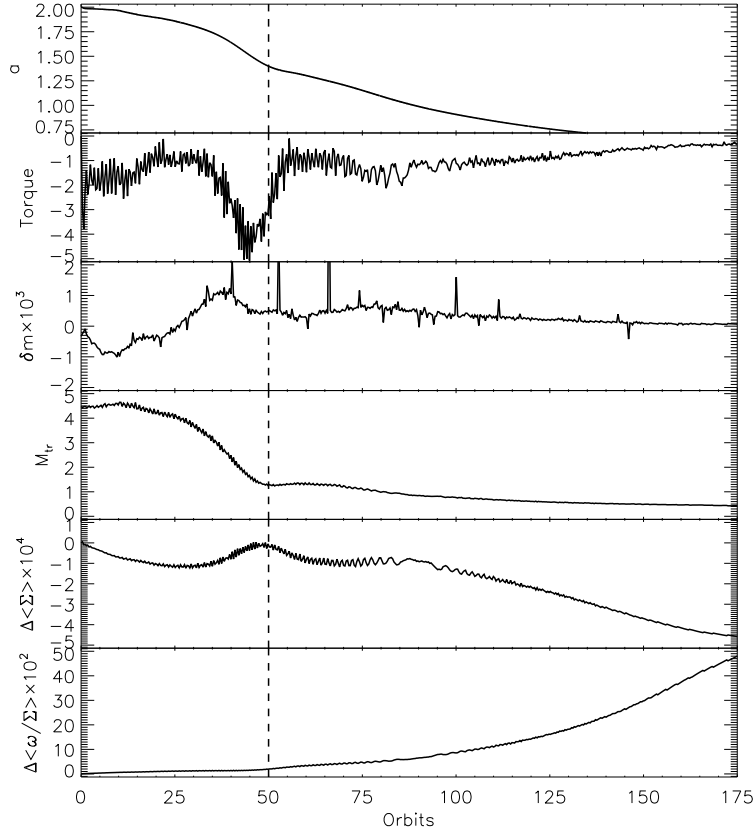
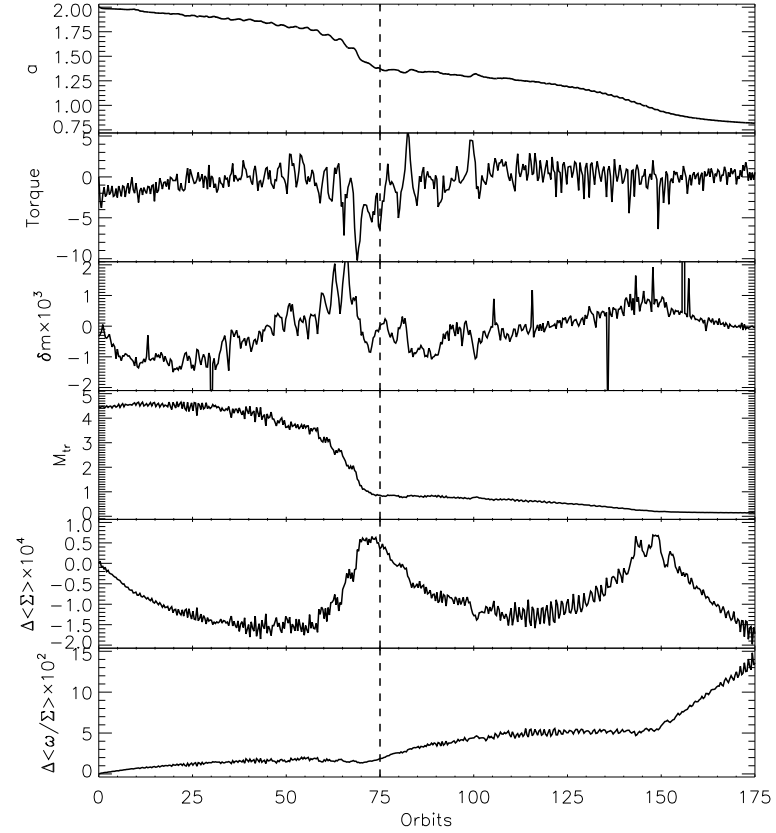
(a) $\nu_0 = 0.5$ (b) $\nu_0 = 0$

Figure 2.18: *Effect of viscosity on the evolution of co-orbital disc properties of a migrating planet. The time evolution is illustrated for the model with $\nu_0 = 0.5$ (left panel) and for the inviscid model (right panel). From top to bottom of each panel the evolution of the semi-major axis, $a(t)$, the co-orbital mass deficit δm , the disc-planet torque, the tracer mass M_{tr} , the surface density $\langle \Sigma \rangle$ and the vortensity $\langle \omega / \Sigma \rangle$ are plotted. Angle brackets denote space averaging over the annulus $[r_p - 2.5r_h, r_p + 2.5r_h]$ and Δ denotes perturbation relative to $t = 0$. The vertical line indicates stalling of migration.*

2.6 The inviscid disc and parameter studies

In the inviscid disc, the development of vortices significantly affect migration. Three phases of migration are apparent from Fig. 2.14: a) $t \lesssim 45P_0$ (slow migration); b) $60P_0 \lesssim t \lesssim 70P_0$ (rapid migration, or a jump in orbital radius); c) $t \sim 75P_0$ (stalling) and d) $75P_0 \lesssim t \lesssim 110P_0$ (second phase of slow migration). Typical migration rates at various times are $\dot{a}(25) \sim -3 \times 10^{-3}$; $\dot{a}(65) \sim -2 \times 10^{-2}$ and $\dot{a}(85) \sim -2 \times 10^{-3}$. The vortex-induced rapid migration (phase b) is almost an order of magnitude faster than the phases a) and d). Hence the latter is referred to as slow migration, but more accurately they are migration during gap formation. They are not necessarily slow in comparison to type I or type II migration.

Different migration phases correspond to different disc structures. Fig. 2.19 shows the global surface density evolution. At $t = 24.75P_0$, during the slow migration (phase a), the planet resides in a partial gap ($|r - r_p| \lesssim 2.3r_h$) with surface density $\sim 20\%$ lower than the uniform background, Σ_0 . A partial gap is necessary for the type III migration mode (Papaloizou & Terquem, 2006) but not sufficient. A surface density asymmetry ahead/behind the planet is needed to provide the net co-orbital torque (Artymowicz, 2004a). At this stage, non-axisymmetry in the gap is too weak. Further azimuthal density asymmetry has developed in the gap by $t = 55P_0$ and there is a factor of ~ 3 variation in the gap surface density, but the asymmetry is still limited.

Strong asymmetry can be provided by large-scale vortices near the gap edges of azimuthal extent $\Delta\phi \sim \pi$ (Fig. 2.19). The outer (inner) vortex rotates clockwise (counter-clockwise) relative to the planet so they are not co-orbital. Their origin was explained as a natural consequence of the instability of the vortensity rings (see §2.4 above). Their occurrence near gap edges has a strong influence on the type III migration mode.

At $t = 65.05P_0$ Fig. 2.19 shows that the planet is just inside the inner gap edge. The surface density contrast in the neighbourhood of the planet is largest as the inner vortex enters the co-orbital region from behind the planet. It exerts a net negative torque as the material crosses the planet orbit and enters (is scattered into) the exterior disc, and this snapshot corresponds to fast migration (phase b). At $t = 75P_0$ the migration stalls and the planet no longer resides inside a gap. The planet has effectively left its gap by scattering vortex material outwards. This completes a single vortex-planet episode. During this time, the outer vortex simply circulates around the original outer gap edge and does not influence the co-orbital dynamics, though it contributes an oscillatory torque on the planet.

2.6 The inviscid disc and parameter studies

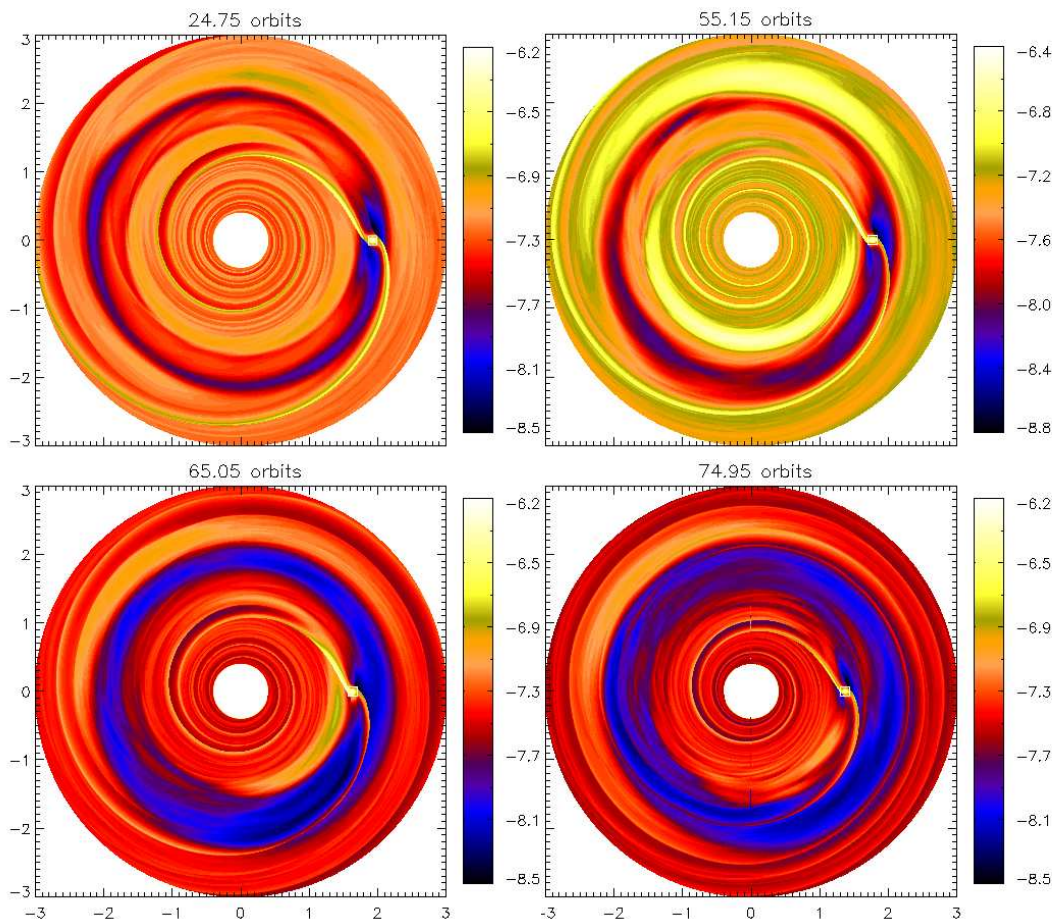


Figure 2.19: *The evolution of $\ln \Sigma$ from early slow migration to stalling (~ 75 orbits). The disc is inviscid.*

The vortex-planet interaction is magnified in Fig. 2.20. At $t = 65P_0$ the vortex circulates at $\sim r_p - 3r_h$ and has radial extent $\sim 3r_h$. The gap depth is largest and the migration is fast. As the planet migrates inwards, the vortex splits with some material entering the co-orbital region while the rest continues to circulate ($t = 70P_0$). Vortex material becomes trapped just behind the planet at $t = 75P_0$. (For large viscosity cases, no trapping is seen.) The flow of vortex material into the co-orbital region is gap filling, thereby lowering the co-orbital mass deficit, discouraging type III migration. Part of the original horseshoe material is replaced with vortex material. Another view is that rapid migration in the first place was provided by the vortex just interior to the co-orbital region. When this vortex is no longer there, rapid migration cannot be sustained.

2. VORTEX-INDUCED TYPE III MIGRATION

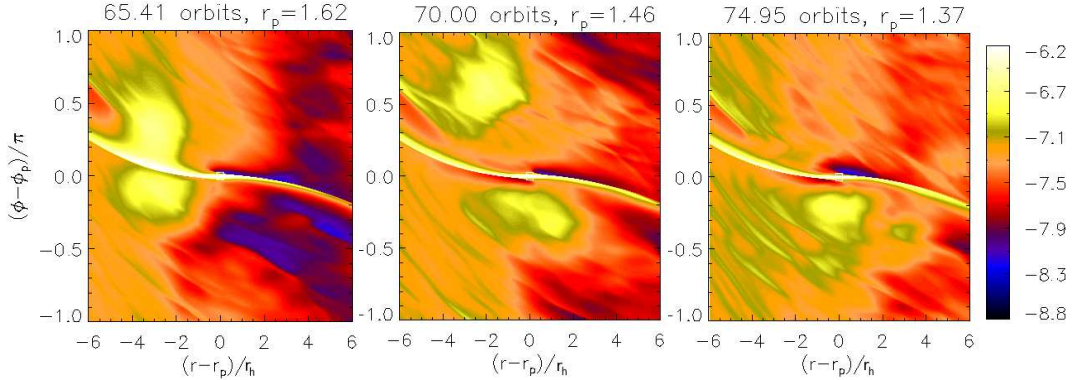


Figure 2.20: *Illustration of the surface density evolution from the start of rapid migration to stalling at $t = 75P_0$. Maps of $\ln \Sigma$ are plotted.*

2.6.1 Evolution of inverse vortensity

It is useful to examine the evolution of Σ/ω , or inverse vortensity, since it defines co-orbital mass deficit δm that drives the type III torque. This is illustrated in Fig. 2.21. In inviscid discs Σ/ω is approximately conserved following a fluid element/streamline so it also tracks material. By $t = 55P_0$ the inner vortex has formed via nonlinear evolution of the gap edge instability (§2.4) and begins to interact with the co-orbital region. Vortensity conservation implies that the small vortices (red blobs) near the outer ring were part of the inner vortex, consistent with inward type III migration (see Fig. 2.21). Each time this vortex passes by the planet, some of its material breaks off and crosses the planet orbit from behind, thereby exerting a negative torque. Rapid migration at $t = 65P_0$ occurs when the bulk vortex body flows across the co-orbital region. The inner ring is disrupted and no longer extends 2π in azimuth.

This contrasts with the usual type III scenario where material simply transfers from inner to outer disc leaving the co-orbital region unaffected. In the inviscid disc, disruption is *necessary* due to the existence of vortensity rings of much higher vortensity than the vortex. Having type III migration while maintaining vortensity conservation means that vortex material must cross the planet orbit without changing its vortensity. This would not be possible if a ring structure is maintained. In this sense, the high vortensity rings oppose the type III mode. Hence, migration is slow until significant ring disruption occurs that is associated with the vortex flowing across.

When migration stalls at $t = 75P_0$, Fig. 2.21 shows that the vortensity rings are much less pronounced compared to the initial phase. The vortex splits into several smaller patches circulating in the original gap. At the planet's new radius, material of

2.6 The inviscid disc and parameter studies

high Σ/ω fills the new co-orbital region, corresponding to lower co-orbital mass deficit. However, by $t = 85P_0$ new vortensity rings are setup near the new orbital radius and are qualitatively similar to those present in the first ring basic state. During the build up to the second rapid migration phase, there is a single vortex associated with inner gap edge ($t = 100P_0$). This simply leads to a repeat of the first rapid migration episode. However, the presence of a vortex inside the co-orbital region, from the first rapid migration phase lowers the co-orbital mass deficit (Fig. 2.18(b)) and hence reduces the migration rate for the second rapid phase (at $t \sim 140P_0$).

2.6.2 Effect of surface density scale

By changing the parameter Σ_0 , and therefore the disc mass, the disc-on-planet torque is scaled accordingly. In the case of vortex-induced migration, changing Σ_0 directly affects the mass contained in a vortex. Here, the inviscid simulations above are repeated with different surface density scales. Simulations here had a resolution $N_r \times N_\phi = 512 \times 1536$.

Fig. 2.22 shows migration as a function of Σ_0 . For $\Sigma_0 = 5$ –9 the planet migrates by the same amount during the first rapid migration phase and stalls at the same radius¹. This is also observed for a second rapid migration phase for $\Sigma_0 = 6$ –9. Although the largest Σ_0 case is unphysical due to the lack of inclusion of disc self-gravity, results are consistent with the notion that rapid migration is initiated by sufficient contrast between co-orbital (gap) and circulating fluid (vortex), which can be measured by δm .

The jump in orbital radius, should it occur, is independent of Σ_0 . As the interaction involves the vortex flowing from the gap edge across the co-orbital region; its change in specific angular momentum is independent of density, because the co-orbital region size is fixed by planet mass. The results then suggest the vortex needs to grow to a mass M_v (or surface density), only dependent on the planet mass, in order to scatter the planet. Since the vortex forms at the gap edge, M_v can be linked to δm because the co-orbital mass deficit depends on the edge surface density.

As the vortex originates from instabilities with growth rate independent of Σ_0 , increasing Σ_0 means less time is needed for the vortex to build up to critical mass or density (to affect migration). Hence, increasing surface density only shortens the ‘waiting time’ before rapid migration. However, if the density is too low, e.g. $\Sigma_0 = 3$ then vortex-induced rapid migration may never occur.

Consider an inner vortex of mass M_v formed by instability at $r_v = r_p - br_h$ ($b > 0$)

¹This contrasts with the classic smooth migration scenarios. For example, in type I migration the disc-planet torque scales with surface density so in a given time interval, planets should migrate further in discs with higher surface density.

2. VORTEX-INDUCED TYPE III MIGRATION

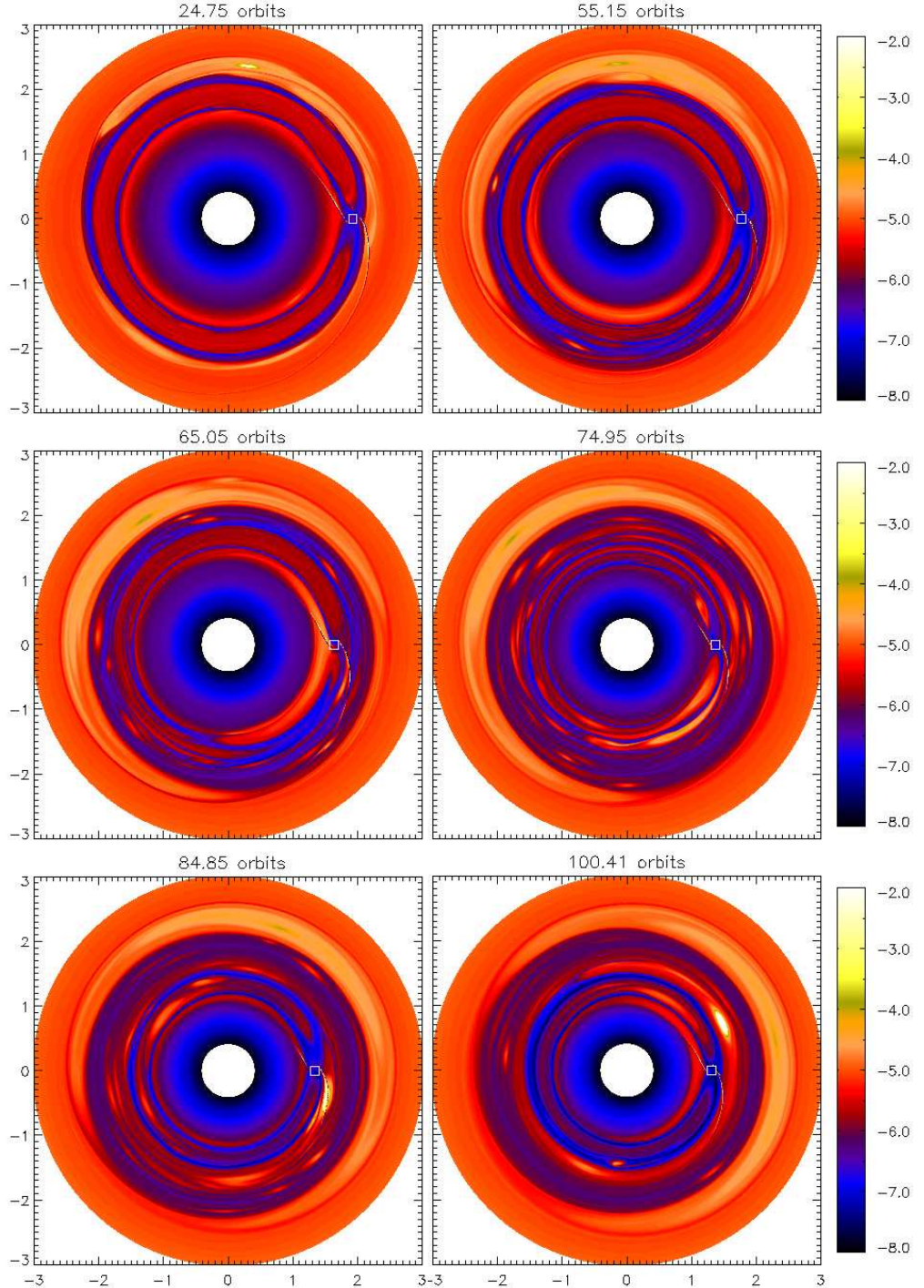


Figure 2.21: Maps of the inverse vortensity $\ln(\Sigma/\omega)$ spanning the time interval from early slow migration to stalling (~ 75 orbits), then to the second phase of slow migration. The passage of the inner vortex material to the outer disc leads to formation of smaller vortices around the outer gap edge (e.g. at $t = 75P_0$).

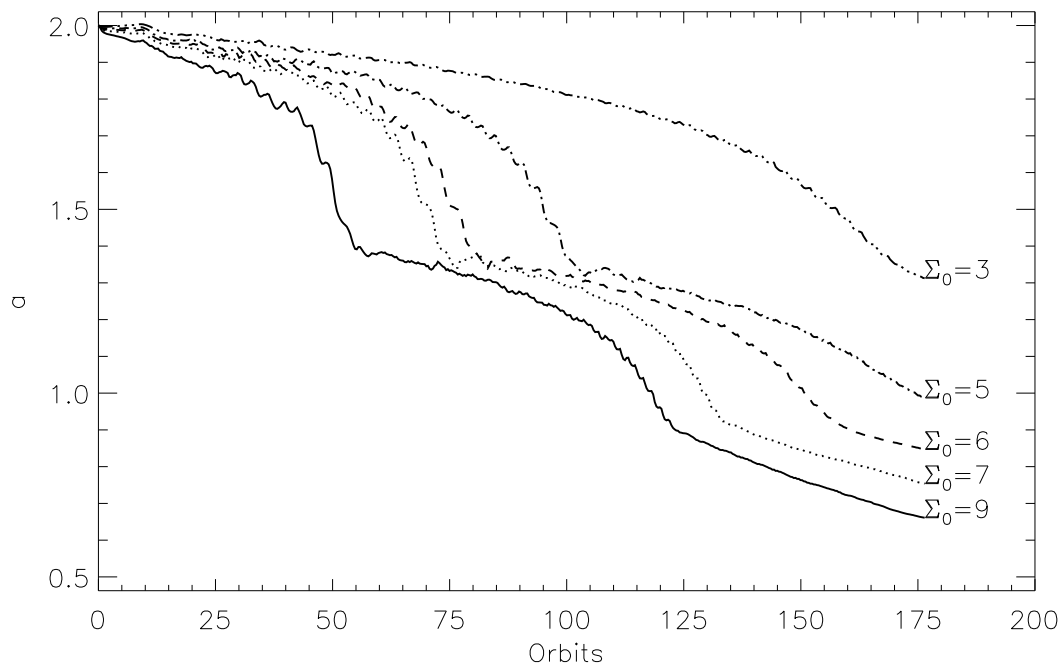


Figure 2.22: *Vortex-induced migration in inviscid discs with different surface density scales. The extent of rapid migration is independent of the initial surface density scaling. The fiducial case is $\Sigma_0 = 7$.*

with width ar_h . M_v is clearly limited by the amount of material that can be gathered into the vortex, so that $M_v < 2\pi ar_h r_v \Sigma$. Taking $M_v/M_p = 3.5$ as critical for rapid migration¹, this means

$$\Sigma_0 > \frac{3.5M_p}{2\pi a f_0 (1 - b f_0) r_p^2} \times 10^4, \quad (2.40)$$

where $f_0 = (M_p/3M_*)^{1/3}$. Taking representative values from simulations of $a = 3$, $b = 4$ and $r_p \sim 2$ gives $\Sigma_0 > 3.5$. For such cases rapid, vortex-induced migration was indeed observed (Fig.2.22), but not for $\Sigma_0 = 3$. This is similar to the usual requirement that in order for type III migration to occur for intermediate planets, the disc should be sufficiently massive (Masset & Papaloizou, 2003). In the inviscid case the limitation is specifically due to the maximum possible vortex mass, which is responsible for most of the migration.

¹The vortex mass can be estimated by monitoring the drop in total mass of an annulus interior to the inner gap edge, during rapid migration.

2. VORTEX-INDUCED TYPE III MIGRATION

2.6.2.1 Critical co-orbital mass deficit

In order to link type III migration and vortex-planet scattering, one can measure the co-orbital mass deficit δm , which amounts to comparing average inverse vortensity of co-orbital fluid just behind the planet, to that of circulating fluid just inside the inner separatrix but also behind the planet:

$$\delta m = 2\pi\sqrt{GM_*x_s}r_p^{-1/2}(\langle\Sigma/\omega\rangle_{\text{circ}} - \langle\Sigma/\omega\rangle_{\text{coorb}}).$$

This is a simplified version of the definition in Masset & Papaloizou (2003), by inserting Keplerian rotation for the Oort constant in Eq. 2.38. Results are shown in Fig. 2.23. The oscillatory nature of δm reflects a vortex circulating at the gap edge, δm maximising when the vortex is within the patch of fluid where averaging is done. As it grows, the vortex contributes to $\langle\Sigma/\omega\rangle_{\text{circ}}$ hence favouring type III migration. Cases with rapid migration share the same evolution of δm . $\delta m/M_p$ increases up to $\sim 4\text{--}5$ before the vortex first induces fast migration. For $\Sigma_0 = 3$, which does not show such a rapid migration phase, typically $\delta m \lesssim 5M_p$ with smaller amplitude variation. During the vortex-planet scattering event, δm rapidly decreases and migration stalls when $\delta m \lesssim 0$. This is because as the vortex flows across the co-orbital radius, it contributes to $\langle\Sigma/\omega\rangle_{\text{coorb}}$, lowering δm .

2.6.3 Effect of background profile

Simulations so far employed a flat initial surface density profile. In this case the stalling and re-start of migration can be expected because the vortex scatters the planet out of its original gap into a previously unperturbed region. If the unperturbed surface density is flat, subsequent migration is a repeat of that occurred previously. However, if the unperturbed surface density increases with decreasing radius, then the vortex-induced inwards migration moves the planet into a region of higher surface density. The background contribution to $\langle\Sigma/\omega\rangle_{\text{circ}}$ may outweigh the vortex contribution to $\langle\Sigma/\omega\rangle_{\text{coorb}}$ just after the vortex-planet interaction. In this case, migration may not stall.

To investigate this, simulations with initial surface density profiles

$$\Sigma = 5\left(\frac{r}{2}\right)^{-p} \times 10^{-4} \quad (2.41)$$

were performed with $p = 0, 0.3, 0.5, 1.0$. Migration is compared in Fig. 2.24. Note that for $p > 0$, the co-orbital mass deficit is positive at $t = 0$, and there is migration

2.6 The inviscid disc and parameter studies

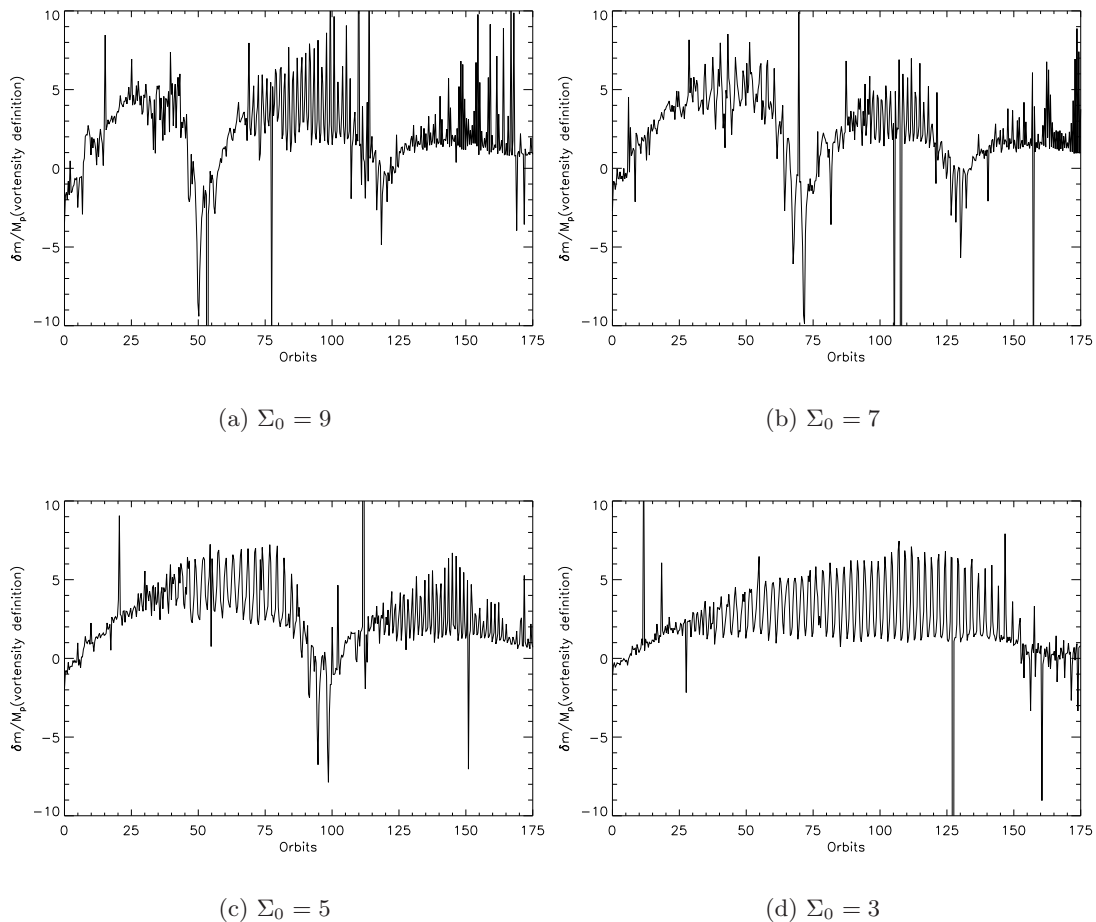


Figure 2.23: Evolution of the coorbital mass deficit as defined by inverse vortensity. The average Σ/ω for coorbital material is taken over $r - r_p = [-2.5r_h, 0]$, $\phi = [\phi_p, \phi_p - \pi/4]$; and that of circulating material is taken over $r - r_p = [-6, -2.5]r_h$, $\phi = [\phi_p, \phi_p - \pi/4]$. Very similar behaviour was obtained when comparing the co-orbital and circulating surface densities instead of the inverse vortensity.

from $t = 0$. With $p = 1$ this initial migration is large and the planet falls to the inner boundary, having no time to open a gap and allow vortex formation. For $p = 0.5$ there is vortex-induced inwards migration at $t \sim 50P_0$. The wait until vortex-planet interaction is shorter for $p = 0.5$ than $p = 0$. $p = 0.5$ does experience deceleration at $t \sim 70P_0$ but unlike $p = 0$, does not stall for long. In this case, gap-filling by the vortex does not effectively stall type III migration. The shortened duration of stalling with increasing p is consistent with the discussion above regarding the variation with the surface density scale (being higher in $r < r_p$ for $p = 0.5$ than for $p = 0$).

2. VORTEX-INDUCED TYPE III MIGRATION

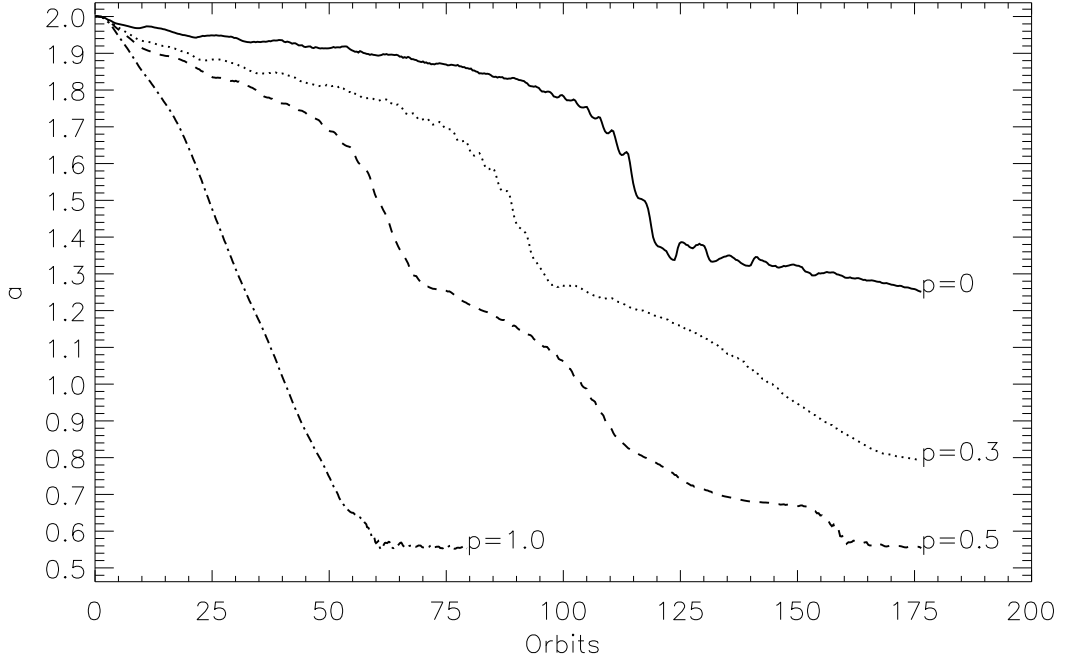


Figure 2.24: *Migration in inviscid discs with variable initial surface density slopes ($\Sigma \propto r^{-p}$).*

2.6.4 Effect of disc temperature

Linear calculations showed that the growth rate of unstable co-rotational (vortex) modes, $|\gamma|$, generally increase when the aspect-ratio h is lowered. Migrating cases with $h = 0.04$ – 0.06 are presented here.

Fig. 2.25 show snapshots of Σ/ω as a function of h and agree with linear theory in that co-rotational modes dominate the region. Both edges become more unstable for lower h . A $m = 4$ mode on $h = 0.04$ inner edge can be seen, which has peak growth rate from theory (Fig. 2.13). A weak $m = 2$ mode is observed for $h = 0.05$ inner edge, consistent with linear theory peak rates ($m = 2$ – 3 is most unstable for $h = 0.05$). As the planet tends to migrate inwards initially, one expects interaction with inner edge vortices to occur sooner for lower h . Disturbances at outer edges have higher growth rates, and have already undergone merging into large vortices.

Fig. 2.26 show corresponding migration curves. Vortensity rings can act to inhibit type III migration. For example, in Fig. 2.26 the stronger vortensity rings associated with $h = 0.04$ results in almost no migration prior to vortex-induced migration. In

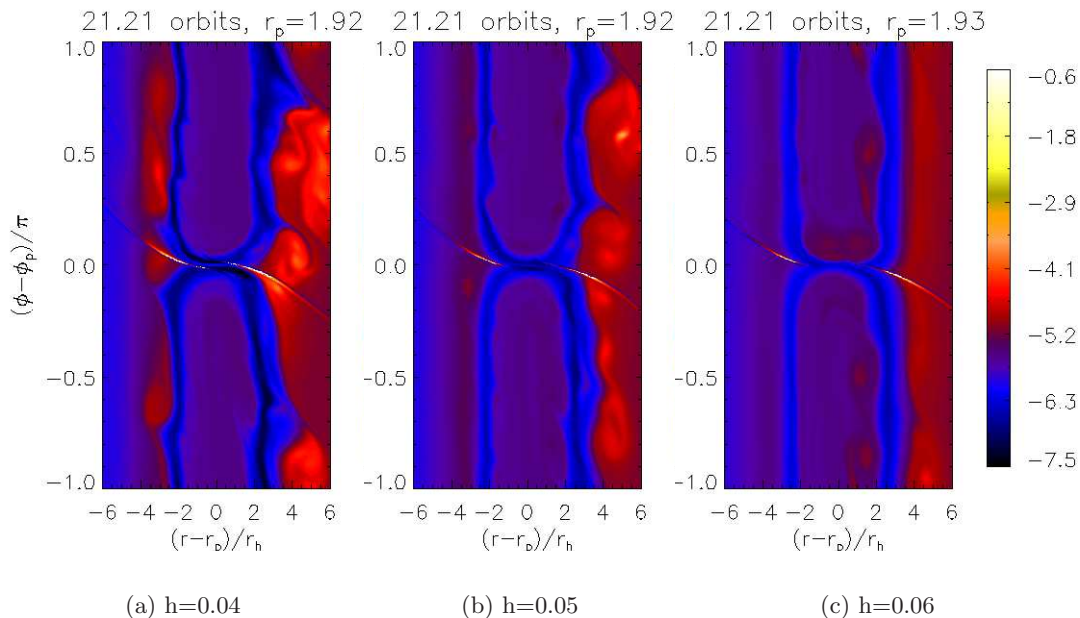


Figure 2.25: Distribution of $\ln(\Sigma/\omega)$ near the planet's co-orbital region, as a function of disc aspect-ratio h .

$h = 0.04$, rapid migration begins at $t = 25P_0$, roughly half the waiting time than $h = 0.05$. The extent of orbital decay during rapid migration is unchanged, consistent with the notion that the vortex implies a co-orbital mass deficit, for which a critical value reached enables rapid migration and determines migration extent. The critical value is simply attained sooner with decreasing h .

More rapid migration episodes occur in $h = 0.04$ than $h = 0.05$, since $h = 0.04$ has higher growth rates so the wait between vortex interaction is shorter. The slow phase of $h = 0.04$ is also non-smooth compared to $h = 0.05$ due oscillatory torques exerted on the planet by vortices. This is not apparent in the migration curve of $h = 0.05$ because perturbations have smaller amplitudes at a given time.

Notice in $h = 0.04$ the waiting times between rapid migration lengthens with time. This is probably because the vortex forming region has width of order $H = hr$ which is decreasing, as is the co-orbital width ($x_s \propto r_p$). The vortex flows across a smaller region as it scatters the planet, which can account for smaller jumps in radius in later interactions.

In $h = 0.06$ no rapid migration occurs within simulation time, in agreement with linear theory where co-rotational modes for the inner edge were not found, and with

2. VORTEX-INDUCED TYPE III MIGRATION

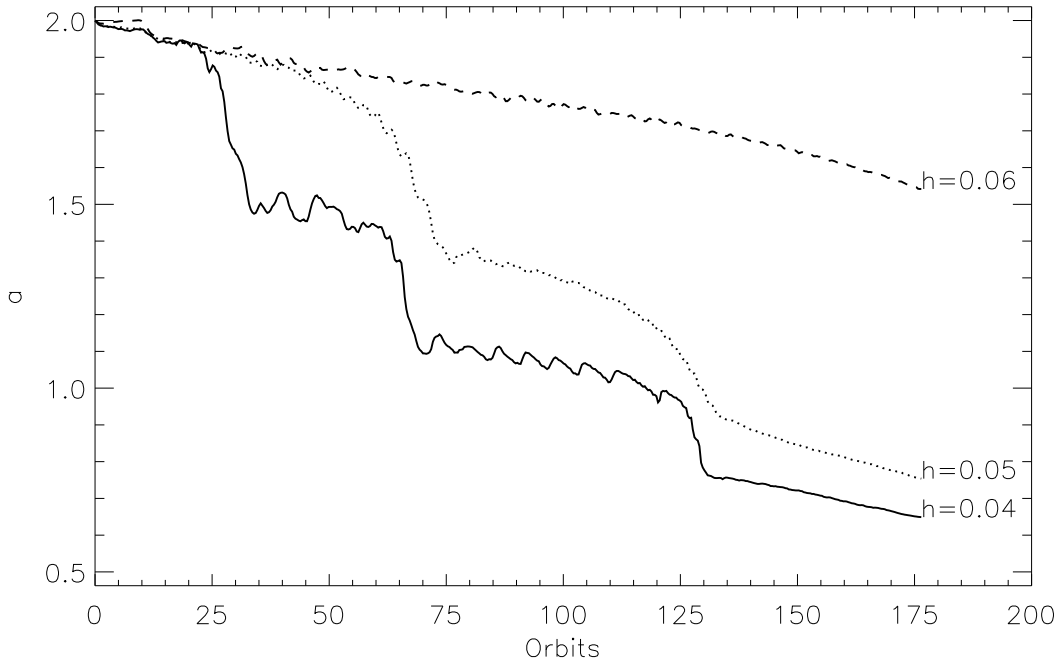


Figure 2.26: *Vortex-induced migration in inviscid discs with different aspect-ratio h but same initial uniform density $\Sigma = 7 \times 10^{-4}$.*

Fig. 2.25 where the inner edge remains stable. There is no vortex that grows to the critical mass to induce rapid migration. This implies that the disc temperature must be sufficiently low for vortex-planet interactions to occur.

2.6.5 Effect of planet mass: comparison to Jupiter

Recently, Pepliński et al. (2008a,b,c) performed numerical simulations of type III migration of a Jupiter-mass planet. Although Pepliński et al. did not include physical viscosity, their Cartesian grid is more diffusive and vortex-formation is suppressed (de Val-Borro et al., 2007). Hence, they obtained smooth migration curves. For completeness, the migration of Jupiter in the presence of vortices is presented below.

Fig. 2.27 compares the migration curves $a(t)$ for $M_p = 10^{-3}M_*$ in inviscid and viscous discs ($\nu_0 = 1, 2$) to the previous case of Saturn in a inviscid disc. Vortex-planet scattering also occur for Jupiter in the inviscid disc. This is seen in Fig. 2.28, where a vortex associated with the inner gap edge builds up and flows across the co-orbital region. Jupiter induces stronger shocks and higher-amplitude vortensity rings, and

2.6 The inviscid disc and parameter studies

are more unstable. The instability growth timescale is therefore shorter than that of Saturn, so vortex-induced migration occurs very soon after the planet is introduced.

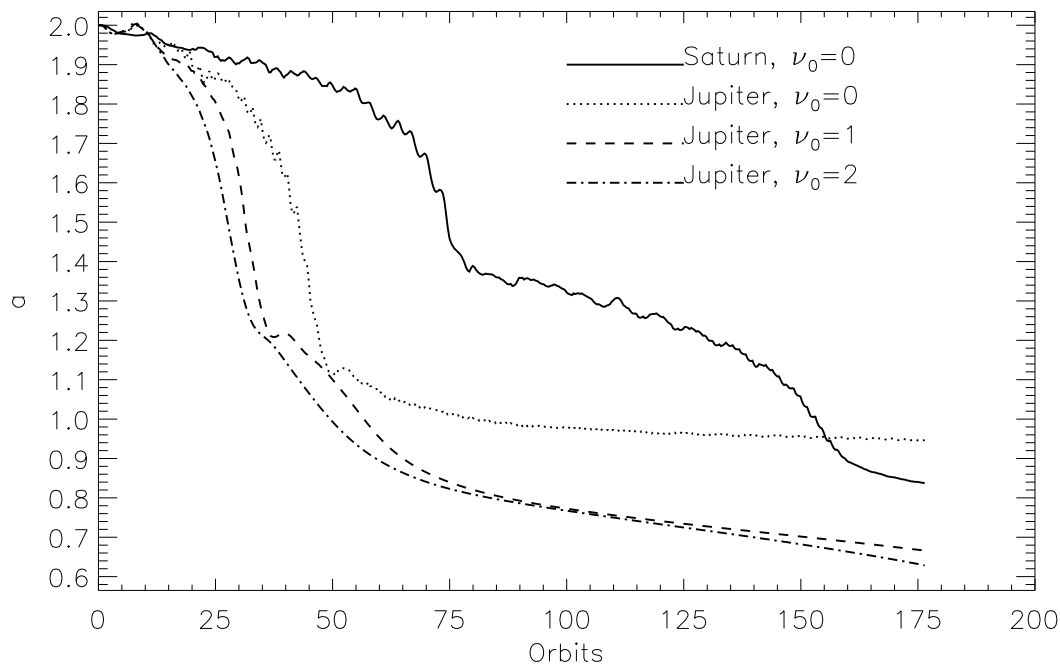


Figure 2.27: *Vortex-induced migration for a Jupiter-mass planet in an inviscid disc (dotted) and viscous disc ($\nu_0 = 1$, dashed; $\nu_0 = 2$, dashed-dot) compared to Saturn-mass planet in a inviscid disc (solid).*

The $a(t)$ for Jupiter in an inviscid disc show two migration phases — fast and slow, and that significant orbital decay occur within $\sim 50P_0$. These features were also observed by Pepliński et al. (2008b). However, the fast phase of Pepliński et al. is almost a linear function of time, here migration is clearly accelerating inwards during $t \lesssim 50P_0$. In this case there is an abrupt transition to the slow phase whereas that of Pepliński et al. is smoother and no vortices were identified.

Interestingly, standard viscosity $\nu_0 = 1$ results in a similar behaviour for the fast phase, though the transition to slow migration is smoother than zero viscosity. The vortensity distribution for $\nu_0 = 1$ was found to be much smoother than $\nu_0 = 0$, without vortices associated with the inner gap edge. This is consistent with de Val-Borro et al. (2007) who found that $\nu = O(10^{-5})$ suppress vortex formation. With $\nu_0 = 1$, flow-through the co-orbital region is a smoother function of time.

A case with $\nu_0 = 2$ is also shown in Fig. 2.27. The orbital decay timescale is again

2. VORTEX-INDUCED TYPE III MIGRATION

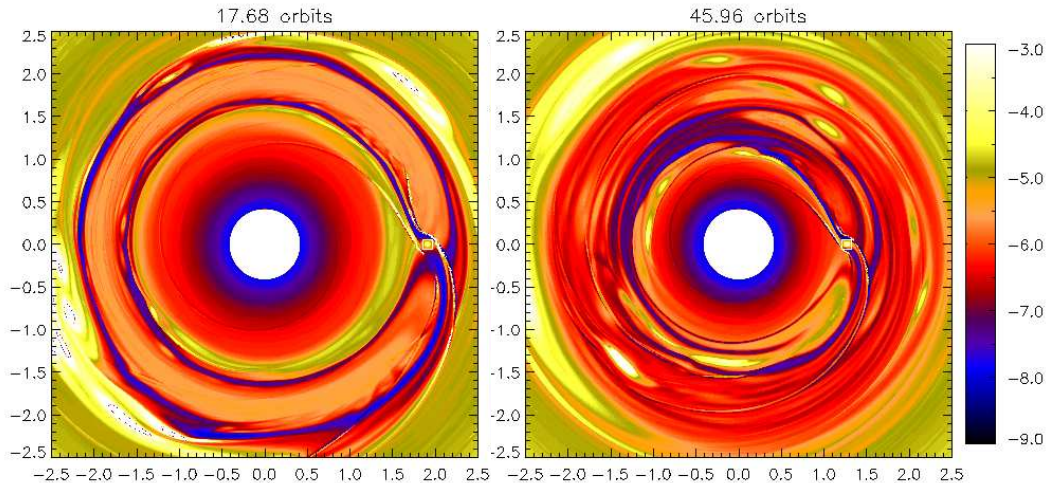


Figure 2.28: *Vortex-induced migration of a Jupiter mass planet in an inviscid disc. Maps of $\ln(\Sigma/\omega)$ are shown. As with Saturn-mass planets, rapid migration is associated with the flow of a vortex across the co-orbital region, simultaneously destroying the thin vortensity rings.*

$\sim 50P_0$, consistent with type III migration, but there could be complications from the fact that a Jovian-mass planet in a viscous disc can undergo type II migration. The transition from fast to slow is smoothed (cf. $\nu_0 = 0, 1$) and migration is qualitatively closer to Pepliński et al. (2008b). These results suggest that inviscid cases here have lower total viscosity than those considered by Pepliński et al., the difference attributed to different numerical grids employed.

2.7 Additional simulations

The simulations above illustrate the physics of vortex-induced migration. It can be understood in the language of type III migration. However, its association with co-orbital flow potentially makes it sensitive to numerics near the planet.

The locally isothermal equation of state allow mass accumulation near the planet, and may lead to spurious torques from within the Hill sphere. The fiducial simulation above was repeated with the equation of state employed by Pepliński et al. (2008a), where the disc temperature increases close to the planet (Eq. 1.18). This reduces mass accumulation onto the planet.

For the fiducial case, employing this EOS with the parameter $h_p = 0.6$ gives a sound-speed about 18% higher than the locally isothermal value at the planet’s initial

position. When $h_p = 0.7$, the increase is about 38%. Migration curves for these h_p are compared to the locally isothermal case in Fig. 2.29. These again yields vortex-planet scattering, which is an indication that significant torque contribution during the interaction is not sensitive to flow close to or within the Hill sphere.

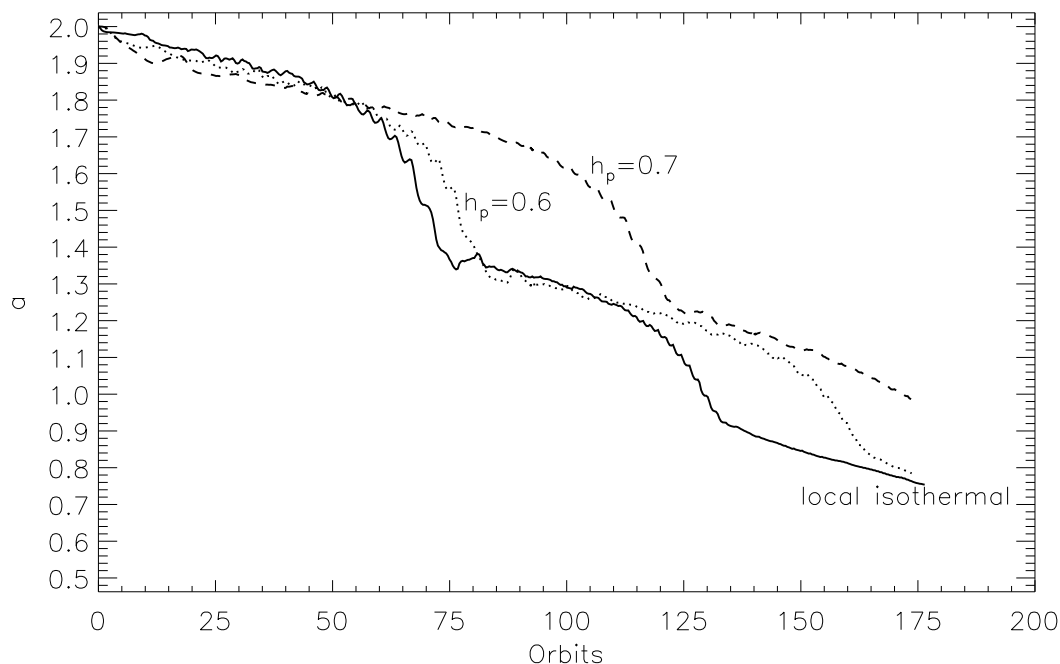


Figure 2.29: *Vortex-induced migration of a Saturn mass planet in inviscid discs with the modified equation of state of Pepliński et al. (2008a). The parameter h_p is a measure of the artificial temperature increase near the planet, relative to the locally isothermal equation of state. curves.*

Finally, there is the issue of softening length. Simulations previous to this section all used $\epsilon_p = 0.6H$. In Fig. 2.30, migrating cases with softening lengths of $0.5H$ and $0.7H$ are displayed. Both cases display vortex-planet interaction, although at earlier times when ϵ is lower. This is expected since smaller softening has the same effect as increasing the planet mass, and thus producing a more unstable disc.

2.8 Summary and discussion

Numerical simulations in this Chapter show that migration of a giant planet induced by large-scale vortices is a robust phenomenon in low viscosity discs. This vortex-planet

2. VORTEX-INDUCED TYPE III MIGRATION

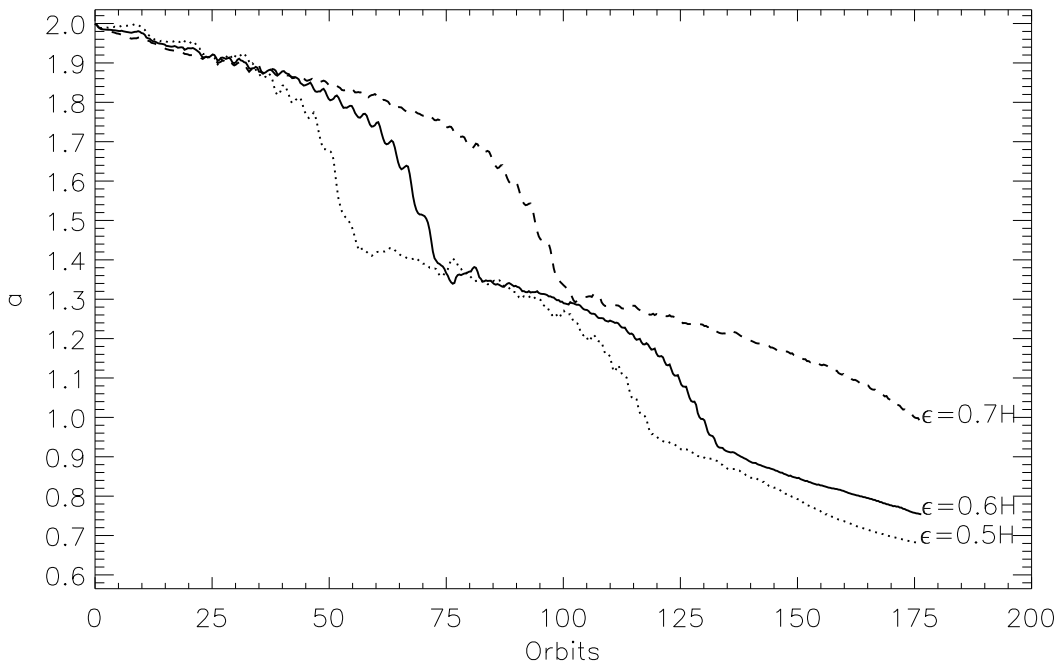


Figure 2.30: *Vortex-induced migration of a Saturn mass planet in inviscid discs with different softening lengths for the planet potential.*

interaction is best understood as a non-smooth form of type III migration.

Planets more massive than about Saturn induce spiral shocks extending into its co-orbital region and generate vortensity rings that delineate the gap. Local vortensity minima associated with gap edges are dynamically unstable to non-axisymmetric perturbations. Dominant unstable modes are radially localised. They develop into vortices and merge, on dynamical timescales, into a single vortex circulating either gap edge.

The effect of vortices on migration is most significant in low viscosity discs ($\nu \lesssim 0.25 \times 10^{-5}$) because the instability is suppressed at higher viscosity¹. The presence of high density vortices at the gap edge produces nonsmooth migration via vortex-planet scattering, with episodes of fast migration where the planet’s orbital radius jumps by a few Hill radii per episode. This is analogous to planet-planet scattering. The vortex is also responsible for stalling migration in discs with initially flat surface density. In this

¹In the inviscid limit only a small numerical viscosity is present. Suppose one considers a physical viscosity of $\nu = O(10^{-6})$, it is still large compared to the numerical viscosity, so one expects the latter to be unimportant. However, $\nu = O(10^{-6})$ is already unable to suppress the instability. Results are therefore unaffected by numerical viscosity of the FARGO code.

case there can be repeated episodes of vortex-induced migration. Viscosity smooths the flow across the co-orbital radius, but has limited effect on the extent of orbital decay via the type III mechanism.

The extent of orbital decay in a single episode of fast migration is independent of surface density scaling. This suggests a critical vortex mass or surface density is required to interact, which can be linked to the concept of the co-orbital mass deficit that drives type III migration. The effect of the background surface density profile, disc aspect-ratio and planetary mass were also examined. The effect of vortices in these cases are all consistent with linear analysis of gap stability and interpretation of vortex-induced migration as a form of type III migration.

2.8.1 Outstanding issues

The vortex instability requires low viscosity. It is then natural to question their existence as long-lived structures in real discs with finite viscosity. Recently, Li et al. (2009) studied migration in nearly-laminar discs in the type I regime. They found effects of lowering viscosity, begin to appear for $\nu \leq 10^{-7}$. It is unclear if viscosity in real discs will be low enough for vortices to develop and affect planets that undergo type I migration. For giant planets considered here, non-smooth migration begins at $\nu = O(10^{-6})$. Vortices can thus interfere with giant planet migration when viscosity is not more than one order of magnitude smaller than the standard value $\nu = 10^{-5}$. This could be satisfied in ‘dead zones’ in protoplanetary discs (Gammie, 1996).

One physical issue is the lack of inclusion of disc self-gravity. It may be important when discussing type III migration since this operates in massive discs (Masset & Papaloizou, 2003), and the vortices here are over-densities. In any case, self-gravity is required for a self-consistent treatment of disc-planet interactions (Baruteau & Masset, 2008). In the next Chapter, the effect of disc self-gravity on the vortex instability is explored.

2. VORTEX-INDUCED TYPE III MIGRATION

Chapter 3

Vortex modes in self-gravitating discs

A self-consistent protoplanetary disc model should account for its gravitational potential, or self-gravity. Even if the unperturbed disc has low mass and can be considered non-self-gravitating, if high density structures such as vortices develop, it may still be important to include self-gravity. In this Chapter, the effects of disc self-gravity on vortex forming instabilities at planetary gaps are examined.

The disc-planet models are first described in §3.1. The linear stability problem is then discussed in §3.2 and linear numerical calculations presented in §3.3. Vortex modes are stabilised by self-gravity through linear perturbations, but destabilised by its effect through the background gap structure. The above effects cause the most unstable vortex modes to shift to higher azimuthal wavenumber m , and hence more vortices to develop, as the disc mass increases. Nonlinear hydrodynamic simulations presented in §3.4 confirm this. Vortices can become self-gravitating with strong over-surface densities. This enables pairs of vortices to undergo coorbital dynamics with horseshoe trajectories, allowing them to survive much longer against merging than in the non-self-gravitating limit. This is strongly suggested by supplementary simulations in §3.5, of interacting Kida-like vortices without an embedded planet.

Most of this study considers fixed-orbit simulations. However, some migrating cases are discussed in §3.6. As in the non-self-gravitating cases (Chapter 2), episodes of rapid migration occur as vortices are scattered by the planet, but longer time intervals between them are expected on account of stabilisation by self-gravity and slowed vortex merging. Finally, results are summarised in §3.7.

3. VORTEX MODES IN SELF-GRAVITATING DISCS

3.1 Disc-planet model

The models for disc-planet systems are described here. Supplementary simulations in §3.5 employ a different set up from that described below.

The protoplanetary disc is now self-gravitating, but still considered two-dimensional. The governing fluid equations are Eq. 1.19, with the disc gravitational potential Φ given by the Poisson integral, Eq. 1.21. The softening length adopted for the disc potential in linear and nonlinear calculations is $\epsilon_g = 0.3H$. The equation of state is locally isothermal (Eq. 1.8) with fixed aspect-ratio $h = 0.05$. The disc has a uniform kinematic viscosity ν .

The disc occupies $r = [r_i, r_o] = [1, 10]$. The initial surface density profile is modified from Armitage & Hansen (1999):

$$\Sigma(r) = \Sigma_0 r^{-3/2} \left(1 - \sqrt{\frac{r_i}{r + H_i}} \right), \quad (3.1)$$

where $H_i = H(r_i)$ is introduced to prevent a singular pressure force at the inner boundary because $\Sigma \rightarrow 0$ there. Determining the disc potential via the Poisson integral does not require explicit boundary conditions to be imposed, but because the integral is only taken over the computational domain, it implicitly assumes no contribution to Φ from disc material outside the computational domain ($\Sigma = 0$ there). Hence, the use of Eq. 3.1, with surface density tapering towards radial boundaries, is for consistency with the Poisson integral.

The surface density scale Σ_0 is chosen via the parameter $Q_o \equiv Q_{\text{Kep}}(r_o)$, where

$$Q_{\text{Kep}}(r) = \frac{c_s \kappa}{\pi G \Sigma} = \frac{h M_*}{\pi r^2 \Sigma(r)}, \quad (3.2)$$

is the Toomre Q parameter for razor thin Keplerian discs with a locally isothermal equation of state. Note that specifying Σ_0 this way renders Q_{Kep} independent of h . Q_o is used to label disc models and $Q_p \equiv Q_{\text{Kep}}(r_p)$ is used to indicate gravitational stability at the planet's initial location. Specifying Q_o also determines the disc-to-star mass ratio, M_d/M_* .

The disc is initialised with the azimuthal velocity required by centrifugal balance. The initial radial velocity is $u_r = 3\nu/r$, corresponding to the initial radial velocity profile of a Keplerian disc with uniform kinematic viscosity and surface density $\propto r^{-3/2}$. Motivated by results in Chapter 2, a viscosity $\nu = 10^{-6}$ in code units is chosen to enable vortex instabilities in disc-planet interactions.

3.2 Linear modes in discs with structured surface density

Most of this Chapter will focus on the stability of the gap induced by an embedded planet. Accordingly the planet is fixed on a circular orbit of radius $r_p = 5$. The planet is introduced at $t = 25P_0$ with azimuthal velocity that takes account the contribution from the gravitational force due to the disc. The planet has mass $M_p = 3 \times 10^{-4}M_*$ and its potential is ramped up over a time period of $10P_0$.

3.2 Linear modes in discs with structured surface density

The stability of interest is that of a surface density gap self-consistently opened by a giant planet in a disc in which self-gravity is not neglected. To simplify linear theory, viscosity, indirect potentials and the planet potential are ignored. As shown in Chapter 2, the planet's role is only to set up the basic state, assumed to be axisymmetric and defined by $\Sigma(r)$, $u_\phi(r)$ and $u_r = 0$.

As usual, perturbations to the disc state variables are assumed to have azimuthal and time dependence through a factor $\exp i(\sigma t + m\phi)$ which is taken as read. Denoting locally isothermal perturbations by δ , the linearised equations of motion for a spatially fixed sound-speed profile gives

$$\delta u_r = -\frac{1}{D} \left[i\bar{\sigma} \left(c_s^2 \frac{dW}{dr} + \frac{d\delta\Phi}{dr} \right) + \frac{2im\Omega}{r} (c_s^2 W + \delta\Phi) \right] \quad (3.3)$$

$$\delta u_\phi = \frac{1}{D} \left[\frac{\kappa^2}{2\Omega} \left(c_s^2 \frac{dW}{dr} + \frac{d\delta\Phi}{dr} \right) + \frac{m\bar{\sigma}}{r} (c_s^2 W + \delta\Phi) \right], \quad (3.4)$$

where $D = \kappa^2 - \bar{\sigma}^2$. Recall $\bar{\sigma} \equiv \sigma + m\Omega(r)$ and $W = \delta\Sigma/\Sigma$. These equations are very similar to the non-self-gravitating case (Eq. 2.17–2.18), except now the disc gravitational potential perturbation $\delta\Phi$ is included, and is given by the linearised Poisson integral

$$\delta\Phi = -G \int_{r_i}^{r_o} K_m(r, r') \Sigma(r') W(r') r' dr', \quad (3.5)$$

$$K_m(r, r') = \int_0^{2\pi} \frac{\cos(m\phi) d\phi}{\sqrt{r^2 + r'^2 - 2rr' \cos(\phi) + \epsilon_g^2(r')}}. \quad (3.6)$$

Using the linearised equations of motion to eliminate the velocity component perturba-

3. VORTEX MODES IN SELF-GRAVITATING DISCS

tions from the linearised continuity equation (Eq. 2.19) yields the governing equation

$$\begin{aligned} \frac{d}{dr} \left[\frac{r\Sigma}{D} \left(c_s^2 \frac{dW}{dr} + \frac{d\delta\Phi}{dr} \right) \right] + \left[\frac{2m}{\bar{\sigma}} \left(\frac{\Sigma\Omega}{D} \right) \frac{dc_s^2}{dr} - r\Sigma \right] W \\ + \left[\frac{2m}{\bar{\sigma}} \frac{d}{dr} \left(\frac{\Sigma\Omega}{D} \right) - \frac{m^2\Sigma}{rD} \right] (c_s^2 W + \delta\Phi) = 0. \end{aligned} \quad (3.7)$$

3.2.1 Analysis for barotropic discs

The linear modes of interest are those found to be localised to gap edges and lead to vortex formation in the nonlinear regime. The effect of self-gravity on these modes can be evaluated analytically with appropriate simplifications. It will be shown that, paradoxically, in view of the fact that increasing self-gravity in general destabilise discs, that self-gravity tends to stabilise vortex modes. To do this, one considers a small change to the strength of self-gravity and apply perturbation theory to determine the consequence of the change for these modes.

To make the analysis tractable, further simplifications are made. The fluid is considered to be barotropic, so that $p = p(\Sigma)$. This includes the strictly isothermal case where c_s is a constant. There should be no significant difference between adopting a strictly isothermal, or barotropic equation of state, and the fixed c_s profile as used in nonlinear simulations. This is because the modes of interest are driven by local features and disturbances are localised, whereas $c_s(r)$ for the adopted locally isothermal equation of state varies on a global scale. For mathematical convenience, the softening prescription is assumed to be such that $K_m(r, r') = K_m(r', r)$ is symmetric, eg. $\epsilon_g = \text{constant}$. Although convenient mathematically, this is not expected to lead to significant changes for the same reason as that given above.

Introducing the variable S

$$S \equiv c_s^2 W + \delta\Phi = c_s^2 W - G \int_{r_i}^{r_o} K_m(r, r') r' \Sigma(r') W(r') dr', \quad (3.8)$$

the governing equation derived from Eq. 3.7 simply by taking c_s to be constant is

$$\begin{aligned} r\Sigma W = \frac{d}{dr} \left[\frac{r\Sigma}{D} \left(\frac{dS}{dr} \right) \right] + \left\{ \frac{m}{\bar{\sigma}} \frac{d}{dr} \left[\frac{1}{\eta(1 - \bar{\nu}^2)} \right] - \frac{m^2\Sigma}{rD} \right\} S \\ \equiv rL(S). \end{aligned} \quad (3.9)$$

where the expression for vortensity $\eta = \kappa^2/2\Omega\Sigma$ has been used, and $\bar{\nu} \equiv \bar{\sigma}/\kappa$. L is a linear operator such that $L(S) = \delta\Sigma$. In fact, these equations also hold for a general

barotropic equation of state.

3.2.2 The limit of negligible self-gravity

When self-gravity is negligible in the linear response, one may set $G = 0$ in Eq. 3.8, giving $W = S/c_s^2$. Substituting this into Eq. 3.9 gives the second order ordinary differential equation for S that governs stability in the limit of zero self-gravity in the form

$$\frac{r\Sigma S}{c_s^2} = \frac{d}{dr} \left[\frac{r\Sigma}{D} \left(\frac{dS}{dr} \right) \right] + \left\{ \frac{m}{\bar{\sigma}} \frac{d}{dr} \left[\frac{1}{\eta(1-\bar{v}^2)} \right] - \frac{m^2\Sigma}{rD} \right\} S. \quad (3.10)$$

This is the same equation studied in Chapter 2 when c_s is regarded constant. If one is concerned with the effect of weak self-gravity, then the vortex modes should still have co-rotation radius r_c , such that $\bar{\sigma}(r_c) = 0$, at or near a vortensity minimum, as found in Chapter 2.

3.2.3 Localised low m modes

When m is small the modes described above are localised around co-rotation and therefore insensitive to distant boundary conditions. These were the dominant modes found in Chapter 2, where self-gravity was ignored. It is then appropriate to adopt these simplifications to calculate the effect of weak self-gravity. For these modes, one can neglect \bar{v} and set $D = \kappa^2$ in Eq. 3.10 to get the simpler equation

$$\frac{r\Sigma S}{c_s^2} = \frac{d}{dr} \left[\frac{r\Sigma}{\kappa^2} \left(\frac{dS}{dr} \right) \right] + \left\{ \frac{m}{\bar{\sigma}} \frac{d}{dr} \left[\frac{1}{\eta} \right] - \frac{m^2\Sigma}{r\kappa^2} \right\} S. \quad (3.11)$$

Localisation occurs because the solutions of Eq. 3.11 can be seen to decay exponentially away from the co-rotation region, where there are large background gradients, on a length scale comparable to H . For Eq. 3.11 to be a good approximation, one requires $|\bar{v}^2| \ll 1$ in the region of localisation which is also comparable to H in extent. This in turn requires $H \ll 2r/(3m)$, a condition which is satisfied for low m . The analysis below assumes localisation so that boundary conditions do not play a role. It therefore only applies for low m .

3. VORTEX MODES IN SELF-GRAVITATING DISCS

3.2.4 Evaluating the effect of small changes to low m modes using perturbation theory

Eq. 3.8 and Eq. 3.9 can be used to investigate the effect of self-gravity. Note that strengthening self-gravity by scaling up the surface density is equivalent to increasing G , provided that the background form remains fixed. Thus although the effective disc gravity is increased, the background disc model shall be assumed unchanged. The background disc model is structured by a perturbing planet and so its response to changing the disc gravity is non-trivial to evaluate analytically. Calculations presented in §3.3.7 show that changes to the background surface density profile induced by incorporating the disc gravity tends to make the vortex modes more unstable. However, the direct effect of self-gravity through the linear response considered below turns out to be more important for localised modes with low m , and acts to stabilise them.

A solution S corresponding to a neutral mode ($\gamma = 0$) with co-rotation radius r_c located at a vortensity minimum, is first assumed ($\bar{\sigma}(r_c) = d\eta/dr|_{r=r_c} = 0$). As the associated σ is real, for this value, the operator L is real and regular everywhere. Now perturb this solution by altering the strength of self-gravity via

$$G \rightarrow G + \delta G$$

so that $\delta G > 0$ corresponds to increasing the importance of self-gravity and vice versa (note that the initial value, G , could be zero). This induces perturbations

$$S \rightarrow S + \delta S, \quad \delta\Sigma \rightarrow \delta\Sigma + \delta\Sigma_1, \quad \sigma \rightarrow \sigma + \delta\sigma, \quad L \rightarrow L + \delta L,$$

with

$$\delta\sigma = \delta\sigma_R + i\gamma, \quad \delta L = \frac{\partial L}{\partial \sigma} \delta\sigma,$$

where $\delta\sigma_R$ and γ are real. Noting that δ denotes a small change and γ is small, one can linearise in terms of these quantities about the assumed original neutral mode and determine γ . The governing equations lead to

$$L(\delta S) + \delta L(S) = \delta\Sigma_1, \tag{3.12}$$

$$\delta S = c_s^2 \frac{\delta\Sigma_1}{\Sigma} - G \int_{r_i}^{r_o} K_m(r, r') r' \delta\Sigma_1(r') dr' - \delta G \int_{r_i}^{r_o} K_m(r, r') r' \delta\Sigma(r') dr'. \tag{3.13}$$

3.2 Linear modes in discs with structured surface density

Next, define the inner product between two functions $U(r)$, $V(r)$ as

$$\langle U, V \rangle \equiv \int_{r_i}^{r_o} r U^*(r) V(r) dr, \quad (3.14)$$

Then, assuming localised functions corresponding to localised modes so that boundary values can be assumed to vanish when integrating by parts,

$$\langle U, L(V) \rangle = \langle L(U), V \rangle, \quad (3.15)$$

which used the fact that L corresponding to the neutral mode is real, making L self-adjoint. Now consider

$$\langle L(S), \delta S \rangle = \langle \delta \Sigma, \delta S \rangle = \langle S, L(\delta S) \rangle. \quad (3.16)$$

The inner products evaluate to

$$\begin{aligned} & \langle \delta \Sigma, \delta S \rangle \\ &= \int_{r_i}^{r_o} r \delta \Sigma^*(r) \left[c_s^2 \frac{\delta \Sigma_1(r)}{\Sigma(r)} - G \int_{r_i}^{r_o} K_m(r, r') r' \delta \Sigma_1(r') dr' - \delta G \int_{r_i}^{r_o} K_m(r, r') r' \delta \Sigma(r') dr' \right] dr, \end{aligned} \quad (3.17)$$

and

$$\begin{aligned} & \langle S, L(\delta S) \rangle = \langle S, \delta \Sigma_1 - \delta L(S) \rangle \\ &= \int_{r_i}^{r_o} r \left[c_s^2 \frac{\delta \Sigma^*(r)}{\Sigma(r)} - G \int_{r_i}^{r_o} K_m(r, r') r' \delta \Sigma^*(r') dr' \right] \delta \Sigma_1(r) dr - \langle S, \delta L(S) \rangle. \end{aligned} \quad (3.18)$$

These expressions can be equated to yield

$$\begin{aligned} \langle S, \delta L(S) \rangle &= \delta G \int_{r_i}^{r_o} \int_{r_i}^{r_o} r r' K_m(r, r') \delta \Sigma^*(r') \delta \Sigma(r) dr' dr. \\ &\equiv \delta G E, \end{aligned} \quad (3.19)$$

where symmetry of the kernel K_m was used. Note that $E > 0$ and is proportional to the magnitude of the gravitational energy of the mode. Following Papaloizou & Lin (1989), one separates out the contribution to the perturbed linear operator δL that is proportional to the vortensity gradient (and potentially singular at co-rotation) and

3. VORTEX MODES IN SELF-GRAVITATING DISCS

write

$$\langle S, \delta L(S) \rangle \equiv \langle S, \delta L_1(S) \rangle + \langle S, \delta L_2(S) \rangle, \quad (3.20)$$

where $\delta L_2(S)$ contains the potentially singular contribution:

$$\langle S, \delta L_2(S) \rangle = -m\delta\sigma \left[\mathcal{P} \left(\int_{r_i}^{r_o} \frac{g(r)}{\bar{\sigma}} dr \right) + \int_{r_i}^{r_o} i\pi\delta(\bar{\sigma})g(r)dr \right] \quad (3.21)$$

$$g(r) \equiv \frac{|S|^2}{\bar{\sigma}} \frac{d}{dr} \left(\frac{1}{\eta} \right). \quad (3.22)$$

Note that the usual Landau prescription for negative γ (growing mode) has been used to deal with the pole at co-rotation and that \mathcal{P} outside the above integral indicates that the principal value is to be taken. Note too that δL_1 accounts for the remainder of δL and is not singular at co-rotation. The contribution from δL_2 is the only one that can lead to an imaginary contribution in the limit $\delta\sigma \rightarrow 0$ because of the pole at co-rotation. No such contributions arise from Lindblad resonances where $D = 0$. This is because, as is well known, these do not constitute effective singularities in a gaseous disc (eg. Papaloizou & Savonije, 1991).

Recalling that γ is small, Eq. 3.19- 3.21 can be combined to give

$$\begin{aligned} \delta GE &= \delta\sigma \left[\frac{\partial \langle S, L_1(S) \rangle}{\partial \sigma} - m\mathcal{P} \left(\int_{r_i}^{r_o} \frac{g(r)}{\bar{\sigma}} dr \right) - im\pi \int_{r_i}^{r_o} \delta(\bar{\sigma})g(r)dr \right] \\ &\equiv (\mathcal{A} - im\zeta)\delta\sigma \end{aligned} \quad (3.23)$$

where \mathcal{A} is the contribution from the L_1 term plus the principle value integral and

$$\zeta \equiv \frac{\pi|S|^2}{m\Omega'|m\Omega'|} \frac{d^2}{dr^2} \left(\frac{1}{\eta} \right) \Big|_{r=r_c}. \quad (3.24)$$

In the limit $\gamma \rightarrow 0_-$, \mathcal{A} is real so

$$\gamma = \frac{m\zeta}{\mathcal{A}^2 + m^2\zeta^2} \delta GE. \quad (3.25)$$

Vortex forming modes are associated with vortensity minima (or maxima of η^{-1}). Consider a marginally stable mode with co-rotation at $\max(\eta^{-1})$. Typical rotation profiles have $\Omega' < 0$, which means $\zeta > 0$ for this mode. For consistency with the assumption of $\gamma < 0$ in deriving Eq. 3.25, one requires $\delta G < 0$ since $E > 0$. Accordingly in order to destabilise this mode, the strength of self-gravity needs to be reduced.

3.2 Linear modes in discs with structured surface density

This leads to the conclusion that: increasing self-gravity *stabilises* low m modes with co-rotation at a vortensity *minimum* (as has been borne out by linear and nonlinear calculations presented below); while increasing self-gravity would *destabilise* modes which had co-rotation at a vortensity *maximum*. This suggests that for sufficiently strong self-gravity, modes associated with vortensity maxima should be favoured. This has been found to be the case but they are not modes leading to vortex formation (see Chapter 4).

Note $\zeta \propto 1/m^2$, suggesting that $|\gamma|$ decreases for large m , so the stabilisation or destabilisation effect of self-gravity diminishes for increasing azimuthal wave-number. This is only speculative at this point, because there are implicit dependencies on m through terms in the integrals and through the original eigenfunction S . Nevertheless, a weakening effect of self-gravity through the potential perturbation is anticipated because increasing m decreases the magnitude of the Poisson kernel K_m .

3.2.5 Association of localised low m normal modes with vortensity minima for the strength of self-gravity below a threshold

Co-rotation radii for localised neutral modes with low m should be at vortensity minima unless self-gravity (or an appropriate mean value of Q) is above (below) a threshold level. Consider a disc with localised steep vortensity gradients and a non-axisymmetric disturbance with co-rotation radius in this region. Multiply Eq. 3.9 by S^* and integrating over the disc. Assuming most of the contribution is from near co-rotation, as is expected for low m modes (see §3.2.3), so that the term that is potentially singular at corotation and proportional to the vortensity gradient is dominant on the RHS, implies the balance

$$\int_{r_i}^{r_o} r c_s^2 \frac{|\delta\Sigma|^2}{\Sigma} dr - G \int_{r_i}^{r_o} \int_{r_i}^{r_o} K_m(r, r') r r' \delta\Sigma^*(r') \delta\Sigma(r) dr' dr \simeq \int_{r_i}^{r_o} \frac{m |S|^2}{\bar{\sigma}} \frac{d}{dr} \left(\frac{1}{\eta} \right) dr. \quad (3.26)$$

If the LHS of Eq. 3.26 can be shown to always be positive then co-rotation for a localised neutral mode must lie at a vortensity minimum, or $\max(\eta^{-1})$ ¹. This holds if the maximum possible value of Λ for any $\delta\Sigma$ is less than unity, where

$$\Lambda = \frac{G \int_{r_i}^{r_o} \int_{r_i}^{r_o} K_m(r, r') r r' \delta\Sigma^*(r') \delta\Sigma(r) dr' dr}{\int_{r_i}^{r_o} r c_s^2 \frac{|\delta\Sigma|^2}{\Sigma} dr} \quad (3.27)$$

¹The RHS is positive if the dominant contribution to the integral comes from the co-rotation ($\bar{\sigma} = 0$) region about a vortensity minimum.

3. VORTEX MODES IN SELF-GRAVITATING DISCS

is the ratio of the magnitude of gravitational potential energy to thermal energy of the disturbance. Consider re-writing the numerator of Λ as

$$\begin{aligned}
& G \int_{r_i}^{r_o} \int_{r_i}^{r_o} \left[\frac{K_m(r, r') \sqrt{r} \sqrt{r'} \Sigma^{1/2}(r) \Sigma^{1/2}(r')}{c_s(r) c_s(r')} \right] \left[\frac{\sqrt{r} c_s(r) \delta \Sigma(r)}{\Sigma^{1/2}(r)} \right] \left[\frac{\sqrt{r'} c_s(r') \delta \Sigma^*(r')}{\Sigma^{1/2}(r')} \right] dr dr' \\
&= \int_{r_i}^{r_o} \frac{\sqrt{r} c_s(r) \delta \Sigma(r)}{\Sigma^{1/2}(r)} \left\{ G \int_{r_i}^{r_o} \left[\frac{K_m(r, r') \sqrt{r} \sqrt{r'} \Sigma^{1/2}(r) \Sigma^{1/2}(r')}{c_s(r) c_s(r')} \right] \left[\frac{\sqrt{r'} c_s(r') \delta \Sigma^*(r')}{\Sigma^{1/2}(r')} \right] dr' \right\} dr \\
&= \int_{r_i}^{r_o} f_1(r) \{f_2^*(r)\} dr. \tag{3.28}
\end{aligned}$$

The Cauchy-Schwartz inequality is

$$\left| \int_{r_i}^{r_o} f_1(r) f_2^*(r) dr \right|^2 \leq \int_{r_i}^{r_o} |f_1|^2 dr \times \int_{r_i}^{r_o} |f_2|^2 dr.$$

So that Eq. 3.27 implies

$$\begin{aligned}
& \Lambda^2 \left(\int_{r_i}^{r_o} r c_s^2 \frac{|\delta \Sigma|^2}{\Sigma} dr \right)^2 \leq \int_{r_i}^{r_o} \frac{r c_s^2(r) |\delta \Sigma(r)|^2}{\Sigma(r)} dr \\
& \times \int_{r_i}^{r_o} \left| G \int_{r_i}^{r_o} \left[\frac{K_m(r, r') \sqrt{r} \sqrt{r'} \Sigma^{1/2}(r) \Sigma^{1/2}(r')}{c_s(r) c_s(r')} \right] \left[\frac{\sqrt{r'} c_s(r') \delta \Sigma^*(r')}{\Sigma^{1/2}(r')} \right] dr' \right|^2 dr. \tag{3.29}
\end{aligned}$$

Applying the Cauchy-Schwartz inequality to the second line,

$$\begin{aligned}
& \Lambda^2 \left(\int_{r_i}^{r_o} r c_s^2 \frac{|\delta \Sigma|^2}{\Sigma} dr \right)^2 \leq \int_{r_i}^{r_o} \frac{r c_s^2(r) |\delta \Sigma(r)|^2}{\Sigma(r)} dr \\
& \times \int_{r_i}^{r_o} \left\{ G^2 \int_{r_i}^{r_o} \frac{K_m^2(r, r') r r' \Sigma(r) \Sigma(r')}{c_s^2(r) c_s^2(r')} dr' \right\} \left\{ \int_{r_i}^{r_o} \frac{r' c_s^2(r') |\delta \Sigma(r')|^2}{\Sigma(r')} dr' \right\} dr. \tag{3.30}
\end{aligned}$$

The integrals involving $\delta \Sigma$ on either sides cancel. It follows that

$$\Lambda \leq G \sqrt{\int_{r_i}^{r_o} \int_{r_i}^{r_o} K_m^2(r, r') \frac{\Sigma(r) \Sigma(r')}{c_s^2(r) c_s^2(r')} r r' dr' dr}. \tag{3.31}$$

Thus if the RHS of Eq. 3.31 is less than unity, corotation of a neutral mode localised at a vortensity extremum must be localised at a vortensity minimum. This condition requires that the strength of self-gravity be below a threshold and this implies a lower bound on an appropriate mean Q value. The fact that this fails for sufficiently strong self-gravity is consistent with the discussion above that led to the conclusion that

increasing self-gravity tends to stabilise the vortex forming instability.

Indeed when self-gravity increases further, instability is transferred to modes with corotation associated with vortensity maxima. These are different in character to vortex forming modes being more global and are referred to as edge modes and they are studied in Chapter 4.

3.3 Linear calculations

Numerical solutions to the linear problem (Eq. 3.7) are now presented. The locally isothermal equation of state $p = h^2 GM \Sigma / r \equiv c_s^2 \Sigma$ and spatially varying softening length $\epsilon_g(r) = 0.3H(r)$, are restored for consistency with numerical simulations used to set up the basic state. The governing equation is regarded as the requirement that an operator \mathcal{L} acting on W should be zero, thus

$$\mathcal{L}(W) = 0. \quad (3.32)$$

One form of \mathcal{L} is

$$\begin{aligned} \mathcal{L} = & r c_s^2 \frac{d^2}{dr^2} + r \mathcal{J}_2 + \left[\frac{1}{r} + \frac{\Sigma'}{\Sigma} - \frac{D'}{D} \right] \left(r c_s^2 \frac{d}{dr} + r \mathcal{J}_1 \right) + r c_s^{2'} \frac{d}{dr} + \left[\frac{2m\Omega c_s^{2'}}{\bar{\sigma}} - rD \right] \\ & + \left\{ \frac{2m}{\bar{\sigma}} \left[\frac{\Sigma' \Omega}{\Sigma} + \Omega' - \frac{\Omega D'}{D} \right] - \frac{m^2}{r} \right\} (c_s^2 + \mathcal{J}_0). \end{aligned} \quad (3.33)$$

Primes in Eq. 3.33 denote d/dr . The integro-differential operators \mathcal{J}_n are such that

$$\mathcal{J}_n(W) = \frac{d^n \delta \Phi}{dr^n}. \quad (3.34)$$

First and second derivatives of the perturbed potential are performed by replacing $K_m(r, r')$ by $\partial K_m / \partial r$ and $\partial^2 K_m / \partial r^2$ in the Poisson integral respectively.

3.3.1 Quadratic approximation

Vortex modes are localised about co-rotation at a vortensity minimum. They extend over a region characterised by small $|\bar{\sigma}|$ in the neighbourhood of r_c and are expected to be insensitive to boundary conditions. Based on these observations, a crude approximation was adopted in the analytical discussion in Chapter 2, where $\bar{\sigma}$ is neglected everywhere except when it appears as $1/\bar{\sigma}$. This enforces locality. However, with self-gravity, the disturbance is potentially more global. A better approximation for

3. VORTEX MODES IN SELF-GRAVITATING DISCS

self-gravitating vortex modes can be made as follows.

Consider multiplying Eq. 3.33 by $\bar{\sigma}D$ and expand the resulting equation in powers of $\bar{\sigma}$. Vortex modes are associated with the region $|\bar{\sigma}| \sim 0$. By neglecting terms proportional to $\bar{\sigma}^3$ and higher powers, one can derive an equation of the form

$$(\sigma^2 \mathcal{L}_2 + \sigma \mathcal{L}_1 + \mathcal{L}_0) W = 0, \quad (3.35)$$

where the operators \mathcal{L}_i are real and only depend on the basic state. This is referred to as the quadratic approximation. Expressions for \mathcal{L}_i are given in Appendix D. The eigenfrequency σ now appears explicitly.

The above procedure is not expected to be valid in general, but it should be valid for modes approximately localised about their corotation radii with the scale of localisation being much less than r_c itself. Solution eigenvalues from Eq. 3.35 can be used as starting values in order to obtain an iterative solution for the eigenvalues for the full Eq. 3.32. The concept of the vortex forming mode as a localised mode *is confirmed* if the final solution of Eq. 3.32 is not significantly changed from the solution of Eq. 3.35. Note that since the operators \mathcal{L}_i are real, eigenfrequencies are either real or occur in complex conjugate pairs.

3.3.2 Distinction between including and excluding self-gravity

It is important to clearly define self-gravitating (SG) and non-self-gravitating cases (NSG). The background state on which linear stability analysis is performed were obtained from nonlinear hydrodynamic simulations (see §3.3.4), which can be run with self-gravity (SGBG) or without self-gravity (NSGBG). In the linear calculations, self-gravity can be disabled by setting $\delta\Phi = 0$, corresponding to two disturbance types: a self-gravitating response (SGRSP) and a non-self-gravitating response (NSGRSP).

The analytical discussion in §3.2.1 applies to the effect of self-gravity through the linear response, assuming the form of the background remains unchanged. In the context of hydrodynamic simulations, including or neglecting self-gravity simply means whether or not the disc gravitational potential is included. If it is included, the background state set up by simulations also depends on self-gravity. Hydrodynamic simulations therefore correspond either to the combination SGBG+SGRSP or to the combination NSGBG+NSGRSP. These are the fully self-gravitating and fully non self-gravitating cases. The advantage of performing linear calculations is that one can distinguish between the effects of self-gravity arising through its effects on the background state and the effects resulting from its influence on the linear response.

3.3.3 Numerical approach

The linear operators are discretised on a grid that divides the radial range $[r_i, r_o]$ into $n_r = 385$ equally spaced grid points at which W is evaluated as $W_j, j = 1, 2..n_r$. The governing equation thus becomes a system of n_r simultaneous equations for the W_j :

$$\sum_{j=1}^{n_r} \mathcal{L}_{ij} W_j = 0, \quad (3.36)$$

where \mathcal{L}_{ij} is the discretised version of the integro-differential operator \mathcal{L} . Eq. 3.36 is a matrix eigenvalue problem for σ . The system of equations incorporates the boundary conditions (replacing the first and last rows of the matrix \mathcal{L}_{ij}). For simplicity, $dW/dr = 0$ is imposed at boundaries. It is known from other linear calculations and simulations that vortex-forming modes are localised and insensitive to boundary conditions (e.g. de Val-Borro et al., 2007). Tests have shown that the boundary condition does not influence the essential effect of self-gravity on localised modes.

The discretised quadratic approximation obtained from Eq. 3.35 leads to a quadratic eigenvalue problem that is equivalent to a $2n_r \times 2n_r$ linear matrix eigenvalue problem. This is solved by standard matrix software to give all eigenvalues σ . The most unstable eigenvalue is then used as a trial solution for the full system (Eq. 3.36), of which non-trivial solutions demand:

$$\det \mathcal{L}_{ij} = 0. \quad (3.37)$$

Although the determinant can be computed without much effort, numerically testing a zero determinant is ambiguous because the value of $\det \mathcal{L}_{ij}$ can change by many orders of magnitude depending on how the original equation is scaled, or equivalently, the choice of units. To obtain a non-dimensional measure of a zero determinant, independent of how the matrix elements \mathcal{L}_{ij} are scaled, σ can be considered as a parameter for the matrix \mathcal{L}_{ij} in the standard eigenvalue problem:

$$\sum_{j=1}^{n_r} \mathcal{L}_{ij}(\sigma) W_j = \lambda W_i. \quad (3.38)$$

Then $\lambda = 0$ is required to be a solution for some σ . Given a trial σ , all the eigenvalues λ can be found using standard methods. The actual eigenvalue, σ , is then solved iteratively using the Newton Raphson method to zero the quantity $\lambda_{\min}/|\lambda_{\max}|$, where subscript ‘min’ and ‘max’ denote the λ of smallest and largest magnitudes, respectively.

3. VORTEX MODES IN SELF-GRAVITATING DISCS

To check the obtained σ indeed makes \mathcal{L} singular, a singular value decomposition (SVD) of \mathcal{L} is performed, and the reciprocal of the condition number computed. This number is zero for a singular matrix. The SVD also gives the solution vector W_i that corresponds to the zero singular value, i.e. the eigenfunction W of Eq. 3.7. Numerical solutions presented below typically have a inverse condition number of $O(10^{-18})$.

3.3.4 Background state

The background states used for linear calculations were set up by running disc-planet simulations for the models described in §3.1. Numerical details will be described in §3.4. The planet is allowed to open a (stable) gap, then azimuthal averages performed to obtain one-dimensional profiles.

Fig.3.1 shows the gap profile, in terms of the vortensity η , opened up by a Saturn-mass planet in a disc with $Q_o = 4$. The formation of vortensity peaks was explained in Chapter 2 as generation of vortensity through shocks induced by the planet. The profiles set up with and without self-gravity are very similar so the formation mechanism is not affected by disc self-gravity. When self-gravity is included, the peaks and troughs have slightly larger amplitudes due to increased effective planet mass (see §3.4.3).

Hydrodynamic simulations indicate vortices predominantly develop at the outer gap edge, henceforth linear modes associated with the outer vortensity extrema are considered. The outer vortensity maximum is located at $r \simeq 5.5$ while the outer vortensity minimum is located at $r \simeq 5.75$. These extrema are separated by about $0.89H$ and are therefore local features.

3.3.5 Solution in the quadratic approximation

Solutions for linear modes in the $Q_o = 4$ disc under the quadratic approximation are shown in Fig. 3.2, which compares the growth rates, γ , and the eigenfunctions, W , obtained for $m = 5$ for the fully self-gravitating case and the fully non self-gravitating case. Growth rates are such that the most unstable mode shifts to higher m when self-gravity is included. Without self-gravity, the dominant mode has $m = 4$, whereas with self-gravity, the dominant mode has $m = 6$ — 7 . The combination of self-gravity acting through the background and the linear response stabilises modes with $m \leq 5$ and destabilises modes with $m \geq 6$. Thus higher m vortex modes are enabled by self-gravity.

The eigenfunctions W with and without self-gravity are similar, and are for the most part localised about co-rotation. This suggests the nature of the instability is

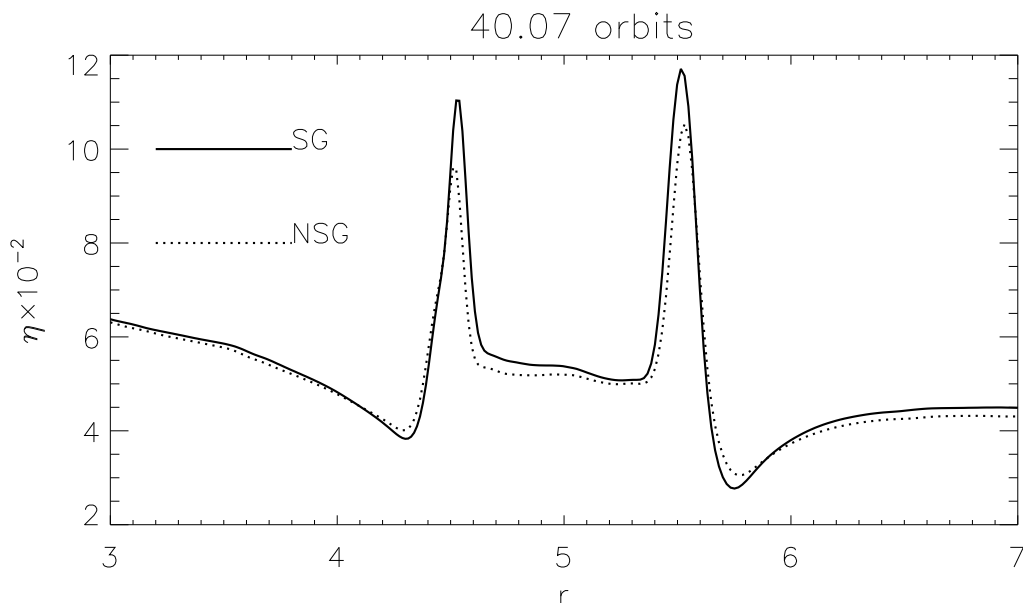


Figure 3.1: Gap profiles opened by a Saturn-mass planet in an initial disc model with $Q_o = 4$. The vortensity, η , obtained with (solid, ‘SG’) and without (dotted, ‘NSG’) self-gravity is shown. Profiles were obtained from azimuthal averages taken from disc-planet simulation outputs. The planet is fixed at $r = 5$.

unchanged in this case. However, the eigenfunction amplitude in $r > 6$, $r < 5.2$ is larger when self-gravity is included. Note too that the amplitude is larger for $r > 6$ than for $r < 5.2$. This is expected since the outer disc has smaller Toomre Q values it is expected to be more responsive.

It is important to remember that the quadratic approximation assumes modes are localised about co-rotation. The global nature of some disturbances may invalidate this approximation. Increasing m eventually quenches modes in the non self-gravitating case, thus localised modes with $m \geq 9$ were not found. For the full governing Eq.3.7, high m modes are expected to have increasing wave-like behaviour and be less focused around corotation, contradicting the quadratic approximation and so it is not surprising that they are not found here.

A localised $m = 1$ vortex mode was not found in the self-gravitating case because it had been stabilised by the inclusion of self-gravity, according to earlier analysis. However, there exists other types of low m mode not captured by the quadratic approximation. Modes with extreme values of m are of less relevance since they do not develop in hydrodynamic simulations, which typically show the number of vortices in agreement with the most unstable m .

3. VORTEX MODES IN SELF-GRAVITATING DISCS

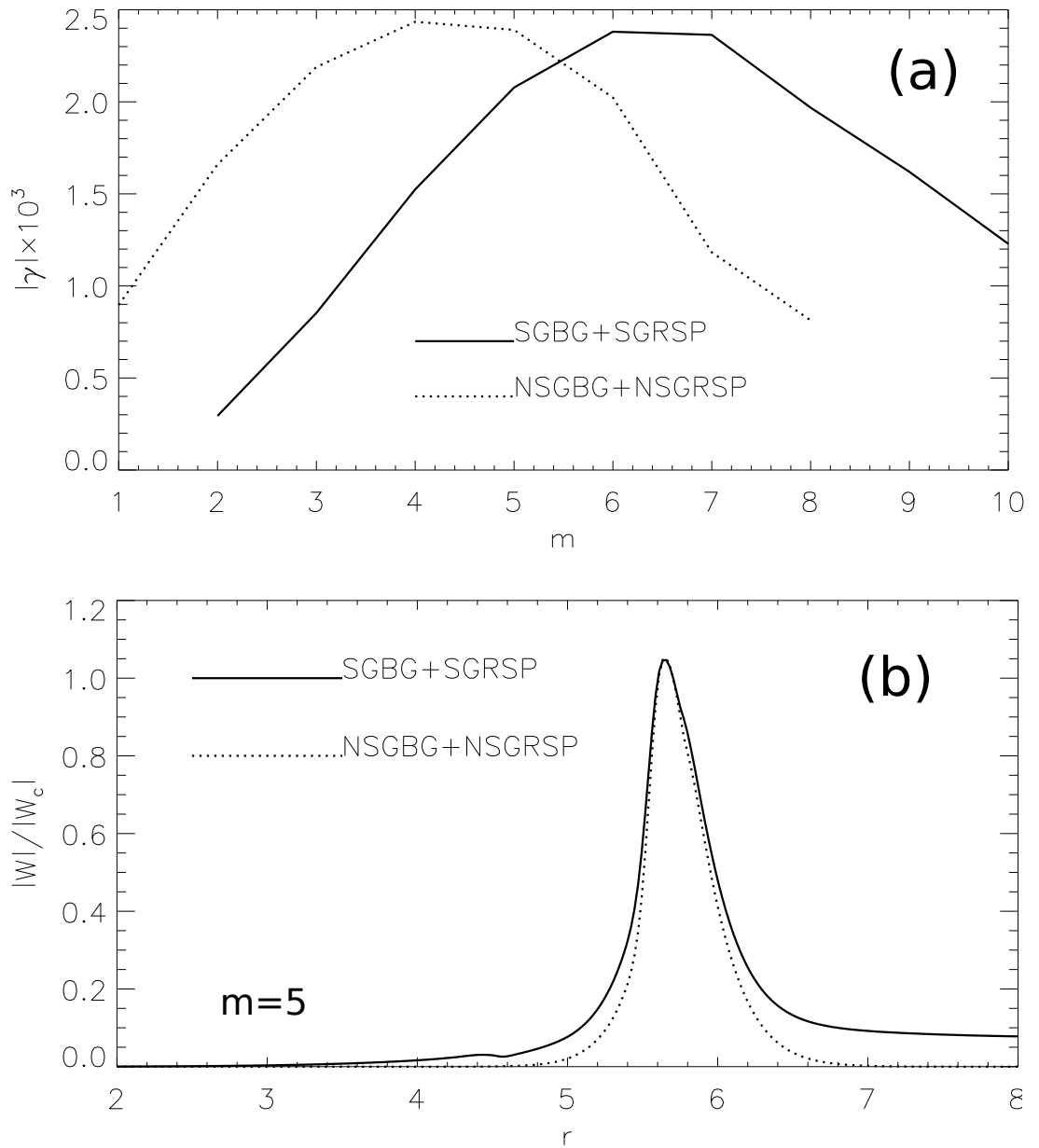


Figure 3.2: Growth rates (a) obtained from the quadratic approximation to the governing equation and the modulus of the $m = 5$ eigenfunction (b). Here $W_c = W(r = 5.7)$. The solid (dotted) lines correspond to the case with (without) self-gravity in both the background state and response. The disc model is $Q_o = 4$.

3.3.6 Solutions to the full governing linear equation

Solutions to the full governing Eq. 3.7, for the fiducial cases with $Q_o = 4$ above are shown in Fig. 3.3. Growth rates and the $m = 5$ eigenfunction are compared. The co-rotation radii for the self-gravitating and non-self gravitating cases are $r_c = 5.88, 5.79$, respectively. These radii are close to the local vortensity minimum. In the non-self-gravitating case, co-rotation almost coincides with the minimum.

The non-self-gravitating case has maximum growth rate for $m = 4$ and the self-gravitating case is most unstable for $m = 6$ – 7 . Modes with $m < 5$ are stabilised while $m > 5$ are destabilised by self-gravity. This results in a shift to higher m modes when self-gravity is included in nonlinear simulations of the model. Modes with the larger m values are destabilised by fully including self-gravity. This is not in contradiction with earlier analysis which assumed small m and did not take account of variations in the form of the basic state. It is shown in §3.3.7 below that the destabilisation of higher m modes is in fact due to the change in the background state.

Results for $m < 5$ are in essential agreement with those obtained in the quadratic approximation giving confidence in view of the importance of the dominance of the corotation (vortensity extremum) region. As with the quadratic approximation, a $m = 1$ mode with dominant disturbance about the gap edge was not found in the self-gravitating case. Thus low m modes are the ones most effectively stabilised by increasing self-gravity.

In the non self-gravitating case, growth rates for $m \geq 8$ are larger than $m = 7$ and do not follow the trend of decreasing growth rates seen from $m = 5 \rightarrow 7$. Non-self-gravitating $m \geq 8$ modes found here are unlikely to be the same type of vortex modes as $m \leq 7$ because they have significant wave-like regions in W and are not concentrated near co-rotation as for $m \leq 7$. In addition, $m \geq 8$ is also where the quadratic approximation appears to fail. These wave-dominated modes demand radiative boundary conditions rather than the simplistic conditions applied here. The $m \geq 8$ non-self-gravitating modes identified here are thus likely to be artifacts of the boundary condition. Fortunately, these modes are irrelevant because they are not the most unstable, nor are they observed in the corresponding nonlinear simulations. By considering the behaviour for $m \leq 7$, one expects a cut-off for vortex modes around $m = 8$ for this model.

The $m = 5$ eigenfunctions in shown in Fig. 3.3 are similar. Both have dominant disturbance around co-rotation. Behaviour in the region $r \leq 4.6$ is essentially identical (since self-gravity weakens with decreasing radius for this disc model). The largest

3. VORTEX MODES IN SELF-GRAVITATING DISCS

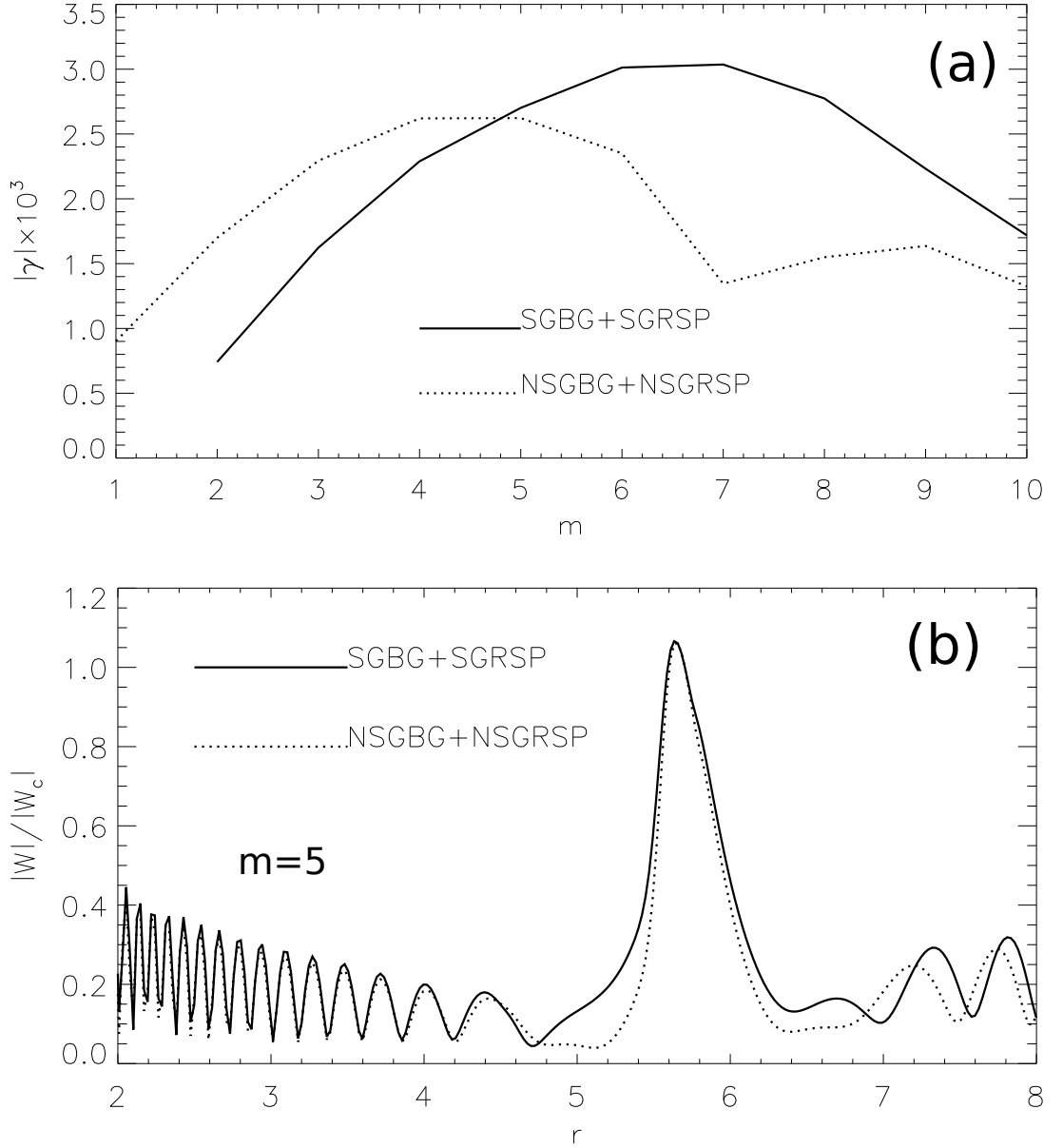


Figure 3.3: Growth rates obtained for the full governing equation (a) and the eigenfunction $|W|$ for $m = 5$ (b). Here $W_c = W(5.7)$. Solid lines are for the fully self-gravitating case and dotted lines are for the fully non self-gravitating case. The disc model is $Q_o = 4$. In (a), the local maximum around $m = 9$ in the dotted curve is most likely caused by a boundary condition effect. Non-self-gravitating modes with $m \geq 7$ are not seen in nonlinear simulations and are not relevant.

difference is found in $r \geq 6.4$, consistent with the general picture that self-gravity is usually more important in outer parts of accretion discs. The disturbance around

co-rotation is also somewhat wider in the self-gravitating case, signifying the global nature of gravity. Comparing with Fig. 3.2, the quadratic approximation captures the main feature of the mode in the co-rotation region. However, it removes the wave-like behaviour in the extended disc.

3.3.7 The role of self-gravity

The calculations above can be compared to hydrodynamic simulations where self-gravity is either enabled or not. The linear problem allows one to examine separately the effect of self-gravity through its influence on the basic state and through its influence on the linear response.

Below, growth rates for the $Q_o = 4$ disc model are compared for a pair of cases, one with and the other without self-gravity in setting up the basic state (i.e. the nonlinear simulations were run with and without disc self-gravity enabled) but both without self-gravity implemented in the linear mode calculation (i.e. setting $\delta\Phi = 0$). Another pair of cases is also compared, one with and the other without self-gravity implemented in the linear response, but with both having the background state calculated with the disc self-gravity incorporated.

Fig. 3.4(a) examines the influence of self-gravity through its modification of the background state. It shows that this modification is destabilising. This is because a deeper gap is set up when self-gravity is included in the simulation. In going from the case without self-gravity to the one with self-gravity, the most unstable mode shifts from $m = 4$ – 5 to $m = 6$. The difference in growth rates decreases as m decreases so the effect due to the modification of the background is smallest for small m . In other words, high m modes are made more unstable by including self-gravity in the basic state.

Fig.3.4(b) compares the growth rates obtained from linear calculations for identical background states but with and without self-gravity implemented in the response. The analysis in §3.2.1 applies to this comparison. Consistent with that, enabling self-gravity in the response decreases $|\gamma|$ and stabilises the system against modes with co-rotation associated with a vortensity minimum for *any value of m* . Unlike the effect acting through the background, the influence of self gravity through the linear response is more significant for low m and can lead to complete stabilisation (e.g. $m = 1$). The most unstable mode without self-gravity has $m = 5$ and including self-gravity shifts it to $m = 6$.

The fact that self-gravity acting in the linear response and background state has

3. VORTEX MODES IN SELF-GRAVITATING DISCS

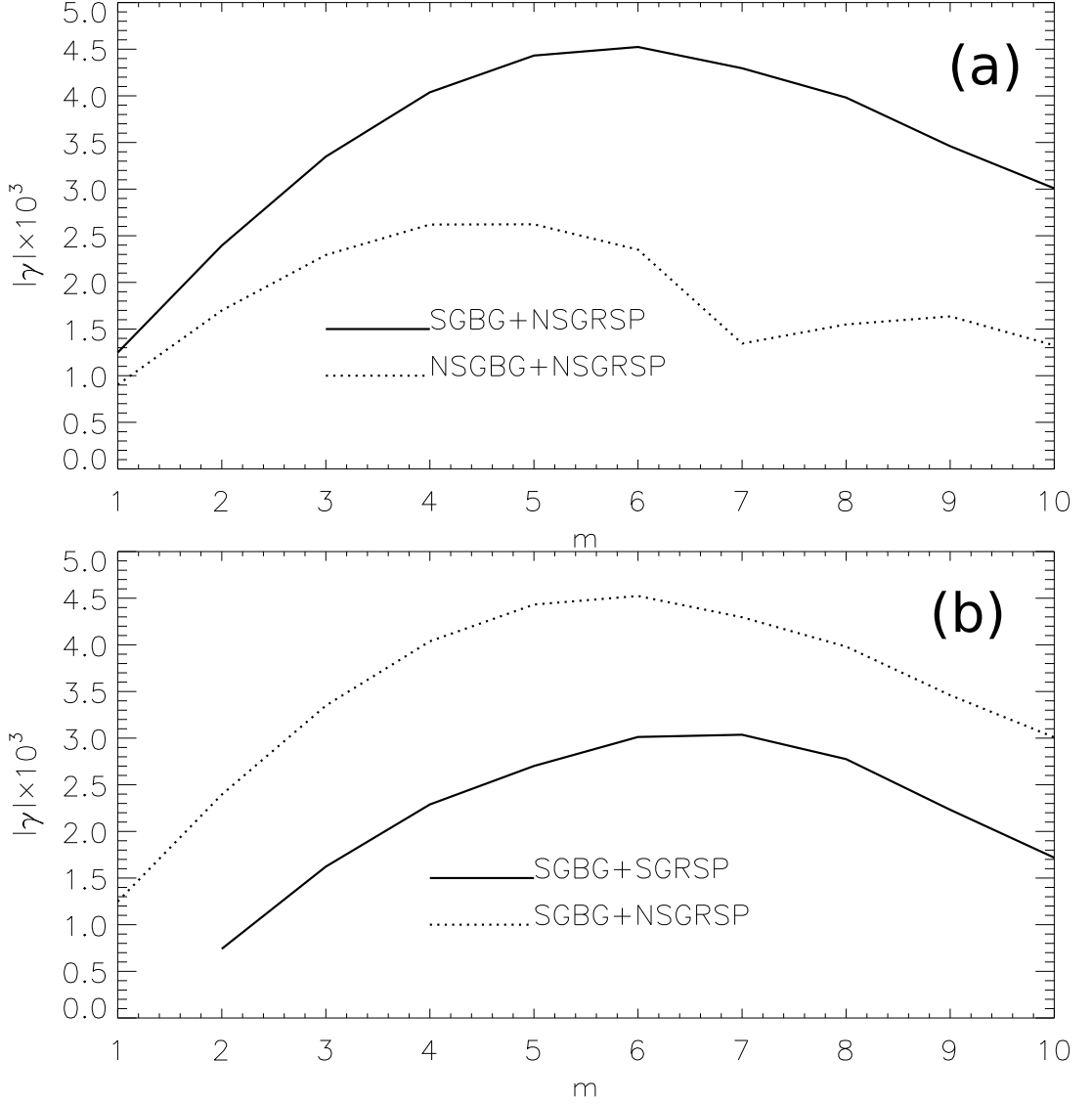


Figure 3.4: *The effect of self-gravity on growth rates of vortex-forming instabilities at gap edges, through the background (a) and through the response (b). The disc model is $Q_o = 4$. In (a), the local maximum around $m = 9$ in the dotted curve is most likely caused by a boundary condition effect. Modes with $m \geq 7$ without self-gravity are not seen in nonlinear simulations and are not relevant.*

opposite effects on growth rates is consistent with the comparison between fully self-gravitating and fully non-self-gravitating cases (see Fig. 3.3, where peak growth rates are approximately equal). The effect of self-gravity through changes to the background and through direct influence on the linear response both contribute to favouring higher

m . However, the background effect is achieved by increasing high m growth rates, whereas the effect via the response works by stabilising low m modes in accordance with the discussion in §3.2.4. Thus low m vortex modes become disfavoured. Overall, one expects more vortices to form, corresponding to increasing m , with increasing self-gravity.

3.3.8 Models with different Q_o

In the calculations below, self-gravity is always included in setting up the basic state. The $Q_o = 4$ disc above is compared to $Q_o = 3$ (a more massive disc) and $Q_o = 8$ (a less massive disc).

Fig. 3.5 show growth rates obtained from the full equation and the quadratic approximation, *without* self-gravity in the linear response. The two sets of curves are similar. As with switching self-gravity on or off, increasing self-gravity in the background destabilises all modes, but this effect diminishes with decreasing m . Although lowering Q_o shifts the most unstable modes to slightly higher m , the low m modes still exist.

Growth rates shown in Fig. 3.6 fully incorporates self-gravity. The plot demonstrates a clear shift of the most rapidly growing mode to higher m as Q_o is lowered. For $Q_o = 8$ the most unstable mode has $m = 5 - 6$ and for $Q_o = 3$ it shifts to $m = 7 - 8$. The shift is accompanied by the stabilisation (or loss) of low m modes. For example, when the disc mass is increased as Q_o changes from $Q_o = 4$ to $Q_o = 3$, the $m = 3$ growth rate decreases by a factor of 3. Modes with the lowest m are stabilised by self-gravity in the response, as $m = 1, 2$ modes were not found for $Q_o = 3$.

Fig. 3.6(b) shows a similar dependence of growth rates on Q_o in the quadratic approximation. The shift to higher m is more apparent. For $Q_o = 3$ no modes with $m \leq 4$ were found. Since the approximation reinforces the localised property of vortex-forming modes, it means that localised modes for low m become increasingly unlikely as Q_o is lowered. Hence, for massive discs only high m vortex forming modes can develop. The agreement between results obtained using the quadratic approximation and the full governing equation indicates a lack of sensitivity to boundary conditions and so is reassuring for the modes of interest.

3. VORTEX MODES IN SELF-GRAVITATING DISCS

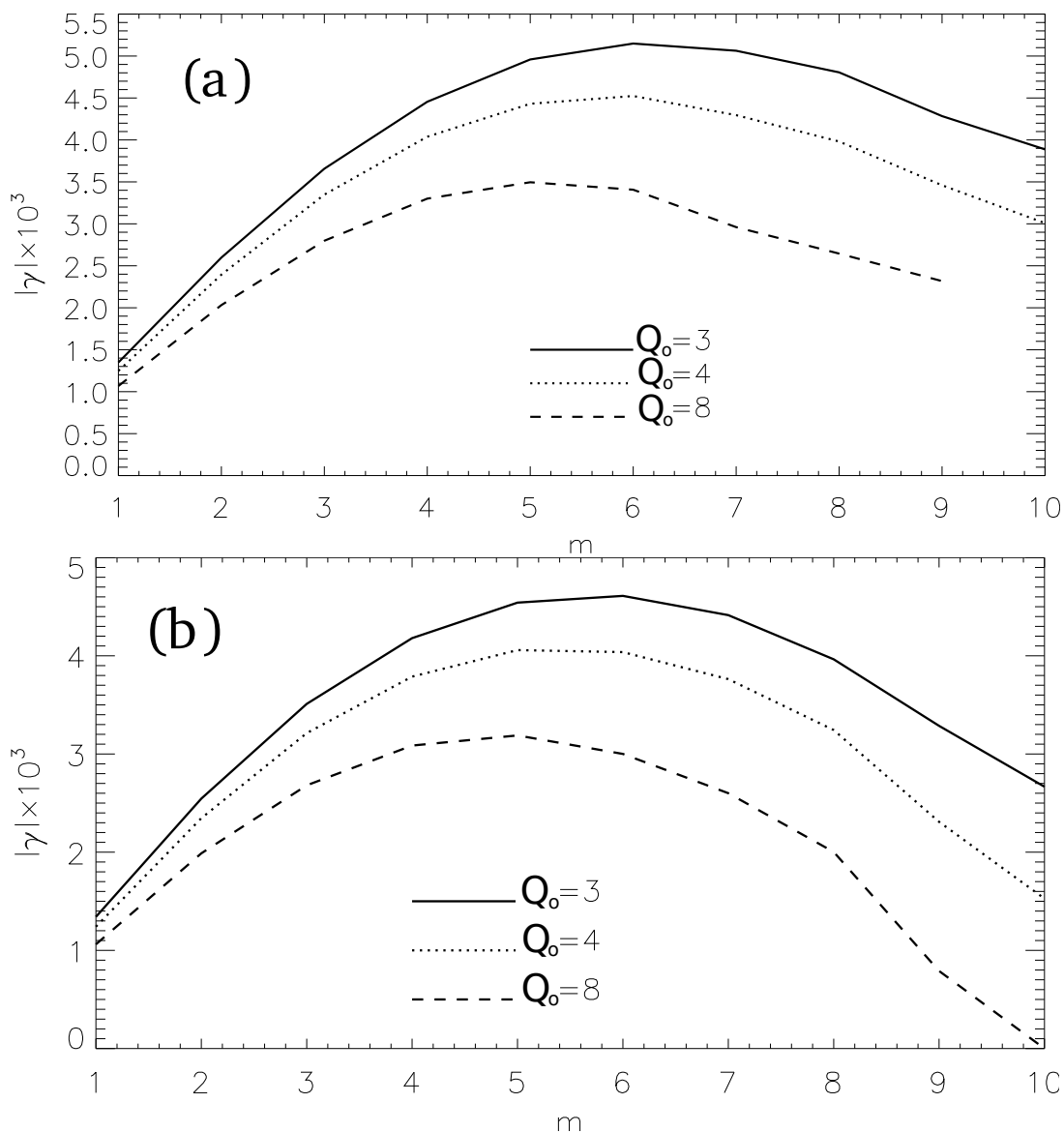


Figure 3.5: Growth rates for unstable modes in background discs with $Q_0 = 3, 4, 8$. (a) is obtained from solving the full linear equation, while (b) is from the quadratic approximation. Self-gravity is only incorporated in the background and not in the linear response. The similarity between (a) and (b) indicate these modes are localised disturbances.

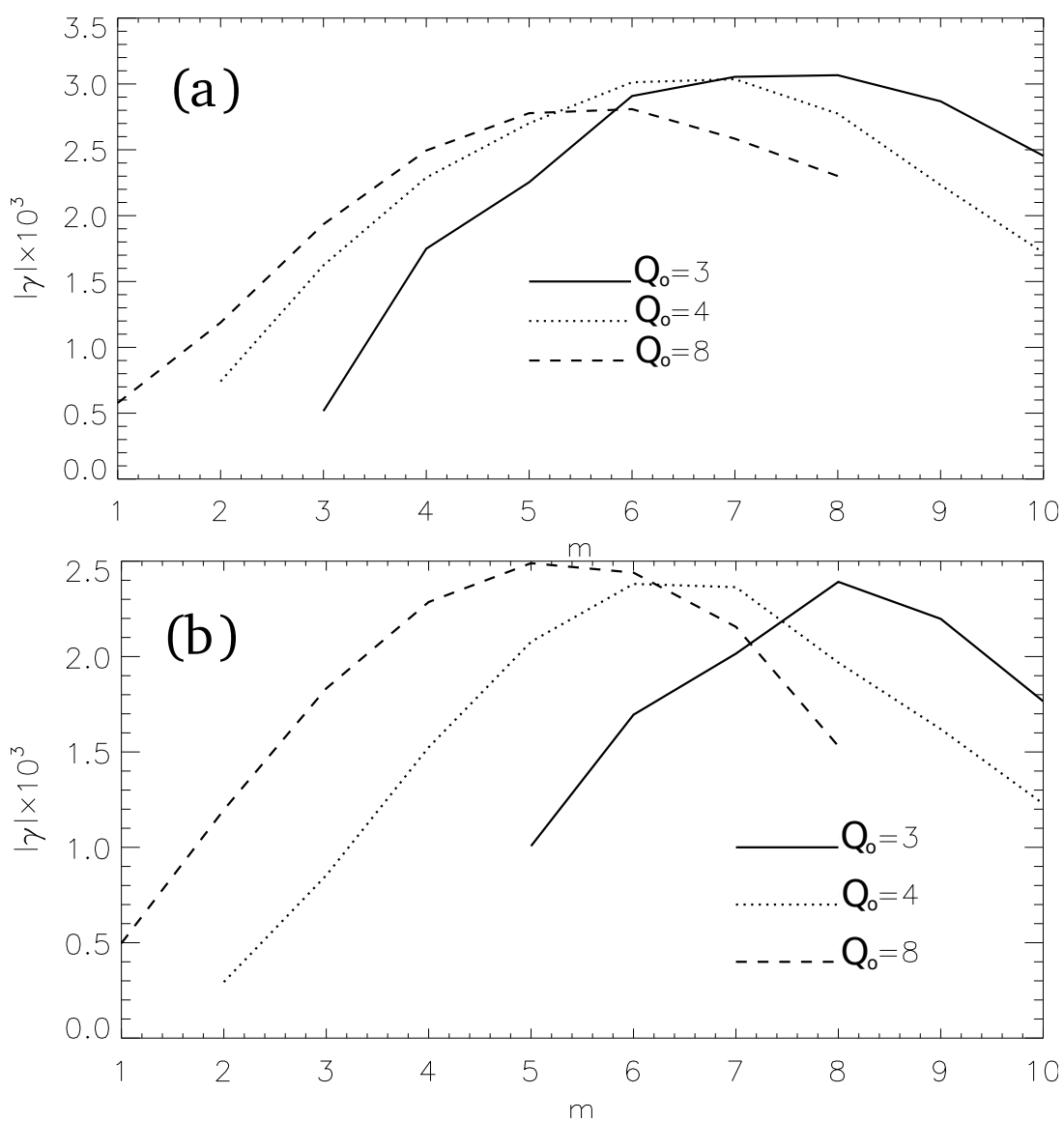


Figure 3.6: Growth rates for unstable modes in background discs with $Q_o = 3, 4, 8$. (a) is obtained from solving the full linear equation, while (b) is from the quadratic approximation. Self-gravity is incorporated in the basic state and the linear response.

3.4 Nonlinear hydrodynamic simulations

Hydrodynamic simulations of vortex formation and evolution at edges of gaps opened by a giant planet were performed for discs of varying masses. The disc-planet models were described in §3.1. The planet mass is $M_p = 3 \times 10^{-4} M_*$ and the disc has uniform kinematic viscosity $\nu = 10^{-6}$ corresponding to an α viscosity of $\alpha = \nu / (c_s H) = 1.8 \times 10^{-4}$ at the planet's fixed orbital radius at $r_p = 5$.

3.4.1 Numerical scheme

The hydrodynamic equations are evolved using an extended version of the **FARGO** code (Masset, 2000a,b) which includes disc self-gravity. A 2D Poisson solver for **FARGO** was implemented by Baruteau & Masset (2008). The gravitational acceleration due to the disc is calculated directly using Fast Fourier Transforms in both azimuth and radius.

The disc is divided into $N_r \times N_\phi = 768 \times 2304$ computational grid cells in radius and azimuth. The grid in the radial (azimuthal) direction is logarithmically (evenly) spaced. An open boundary is imposed at $r = r_i$ and non-reflecting boundary as used by Zhang et al. (2008) at $r = r_o$ (Godon, 1996). Since vortices are localised features, as long as gap edges are far from boundaries, boundary conditions can only have a limited effect. Test simulations with damping boundary conditions (de Val-Borro et al., 2007) and open outer boundaries yield similar results to those presented below.

3.4.2 The effect of self-gravity

The standard distinction between self-gravitating and non-self-gravitating disc-planet simulations is either including or excluding the disc gravitational potential in the total potential calculation (Nelson & Benz, 2003; Zhang et al., 2008). A pair of such cases for the $Q_o = 4$ disc are compared. The disc mass is $M_d = 0.024 M_*$. Note the response of a non-self-gravitating disc to the planet is independent of Q_o .

Fig. 3.7 shows vortensity contours at the onset of vortex formation at the outer gap edge. The local vortensity maximum at about $r = 5.5$ remains largely undisturbed indicating that instability is associated with structure outside it and associated with the vortensity minimum, consistent with earlier analytical arguments. More vortices are excited when self-gravity is included. In that case the $m = 6$ vortex mode dominates whereas the $m = 3$ mode dominates in the non-self-gravitating case. This contrast is consistent with linear calculations.

Fig. 3.7 also shows that vortices have radial length-scales comparable to the local

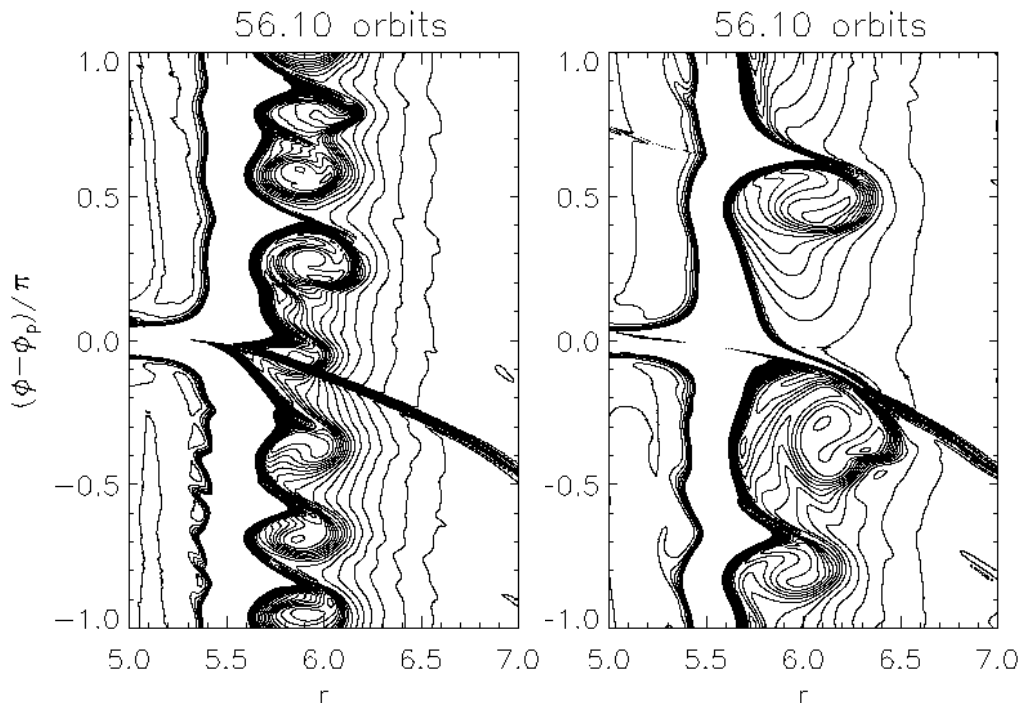


Figure 3.7: Vortensity contours for $Q_o = 4$ with (left) and without (right) self-gravity. The planet is located at $r = 5$, $\phi = \phi_p$. The thick lines crossing the outer radial boundaries correspond to spiral shocks induced by the planet. Local vortensity maxima ($r \sim 5.5$) are also produced by such shocks.

scale-height ($H(6) = 0.3$). The vortices have radial sizes of $\sim 2.2H(6)$ without self-gravity and $\sim 1.3H(6)$ with self-gravity. The vortices are of smaller radial extent when self-gravity is included because of the preference for higher m . A decrease in radial size can be expected from the decrease in width of the WKBJ evanescent zone centred on corotation. This is determined by the condition

$$(\sigma + m\Omega)^2 = \kappa^2(1 - 1/Q^2). \quad (3.39)$$

Writing $\sigma = -m\Omega(r_c)$, it is straight forward to show that the radial width of the evanescent zone approximately scales as $1/m$. Since without self-gravity the preferred m is a factor of two smaller, vortices in this case are expected to be approximately double the size of those in the self-gravitating case.

With self-gravity, the vortices are approximately centred along the radius $r = 5.9$, close to the corotation radius expected from linear calculations $r_c = 5.88$. In the non-self-gravitating case, linear calculation gives $r_c = 5.81$ for $m = 3$, but the vortices are

3. VORTEX MODES IN SELF-GRAVITATING DISCS

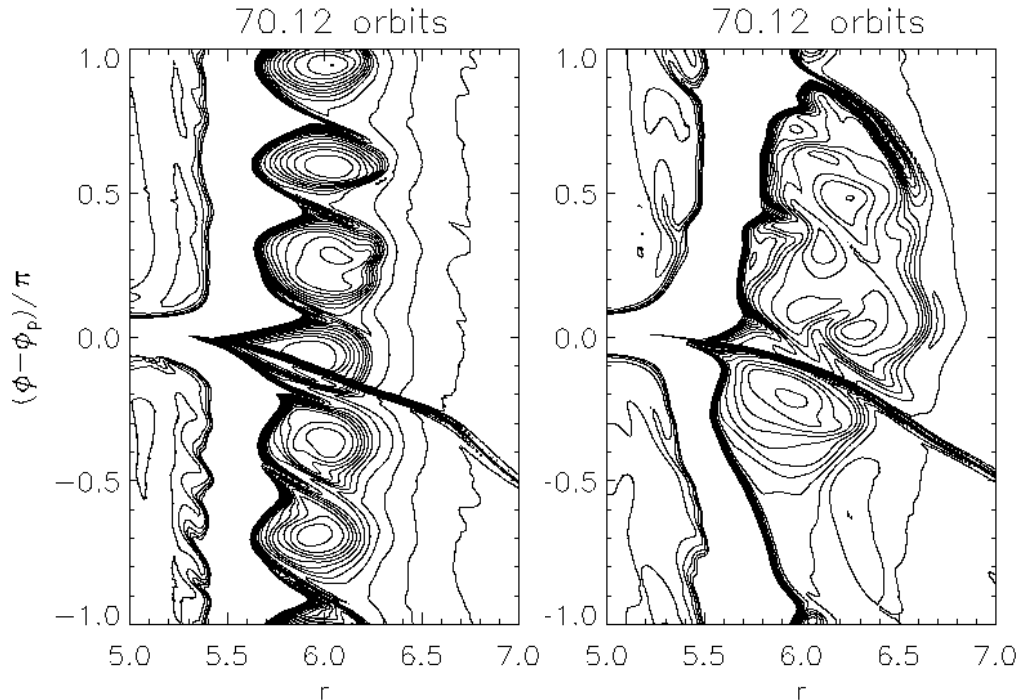


Figure 3.8: *Same as Fig. 3.7 but at a later time. The left panel includes self-gravity whereas the right panel is non-self-gravitating.*

approximately centred along $r = 6$. However, perturbations here are already in the nonlinear regime and interaction between vortices or with the spiral shock may shift the vortices around. In this regard there is more variation in the radial locations of the vortices as compared to the case with self-gravity.

Evolving the systems further, Fig. 3.8 shows that in the non-self-gravitating case, a single vortex has formed through merging. This is a standard result for gap edge vortices (e.g de Val-Borro et al., 2007). However, in the self-gravitating case 6 vortices remain. Resisted merging is explored in more detail later on, but this is not too surprising if one regards these vortices as planets in mean motion resonance.

3.4.3 Varying disc mass: gap profiles

Self-gravitating disc models with $1.5 \leq Q_o \leq 8$ are presented below. These are equivalent to $2.6 \leq Q_p \leq 14$ or $0.063 \geq M_d/M_* \geq 0.012$. The equilibrium gap profiles have a range of Q values with local extrema of $1.5 \leq \min(Q) \leq 9.5$ and $4.8 \leq \max(Q) \leq 21.6$ near the outer gap edge. The Q profiles behave similarly to the vortensity profiles (see Chapter 4 for an explicit illustration).

The gap profiles opened by the planet are shown in Fig. 3.9 for the range of Q_o that develop vortices. Outside the plotted region the curves are indistinguishable. The gap deepens with decreasing Q_o and gap edges become steeper. In going from $Q_o = 8 \rightarrow 2$ the gap depth $|\Delta\Sigma/\Sigma|$ increases by about 0.05—0.08, similarly the bumps near gap edges increase by 0.05—0.06.

Self-gravity affects gap structure on a local scale by increasing the effective planet mass so that $M_p \rightarrow M'_p$. A straightforward estimate, based on the unperturbed disc model, of the expected mass within the Hill radius, r_h , of a point mass planet with $M_p = 3 \times 10^{-4} M_*$ is $M_H = 0.047 M_p$ for $Q_o = 8$ and $M_H = 0.17 M_p$ for $Q_o = 2$. It is likely that M_H , or at least some significant fraction of it adds to the effective mass of the planet acting on the disc when self-gravity is self-consistently included. Thus M'_p , and therefore also the gap depth and steepness of gap edges, are expected to increase with disc mass.

Without carefully tuning M_p , self-gravitating disc-planet calculations with different surface density scales will always differ in M'_p . Since the gap profiles differ, stability is affected also. Note that the gap width w in Fig. 3.9 does not change greatly with Q_o , which is consistent with the scaling $w \propto r_h \propto M_p^{1/3}$. However, the peaks/troughs for $Q_o = 2$ do lie slightly closer to the orbital radius r_p than other cases. This is because shocks are induced closer to the planet due to increased M'_p (Chapter 2).

3.4.4 Varying disc mass: gap stability

Fig. 3.10 shows snapshots of the relative surface density perturbation as instability sets in. Consider $Q_o \geq 2$ first. The instability is associated with the outer gap edge while the inner edge remains relatively stable. In the least massive disc $Q_o = 8$, 3—4 vortices form at the outer gap edge, similar to what happens in the non-self-gravitating disc in §3.4.2. As Q_o is lowered, more vortices develop. By $Q_o = 2$, 8 surface density maxima can be identified ¹.

In moving from $Q_o = 8 \rightarrow Q_o = 2$, vortices become smaller and their centres move inwards. When $Q_o = 8$ local surface density maxima lie just outside $r = 6$ while for $Q_o = 2$ they lie just interior to $r = 6$. The shift to smaller radial location is consistent with analytical expectations that a vortensity maximum is favoured against a vortensity minimum as self-gravity is increased, and for planetary outer gap edges the local vortensity maximum lies closer to the planet than the vortensity minimum (Fig. 3.1). Increasing M_d makes the perturbation more global. Vortices in massive discs

¹Note that a vortex may be obscured by the planetary wake.

3. VORTEX MODES IN SELF-GRAVITATING DISCS

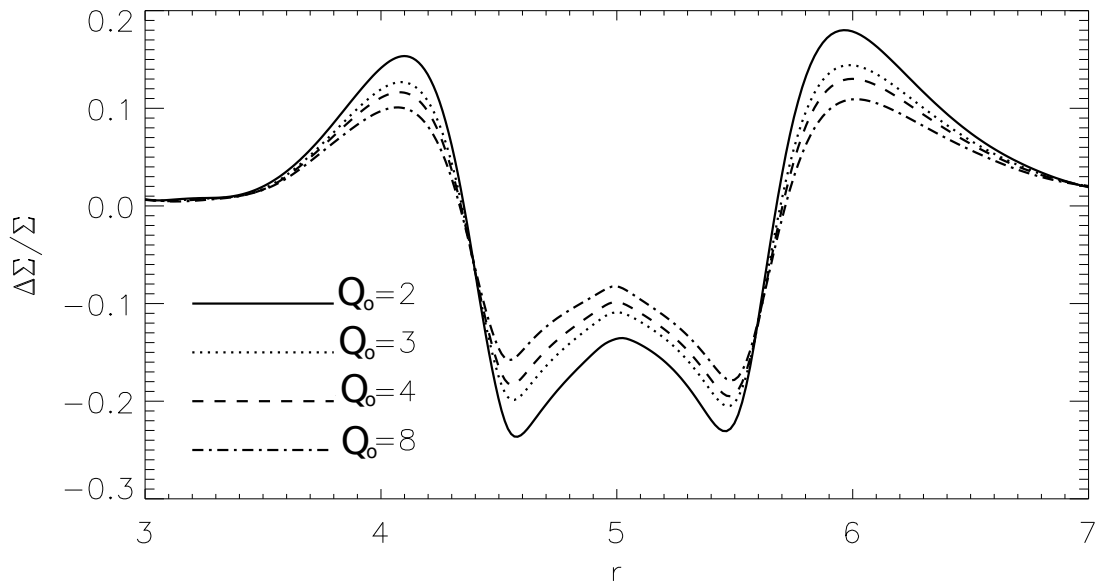


Figure 3.9: Gap profiles opened by a Saturn mass planet in self-gravitating discs as a function of disc mass, parametrised by Q_o . The azimuthally averaged relative surface density perturbation is shown. The planet located at $r = 5$.

can gravitationally perturb parts of the disc further out, similar to a planet. Lowering Q_o , vortices develop longer, more prominent trailing spirals exterior to them. This is most notable with $Q_o = 2$, where the vortex spirals can have comparable amplitudes as the planet wake. However, lowering Q_o even further, a global instability eventually develops as the edge disturbances perturb the disc strongly via self-gravity.

The $Q_o = 1.5$ disc does not develop vortices. This is consistent with the stabilisation effect of self-gravity on the vortex instability through the linear response. Instead, $Q_o = 1.5$ develops global spiral instabilities associated with the gap edge. These are referred to as edge modes and are discussed in detail in Chapter 4. One important difference from vortex modes is that edge modes are associated with local vortensity *maxima*, as expected from energy arguments (§3.2.5).

The Fourier amplitudes of the surface density averaged over the region $r \in [5, 10]$, as a function of m , is shown in Fig. 3.11 for $Q_o = 2, 4, 8$. The snapshot is taken at $t = 56P_0$. The shift to higher m vortex modes with increasing disc mass is evident as expected from linear calculations. For $Q_o = 8, 4, 2$ the preferred vortex modes have respectively $m = 4, 5-7$ and $7-9$, with average amplitudes that decrease with Q_o . The latter is consistent the stabilisation effect of self-gravity on linear modes with low m .

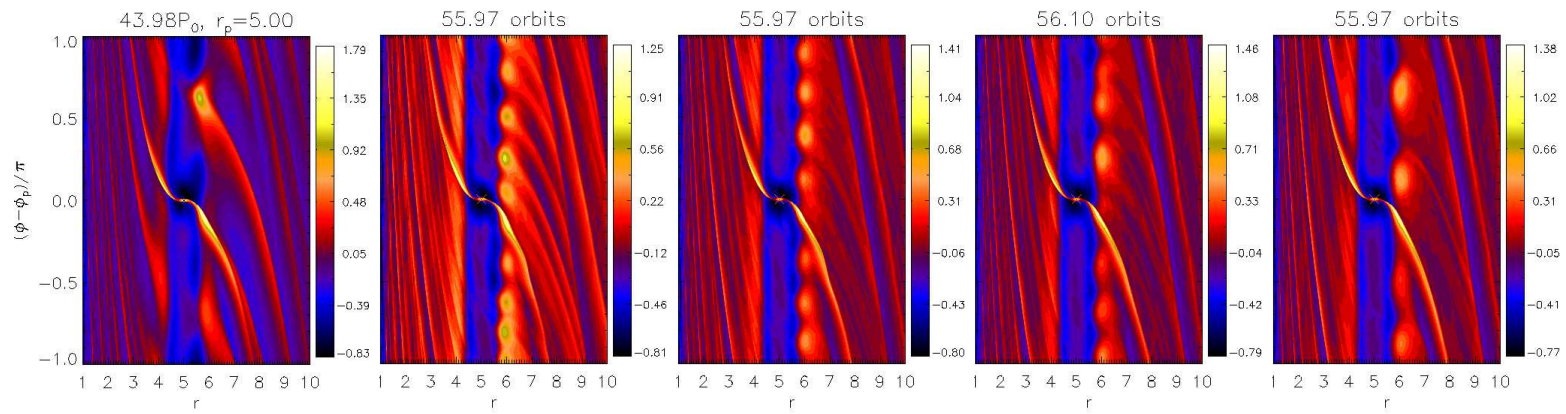


Figure 3.10: *Instability at the outer gap edge of a Saturn-mass planet, in self-gravitating discs with minimum Toomre parameter, from left to right, of $Q_o = 1.5, 2, 3, 4, 8$. The relative surface density perturbation is shown.*

3. VORTEX MODES IN SELF-GRAVITATING DISCS

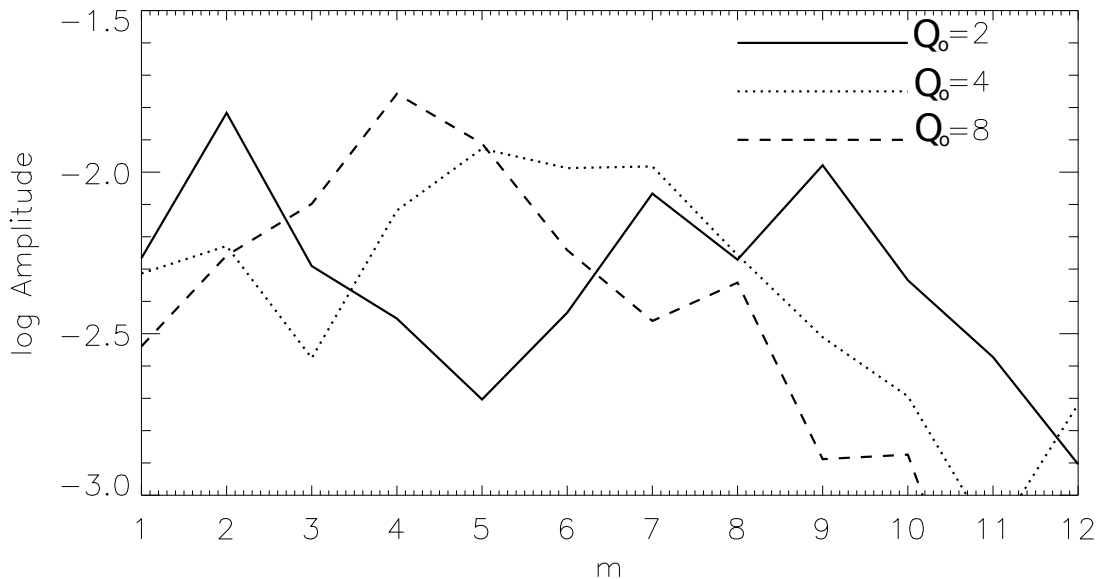


Figure 3.11: *Fourier amplitudes of the surface density in the outer disc $r \in [5, 10]$, normalised by the axisymmetric component, for some of the models illustrated in Fig. 3.10. The plotted quantity is $\log |C_m/C_0|$, where $C_m = \int a_m(r)dr$ and $a_m(r) = \int \Sigma \exp(-im\phi)d\phi$.*

An important region in Fig. 3.11 is $m \leq 3$. There is a peak in amplitude at $m = 2$ for $Q_o = 2, 4$. These are the edge modes described above. They were not seen in linear calculations because those calculations focused on finding vortex modes, which have corotation at or close to local vortensity minima. The loss of low m vortex modes with increasing self-gravity observed in linear calculations, is replaced by the increasing prominence of global edge modes. Fig. 3.11 shows the $m = 2$ amplitude becomes more significant with increasing self-gravity.

The $Q_o = 2$ simulation does display evidence of global disturbances hindering vortex evolution. This transition case is difficult to analyse since both vortices and spirals develop, as indicated by evolutionary snapshots of $Q_o = 2$ in Fig. 3.12. Fig. 3.11 shows the low m and high m modes have comparable amplitudes in $r > 5$. However, if only averaged over the region $r \geq 7$, then in $Q_o = 2$ the $m = 2$ mode was found to be dominant, because this spiral mode is global whereas vortex mode disturbances are localised to the edge ($r \simeq 6$). Increasing self-gravity further, eventually edge modes become dominant, this is seen in the $Q_o = 1.5$ case in Fig. 3.10.

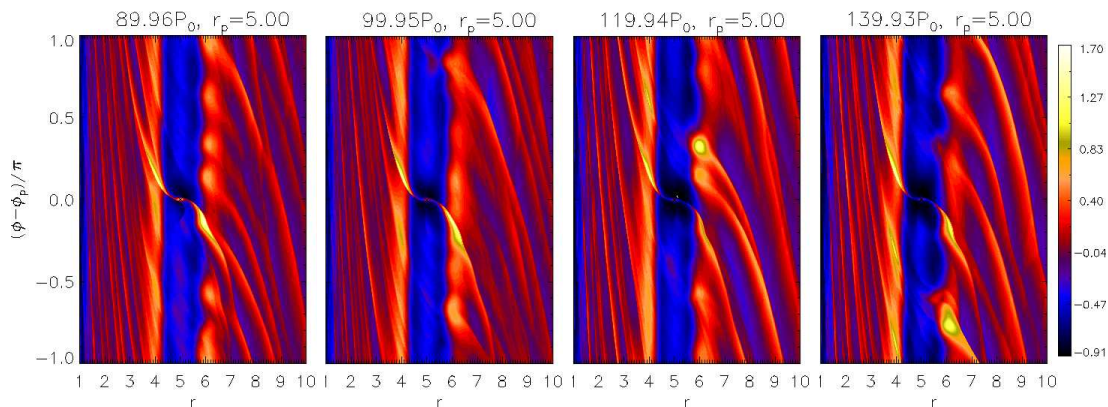


Figure 3.12: *Evolution of gap edge instabilities for the $Q_o = 2$ case, for which both vortex and global spirals develop. Although Fourier amplitudes suggest $m = 2$ global modes to be present early on, 3–4 global spirals can be identified in the last snapshot. These higher m spiral modes may have resulted from nonlinear evolution of vortices perturbing the disc globally, rather than growth from linear instability of edge modes.*

3.4.5 Evolution and merging of self-gravitating vortices

Disc self-gravity also affects nonlinear interaction between vortices. Fig. 3.13 compares surface density perturbations for a range of Q_o at $t = 100P_0$. Since the vortices emerge at roughly $t = 56P_0$, they have evolved for a similar time. Typical growth times for vortex modes are $t_g \sim 6P_0$, so the vortices have evolved for about $7t_g$, well into the nonlinear regime.

The snapshot for $Q_o = 8$ shows a single vortex resulting from merging of the initial vortices. The vortex disturbance is largely confined to within a local scale-height of the gap edge. A single large vortex is the typical result for simulations with no self-gravity (e.g. de Val-Borro et al., 2007, , and in Chapter 2) , although the self-gravitating vortex here is thinner than the completely non-self-gravitating simulation (Fig. 3.8). The $Q_o = 4$ disc shows vortex merging taking place, as individual surface density maxima can still be identified within the large vortex behind the shock. For $Q_o = 4$, a single vortex forms at $t \simeq 110P_0$.

Fig. 3.13 shows that increasing the disc mass delays vortex merging. For $Q_o = 3.5$, 5 vortices remain and for $Q_o = 3$ and $Q_o = 2.5$, 7 vortices remain. They have not merged into a single vortex as happened when $Q_o = 8$. A 25% increase in disc mass as $Q_o = 4 \rightarrow 3$ causes merging to be delayed by $50P_0$. This suggests that increased gravitational interaction between vortices opposes merging.

As Q_o is lowered the inter-vortex distance increases. When $Q_o = 2.5$ their azimuthal

3. VORTEX MODES IN SELF-GRAVITATING DISCS

separation can be larger than the vortex itself. Also, vortices become less elongated and more localised being symptomatic of gravitational condensation. Trailing wakes from vortices become more prominent as the vortices more strongly perturb the disc via their self-gravity. In fact, the planetary wake becomes less identifiable among the vortex wakes. Notice vortex wakes are mostly identified with vortices upstream of the planetary wake, rather than just downstream. Vortex wakes appear to detach from the vortex after passing through the planetary shock. There is a sharp contrast in gap structure between the $Q_o = 2.5$ and $Q_o = 8$ cases. The single vortex that results when $Q_o = 8$ is aligned along the outer gap edge, which is still approximately identified as circle $r = 6$. However, when $Q_o = 2.5$ the vortices and their wakes intersect the circle $r = 6$ making the radius of the outer gap edge less well-defined.

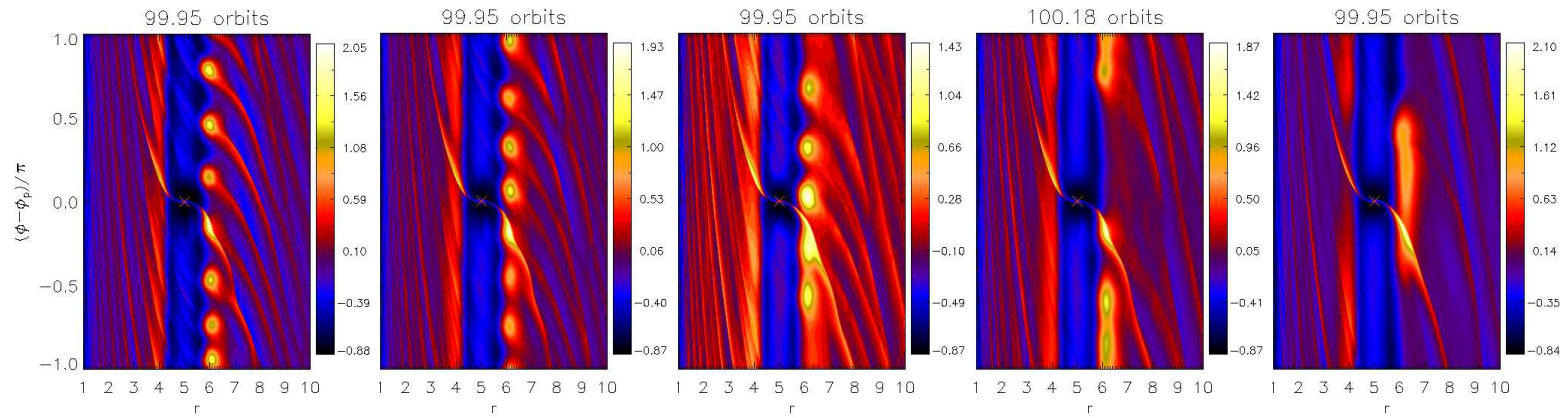


Figure 3.13: *Non-linear evolution of vortex instabilities at the outer gap edge of a Saturn mass planet as a function of disc mass, parametrised by the minimum Toomre parameter Q_o , from left to right, of $Q_o = 2.5, 3.0, 3.5, 4.0, 8.0$.*

3. VORTEX MODES IN SELF-GRAVITATING DISCS

An important feature of vortices is their wakes are associated with the excitation of density waves which transport angular momentum transport outwards (see e.g. Paardekooper et al., 2010a), even without self-gravity. A measure of this can be made through the α viscosity parameter. Here, it is defined as

$$\alpha(r, \phi) \equiv \Delta u_r \Delta u_\phi / c_s^2(r) \quad (3.40)$$

where Δ denote deviation from azimuthal averaged values. For diagnostics, α is spatially averaged over the vortex region $r \in [5.7, 7.1]$ and its running-time average $\langle \alpha \rangle$ is plotted in Fig. 3.14. The parameter $\langle \alpha \rangle$ associated with vortices is $O(10^{-3})$, an order of magnitude larger than the imposed value associated with ν . When $Q_o = 4$, $\langle \alpha \rangle$ decays steadily after an initial transient growth. The case $Q_o = 2$ has a roughly constant $\langle \alpha \rangle \simeq 1.9 \times 10^{-3}$ for $t \gtrsim 80P_0$ after a brief growth around $t \sim 70P_0$. Interestingly, for intermediate strengths of self-gravity, $Q_o = 2.5\text{--}3.5$, there is growth in $\langle \alpha \rangle$ over several tens of orbits.

The behaviour of $\langle \alpha \rangle$ as a function of Q_o is consistent with the general picture of vortex formation and evolution (Fig. 3.13). The decay in $\langle \alpha \rangle$ is associated with vortex merging leading to fewer vortices. This is the case for $Q_o = 4$. Vortex merging happens more readily for lower disc masses, hence although multiple vortices develop from the instability, this phase does not last long enough for the multi-vortex configuration to significantly transport angular momentum. Increasing self-gravity to $Q_o = 3.5 \rightarrow 2.5$, vortices become less prone to merging and the multi-vortex phase lasts longer. They have time to evolve into compact objects that further perturb the disc. This is consistent with fact that $\langle \alpha \rangle$ -growth is prolonged with increased disc mass, as merging is delayed. However, if self-gravity is too strong, such as when $Q_o = 2$, growth is again limited, because the $m = 2$ global spiral mode develops and hinders vortex evolution.

3.4.6 Long term evolution of gap edge vortices

The long term evolution of vortices in the $Q_o = 3$ disc is shown in Fig. 3.15. It shows the instantaneous α viscosity parameter, average over-density (defined in regions where the relative surface density perturbation is positive) and contour plots of relative surface density perturbation in the vortex region.

The viscosity parameter α grows from $t = 100P_0$ to $t = 170P_0$ with $\max(\alpha) \simeq 8 \times 10^{-3}$, which is ~ 50 times larger than the background viscosity. At $t = 150P_0$ there remains 6 distinct vortices (one vortex just passing through the planetary wake).

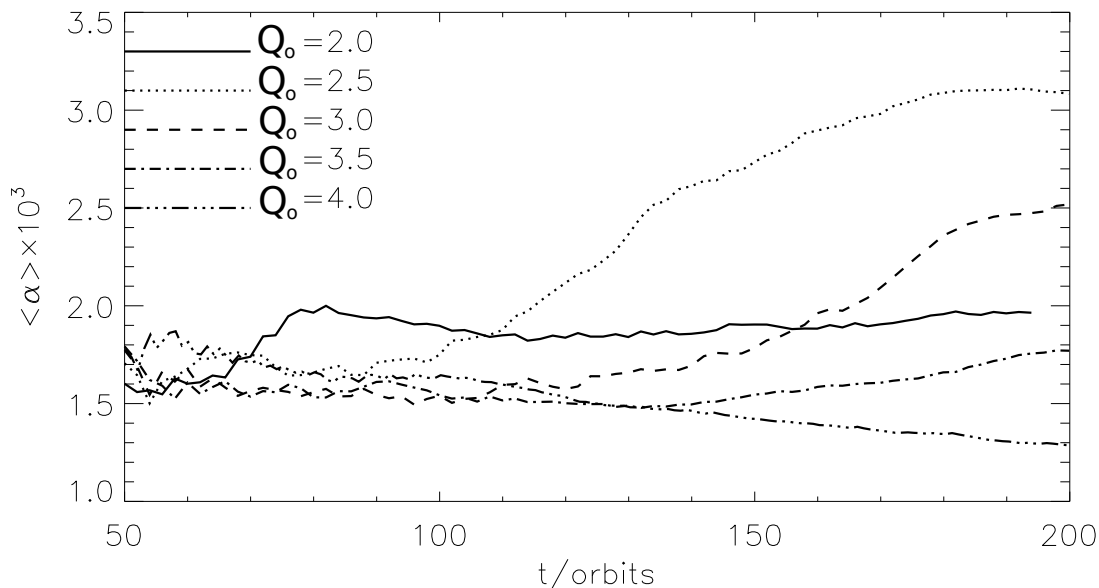


Figure 3.14: Running time average of the viscosity parameter α averaged over the vortex region for a range of disc models.

The multi-vortex configuration has been maintained for a further $\sim 50P_0$ since Fig. 3.13. With weak self-gravity, a single vortex would have formed through merging. Delayed merging allows individual vortices to evolve and become planet-like. The typical over-density in a vortex is ~ 1 at $t = 100P_0$ and increases to > 1.5 by $150P_0$. As a consequence one expects increased disturbance in the surrounding disc. The phase of α -growth correlates with a linear increase in the average over-density in the vortex region.

At $t = 170P_0$, Fig. 3.15 shows 6 vortices still remain, and the parameter α is a maximum. However, after a burst of vortex merging events, α decreases rapidly. The snapshot at $t = 180P_0$ shows a quieter disc with only 3 vortices, but of similar size to those before merging. This is unlike cases with weak self-gravity where a larger vortex results from merging. Fig. 3.16 compares post-merging vortices in discs with $Q_o = 3$ and $Q_o = 4$. For $Q_o = 3$, the post-merging vortices are localised in azimuth, with $Q \sim 1$ and the contour plot shows the densest vortex has an over-density of $\simeq 3.75$. However, for the disc with $Q_o = 4$, a single vortex forms that extends about half the total azimuth and has $Q > 2$. This hints at gravitational collapse of vortices in the $Q_o = 3$ disc.

Note that the mass of the $Q_o = 3$ disc is only $M_d = 0.031M_*$, usually considered insufficient for fragmentation into bound objects by classic Toomre instability (though it

3. VORTEX MODES IN SELF-GRAVITATING DISCS

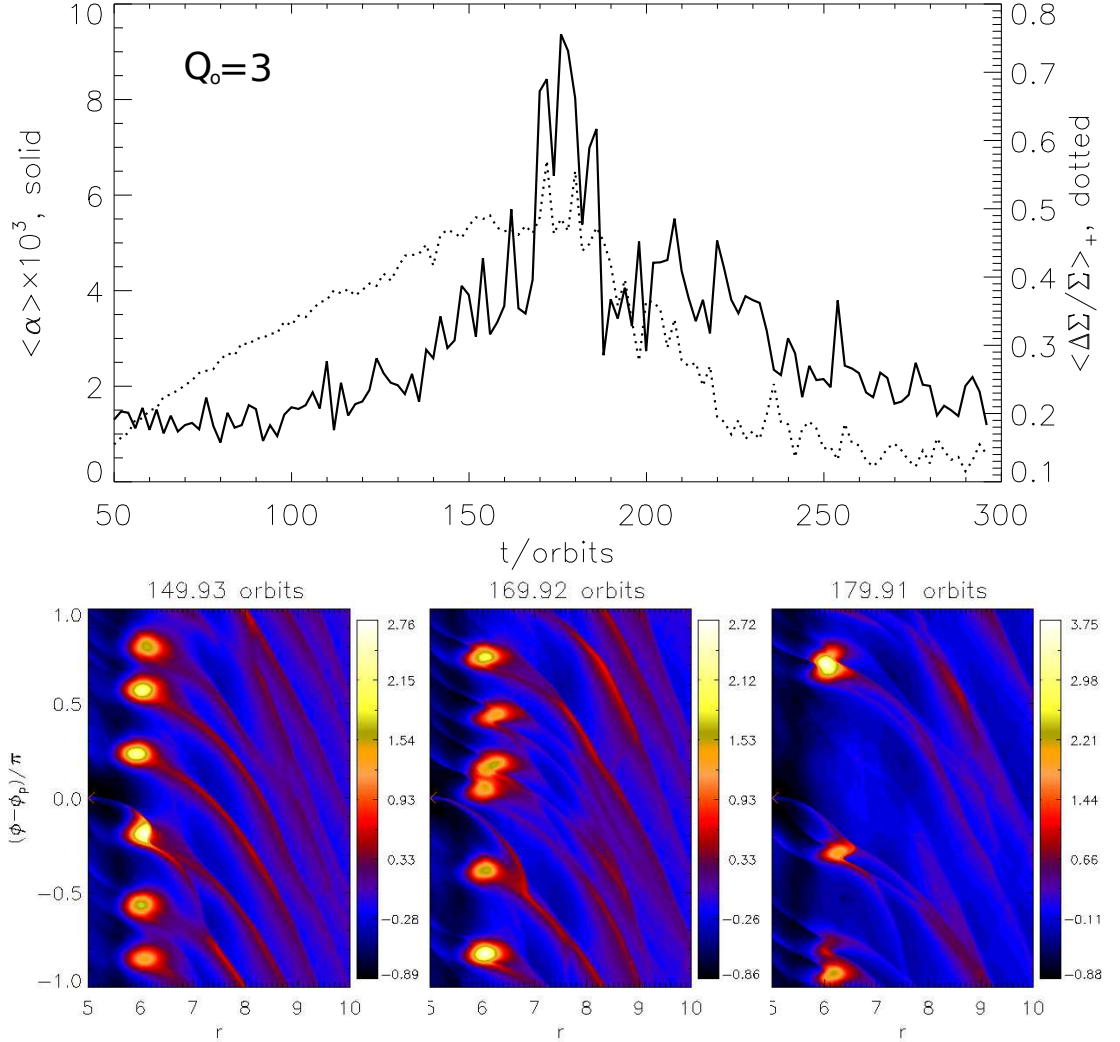


Figure 3.15: Vortex evolution for the disc with $Q_o = 3$. Line plot: instantaneous α viscosities (solid) and average over-density (dotted). Contour plots: relative surface density perturbations before and after vortex merging.

is still massive compared to more typical protoplanetary disc masses of $M_d \sim 0.01M_*$). However, the comparison above shows that adding self-gravity to the vortex instability enables collapse into compact objects that survive against shear.

Vortices in $Q_o = 3$ noticeably affect the gap structure. Fig. 3.17 shows several snapshots of the gap structure from $t = 100P_0$, when there were multiple vortices, to the end of the simulation when a vortex pair remained. The quasi-steady gap profile

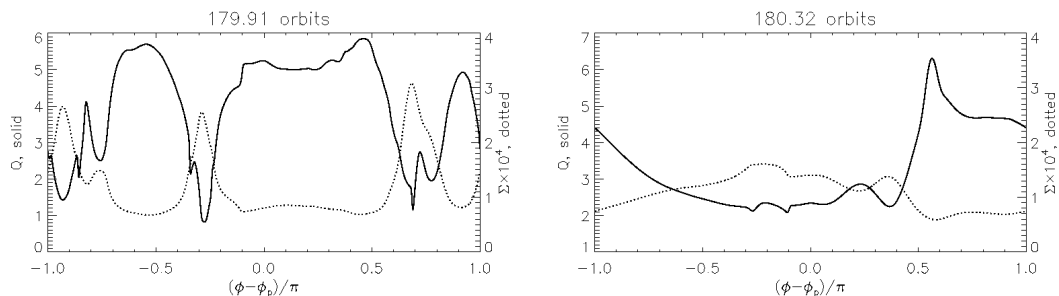


Figure 3.16: Comparison of post-merging vortices for discs with $Q_o = 3$ (left) and $Q_o = 4$ (right). The Toomre Q (solid) and surface density (dotted) averaged over $r \in [6, 6.5]$ is shown as a function of azimuth relative to the planet.

for $Q_o = 4$ is also shown for comparison. Neither the single vortex in $Q_o = 4$ nor multiple vortices with $Q_o = 3$ affect the one-dimensional gap profile at $t = 100P_0$ because the disturbances only redistribute mass in the azimuthal direction. However, in the $Q_o = 3$ disc, the original bump at the outer gap edge is diminished after vortex merging takes place at $t = 200P_0$. A surface density depression of $\Delta\Sigma/\Sigma \simeq -0.1$ then develops at $r = 7.3$ and a bump develops at $r = 8.5$. These features last to the end of the simulation.

Self-gravitating vortices behave like planets. Assuming vortices lie near the surface density maximum at $r = 6$, the creation of the surface density deficit in $r \in [7, 7.5]$ and the new surface density maximum at $r = 8.5$ could be induced, in a similar manner as it would be by a planet, through the outward transport of angular momentum by the density waves launched by the vortices.

Roughly speaking, treating m vortices at $r = r_v$ as a gravitational perturbation on the exterior disc, it exerts a positive torque at $r_L = r_v(1 + 1/m)^{2/3}$. Inserting $r_v \sim 6$ and $m \sim 3$ (Fig. 3.15—3.16) gives $r_L \sim 7.3$. The material removed from the region of the deficit, which has gained angular momentum, ends up contributing to the new surface density maximum at $r = 8.5$.

3.4.7 Anticyclonic vortices

For the $Q_o = 3$ disc, a vortex-pair forms at $t \simeq 230P_0$ and lasts until the end of the simulation ($t \simeq 300P_0$) and is reminiscent of co-orbital planets. The pair survives on a timescale well beyond which merging occurs in non-self-gravitating simulations. A snapshot is shown in Fig. 3.18(a). Two blobs can be identified along the gap edge, the upper vortex being more over-dense than the lower one. They lie at a radius

3. VORTEX MODES IN SELF-GRAVITATING DISCS

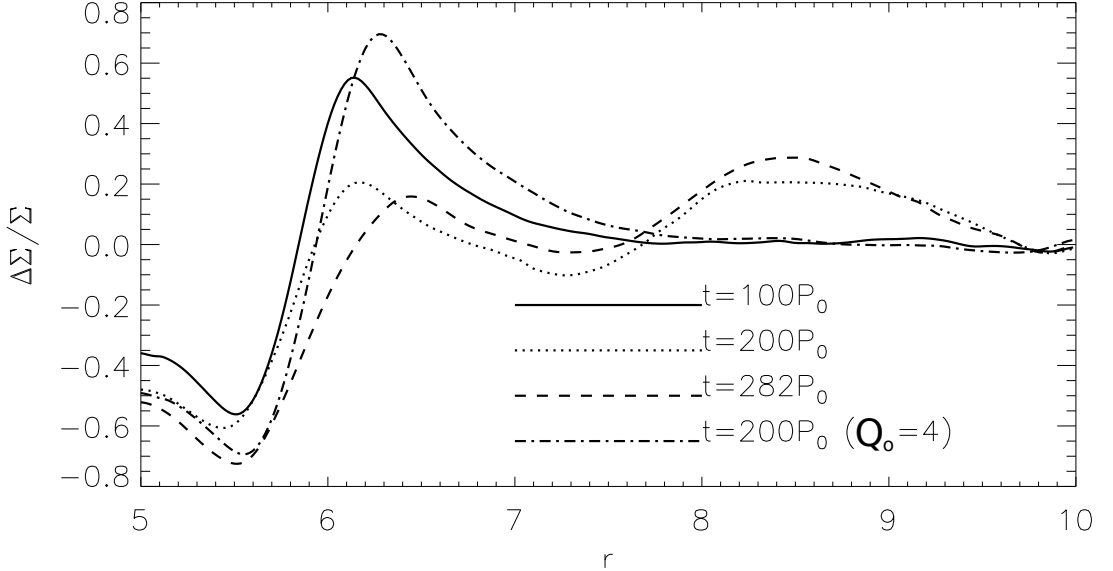


Figure 3.17: The gap profile for the $Q_o = 3$ disc at different times (solid, dotted and dashed) as shown through relative surface density perturbation. A snapshot of the profile for the disc with $Q_o = 4$ (dash-dot) is also plotted for comparison.

$r_v = 6.4$, corresponding to local surface density maximum in azimuthally averaged one-dimensional profiles, which is expected to be neutral for vortex migration (Paardekooper et al., 2010a).

The upper vortex in Fig. 3.18(a) is different to the pre-merging vortices or those with weak self-gravity. It has two spiral disturbances extending from the vortex to $(r, \phi - \phi_p) = (9, -0.2\pi)$, whereas the pre-merging vortices have one trailing spiral. Its vortensity field is shown in Fig. 3.18(b). The vortex core has $\eta < 0$, whereas the final large vortex in the disc with $Q_o = 4$ has $\eta > 0$ in its core (though it is still a local minimum).

The region with $\eta < 0$ has a mass of $M_v \simeq 5.46 \times 10^{-5} M_* \simeq 0.18 M_p$ and average radius $\bar{r} \simeq 0.92H(r_v)$ about the vortex centroid. It has semi-major and semi-minor axes of $\sim 1.57H(r_v)$ and $\simeq 0.55H(r_v)$ respectively, corresponding to aspect-ratio of 2.9. This is about 18 Earth masses if M_* is Solar. Denoting the mean square relative velocity (to the vortex centre) as Δv^2 , $GM_v/\bar{r}\Delta v^2 \sim 3.9$. This region is gravitationally bound. Although planets of such mass are not expected to open significant gaps, the surface density deficit in $r = [7, 7.5]$ in Fig. 3.18(a) indicates vortices may do so. The process could be assisted by the fact that a vortex of size H produces a perturbation of a magnitude similar what would be produced by Saturn when $h = 0.05$ (Paardekooper et al., 2010a).

3.5 Simulations of co-orbital Kida vortices

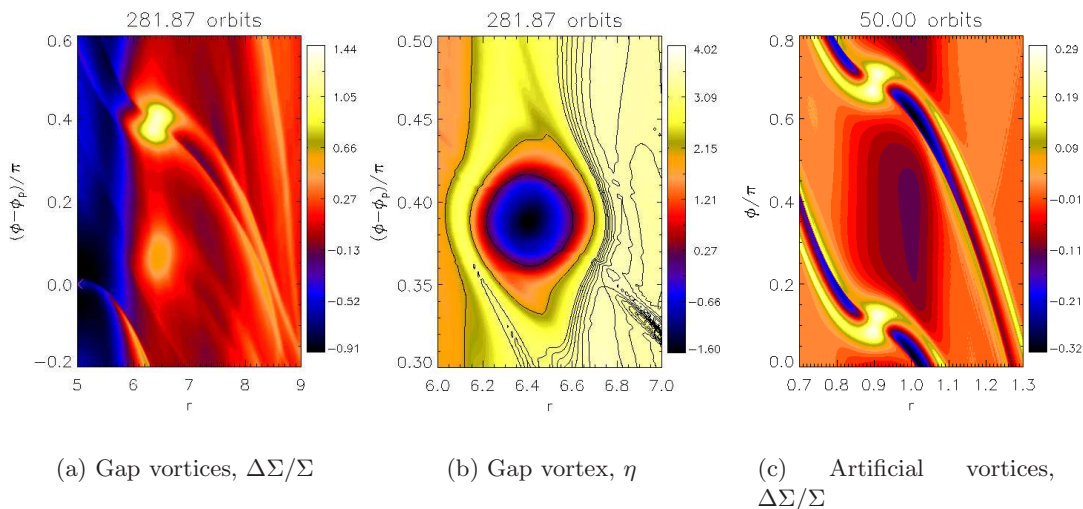


Figure 3.18: *The final vortex pair at the end of the disc-planet simulation for $Q_o = 3$: (a) Relative surface density perturbation contour plot (b) Vortensity contours for the upper vortex. Shown in (c) is the relative surface density perturbation for a pair of vortices formed by imposing Kida vortex solutions as a perturbation in a standard power-law disc.*

Finally, Fig. 3.18(c) show $\Delta\Sigma/\Sigma$ for a pair of Kida-like vortices placed in a disc. Note the similarity between the Kida-like vortex and the upper vortex in Fig. 3.18(a), particularly the double wake structure, the tilted core and the surface density deficit just outside the vortices. This is a surprising coincidence, given that the simulation setups that produced them were completely different (for more details see below).

3.5 Simulations of co-orbital Kida vortices

One of the effects of self-gravity shown in simulations above is to delay vortex merging. This effect is significant when vortices develop into compact structures. Furthermore, the similarity between the final vortex in the disc with $Q_o = 3$ and a Kida-like vortex (Kida, 1981) motivates one to consider the effect of self-gravity on the interaction between two Kida vortices.

Simulations presented here are supplementary and focus on the interaction between vortices without a planet. This isolates effects due to vortex-vortex interactions and the influence of self-gravity without interference from the planet.

3. VORTEX MODES IN SELF-GRAVITATING DISCS

Table 3.1: *Parameters for two-vortex interaction simulations. The first column gives the nomenclature for the simulations. The initial radial separation is X_0 and Q_1 is the Keplerian Toomre parameter Q_{Kep} at unit radius. The fourth column indicates whether self-gravity was included.*

Case	X_0/H	Q_1	self-gravity
Fnsg	0	8.0	NO
Fsg	0	8.0	YES
M1	0	15.9	YES
M2	0	5.3	YES
S1	0.1	8.0	YES
S2	0.2	8.0	YES
S3	0.3	8.0	YES
S4	0.6	8.0	YES

3.5.1 Numerical setup

A standard power law disc is adopted, with initial surface density profile $\Sigma = \Sigma_0 r^{-1/2}$, where Σ_0 is a scaling constant. The disc occupies $r \in [0.25, 2.5]$. The value of Σ_0 is chosen to provide a specified Q_{Kep} at a specified radius as before. The initial azimuthal velocity is chosen so that the disc is initially in hydrostatic equilibrium. The initial radial velocity is zero. The disc is locally isothermal with $h = 0.05$. The viscosity is $\nu = 10^{-9}$ which is essentially the inviscid limit. No planets are introduced and the notation (r_p, ϕ_p) is used to denote a vortex centroid.

Artificial vortices are set up by following Lesur & Papaloizou (2009). Details of the implementation is described in Appendix E. Two vortex perturbations are imposed with initial angular separation $\theta = \pi/2^1$. Table 3.1 summarises the numerical experiments, which considers switching self-gravity on or off, a range of disc masses (here specified through $Q_1 \equiv Q_{\text{Kep}}(1)$) and initial radial separations X_0 of the two vortices (so that for one of the vortices, $r_p \rightarrow r_p + X_0$).

The fluid equations are evolved using the FARGO code with resolution $N_r \times N_\phi = 800 \times 2400$ giving $\simeq 17$ grid points per scale height. The vortices are typically of that size. Damping boundary conditions are applied (de Val-Borro, 2006).

3.5.2 Vortex pairs

As an example of vortices formed by the above procedure, Fig. 3.19 shows the vortensity field for the case Fnsg. The vortex pair circulates at $r \simeq 0.96$ and are localised in radius

¹Not to be confused with the polar angle in 3D.

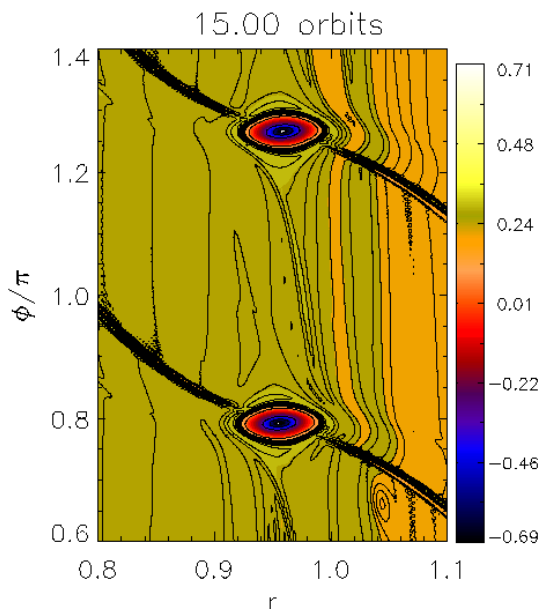


Figure 3.19: Vortices formed by imposing Kida vortex solutions as a perturbation to a global disc. The scaled vortensity field is shown.

and azimuth. The vortex centroids have vorticity -1.57 (dimensionless units) in the non-rotating frame, close to the local Keplerian shear ($-1.5r_p^{-3/2} \simeq -1.59$). The upper vortex has a long tail reaching the outer part of the lower vortex rather than its centroid. The wakes associated with each vortex are similar to those induced by a planet and is responsible for vortex migration (Paardekooper et al., 2010a).

Focusing on one vortex, the region with negative absolute vorticity has half width in radius and azimuth of $\simeq 0.6H$ and $1.3H$ respectively. This region has mass $m_v \sim 1.2 \times 10^{-5} M_*$, or about $4M_\oplus$ if $M_* = 1M_\odot$. The estimate is not far from the assumption that the final vortex has size H with average density that of the unperturbed disc at the location it was set up, which gives $m_v \sim 1.6 \times 10^{-5} M_*$. If self-gravity is enabled, one can expect gravitational interaction between vortices to behave like that between two co-orbital planets of at least a few Earth masses.

3.5.3 Simulation results

Consider first the cases Fsg and Fnsg. These are for the disc with $Q_1 = 8$ with and without self-gravity respectively. The distance between vortex centroids as a function of time is shown in Fig. 3.21(a). It takes almost 5 times as long for the self-gravitating

3. VORTEX MODES IN SELF-GRAVITATING DISCS

vortices to merge. A merging event is shown in Fig. 3.20.

By examining the vortensity distribution, non-self-gravitating vortices were observed to begin merging at $t = 28P_0$, but self-gravitating vortices at $t = 126P_0$. For F_{nsg} , vortices approach each other within $\sim 23P_0$ of formation and merge. Assuming they move relative to each other because of Keplerian shear, the time taken to merge implies that their radial separation at vortex formation must be $\simeq 0.14H$. This length-scale is only resolved by about 2.5 grid cells indicating that a non-zero initial separation is generated by grid effects, despite the vortex perturbations being imposed at the same radius. This is actually not unphysical because vortex formation through instability is not expected to occur at exactly the same radii.

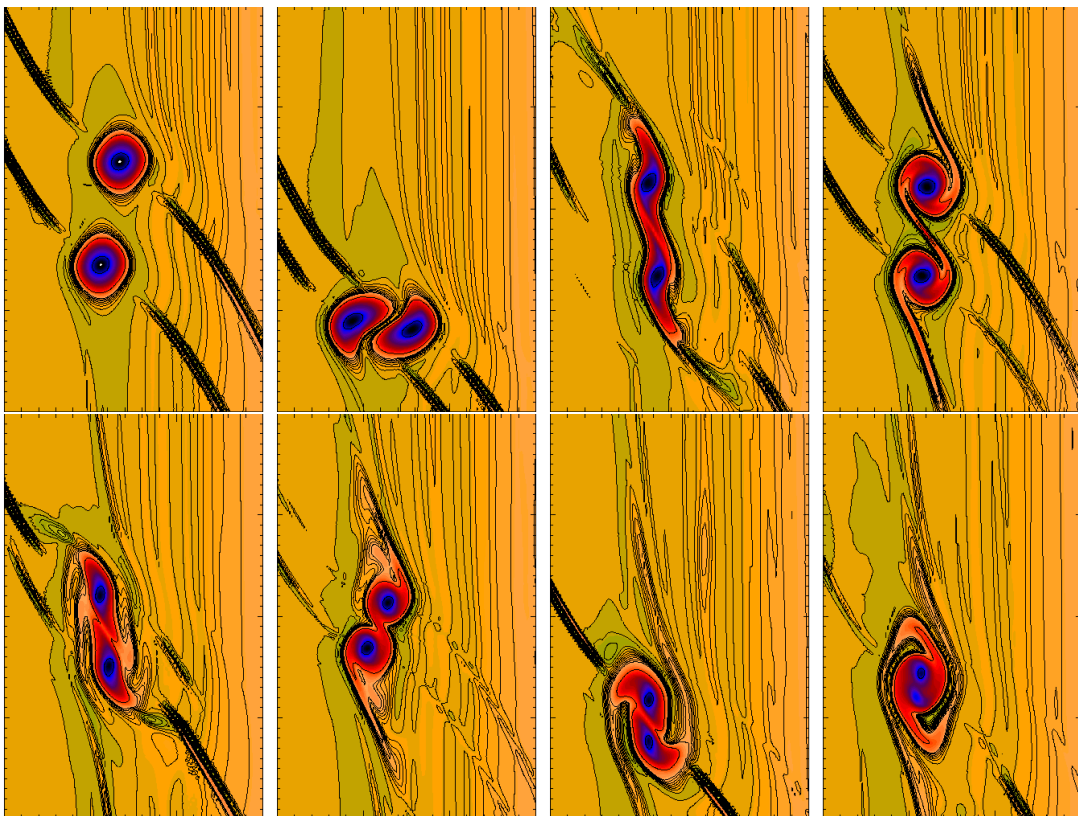


Figure 3.20: *Typical merging of a vortex pair. There is no qualitative difference in this process between self-gravitating or non-self-gravitating vortices. The scaled vortensity field is shown. The vertical direction is azimuthal, and horizontal is radial. The vortices have size $\sim H$.*

In the self-gravitating case F_{srg} , the inter-vortex distance oscillates with period $50P_0$ and there are two close-encounters before merging. This is analogous to the survival

3.5 Simulations of co-orbital Kida vortices

of the vortex pair towards the end of the disc-planet simulation for $Q_o = 3$. Given a maximal separation of $\simeq 1.4$ and that the vortices are circulating near unit radius, the maximal angular separation is $\simeq \pi/2$. The minimal separation during the first two close encounters is $\simeq 0.6$, or about $12H$ which implies that the vortices are on tadpole orbits. As the minimum separation is much larger than the typical vortex size, merging does not occur during the first two close encounters.

The experiment in Fig. 3.21(b) varies the strength of self-gravity via Q_1 . The reference case has $Q_1 = 8$. Doubling Q_1 to $Q_1 = 15.9$ (case M1) weakens self-gravity and vortices merge within few tens of P_0 . For $Q_1 = 5.3$ (case M2) the oscillation period is $\simeq 40P_0$ and maximal separation is 1.55, larger than for $Q_1 = 8$, as is the first minimum separation: the increased self-gravity has enhanced the mutual repulsion of vortices. The two-vortex configuration in M2 lasts until the end of the simulation, but there is a secular decrease in the minimum separation (of 0.6 at $t = 20P_0$ and 0.45 at $t = 180P_0$) due to vortex migration. Merging will likely occur eventually. Consistent with the behaviour seen for gap edge vortices, increasing self-gravity delays merging.

The final experiment varies the initial radial separation of the vortex perturbations. Results are shown in Fig. 3.21(c). Increasing X_0 to $0.1H$ from the reference case, the first minimum separation decreases to 0.45 from 0.6: vortices approach each other more closely, but still repel and undergo co-orbital dynamics. The increased oscillation amplitudes imply a larger tadpole orbit. For $X_0 = 0.2H$ and $0.3H$, vortex separation decreases more rapidly and become small enough for vortex merging. However, for $X_0 = 0.6H$, vortices simply circulate past by each other and no merging occurs, despite reaching a similar minimal separation as when $X_0 = 0.2H, 0.3H$. Vortices should collide head on if they are to merge.

3.5.4 Vortices as co-orbital planets

Simulations above indicate that self-gravitating vortex pairs behave like co-orbital planets. This means there exists an initial radial separation (or impact parameter) below which vortices execute horseshoe orbits relative to each other. Analytical and numerical work indicates that vortices merge if their centroids are within $d \sim 3s$ of each other (e.g. Zabusky et al., 1979; Melander et al., 1988), where s is the vortex size. Although one cannot assume the same critical d/s apply here because the situations are very different, it is still reasonable to expect merging if vortices can reach within some critical distance of one another.

The results above can be anticipated from existing treatments of co-orbital dynam-

3. VORTEX MODES IN SELF-GRAVITATING DISCS

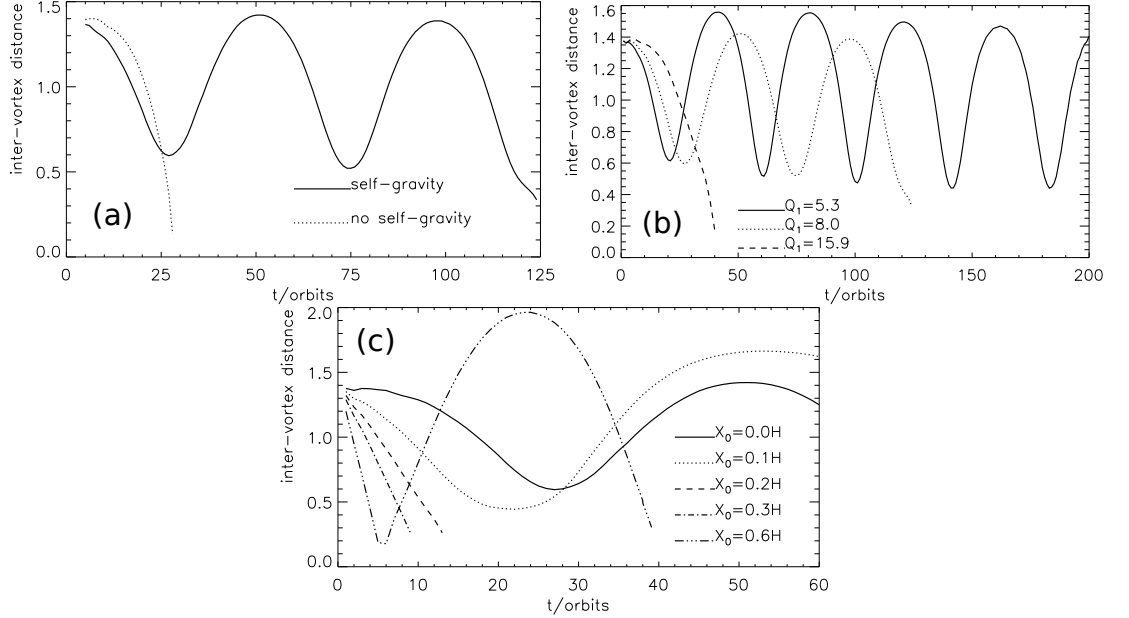


Figure 3.21: Inter-vortex distances for vortex pair simulations: (a) With and without self-gravity for $Q_1 = 8$ (simulations *Fsg* and *Fnsg*) (b) For discs of different masses characterised by Q_1 , the Keplerian Toomre parameter at $r = 1$, self-gravity being included (simulations *Fsg*, *M1*, and *M2*). (c) Vortices input with varying initial radial separation, X_0 , with self-gravity included for disc models with $Q_o = 8$ (simulations *Fsg*, *S1*, *S2*, *S3*, and *S4*).

ics. The first is Murray & Dermott (2000)’s model of the co-orbital satellites of Saturn, the Janus-Epimetheus system. The governing equation from Murray & Dermott (2000) gives a relationship between two configurations, the ‘final’ configuration (subscript f) and the ‘initial’ configuration (subscript i) in the form

$$\left(\frac{\Delta r_i}{r_0}\right)^2 - \left(\frac{\Delta r_f}{r_0}\right)^2 = -\frac{4}{3}q[H(\theta_i) - H(\theta_f)], \quad (3.41)$$

$$H(\theta) = [\sin(\theta/2)]^{-1} - 2\cos\theta - 2$$

where Δr is the radial separation of the satellites, r_0 the average orbital radius (assumed fixed, thereby ignoring migration), θ their angular separation and $q = \mathcal{M}/M_*$, with \mathcal{M} being the sum of the satellite masses. It is assumed $q \ll 1$.

Eq. 3.41 should apply to vortex pairs if there is significant mutual gravitational interaction. Their final radial separation is assumed zero, $\Delta r_f = 0$, when the vortices are at their minimum angular separation. Inserting $\theta_i = \pi/2$, $\Delta r_i = 7.1 \times 10^{-3}$ (corresponding to the initial conditions deduced for *Fnsg*), $r_0 = 1$ and $q = 2.3 \times 10^{-5}$

3.5 Simulations of co-orbital Kida vortices

(corresponding to the measured mass of the negative absolute vorticity region in Fsg), Eq. 3.41 gives $\theta_f \simeq 0.41$ or a minimal angular separation of $\sim 8H$ so merging is not expected if vortices have size H . This estimate is lower than the observed value of 0.6, but this q is a lower limit on the vortex mass. Inserting $q = 1.2 \times 10^{-4}$ gives $\theta_f = 0.6$ assuming all other parameters remain the same. This implies the effective gravitational mass of the vortex should be 20 Earth masses.

Eq. 3.41 is illustrated in Fig. 3.22. For fixed X_0 and increasing q , (e.g. as a by product of increasing the disc mass) the minimal separation Y increases, eventually becoming too large for merging to occur. For fixed q , Y decreases as X_0 increases. This is similar to a test particle on a horseshoe orbit in the restricted three body problem. The larger its impact parameter, the closer it approaches the secondary mass.

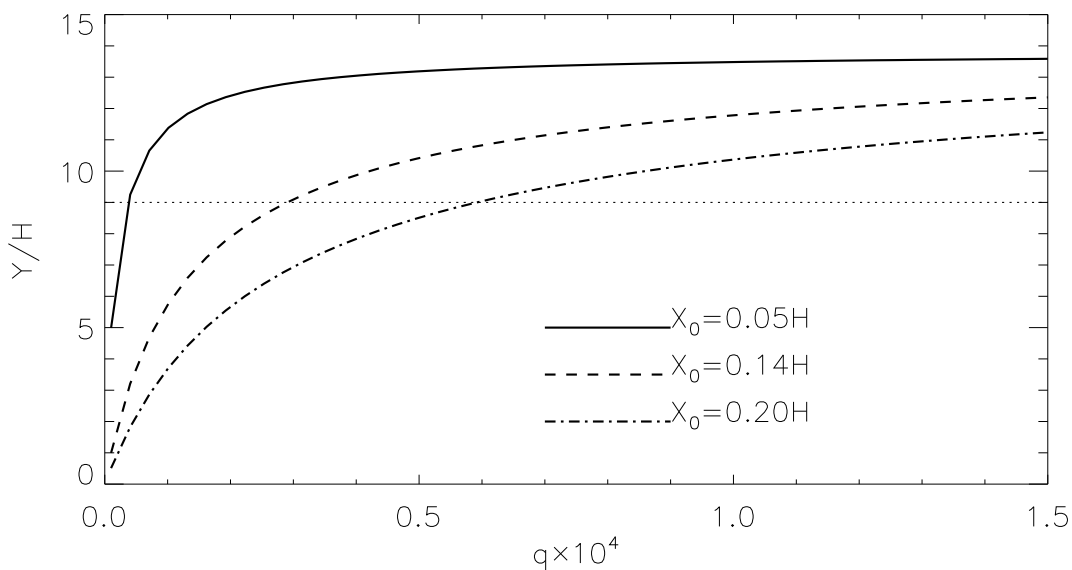


Figure 3.22: *Merging conditions based on Murray & Dermott's model of co-orbital satellites. Y is the minimum inter-vortex separation and the horizontal line is a hypothetical critical separation below which merging occurs. Thus merging is less likely for larger q .*

Mutual horseshoe turns may also be analysed in a shearing sheet approximation. This is just the generalisation of the result for a test particle and a planet, to the two-planet case (Appendix A.1). Since the vortices have characteristic sizes of the local scale-height, it is convenient to non-dimensionalise lengths by H , in the expression for

3. VORTEX MODES IN SELF-GRAVITATING DISCS

the maximum initial radial separation of vortices that allow horseshoe turns:

$$\hat{x}_0 < \frac{1}{h} \left(\frac{8q}{3\hat{y}h} - \frac{q^2}{3\hat{y}^4 h^4} \right)^{1/2}, \quad (3.42)$$

where $\hat{x}_0 = X_0/H$, $\hat{y} = Y/H$. Given X_0 , Eq. 3.42 can be inverted to give the range of possible Y . This inequality is displayed in Fig. 3.23. There exists a critical $X_0 = X_s$ beyond which vortices do not execute horseshoe turns and instead circulate past each other. There may also exist a value $X_{0,c}$ such that $X_{0,c} < X_s$ and for $X_{0,c} < X_0 < X_s$, vortices execute U-turns and the values of Y attained are sufficiently small to allow merging. This happens only when the vortices have small enough mass. When $X_0 < X_{0,c}$, Y can be larger than critical, so that merging does not occur.

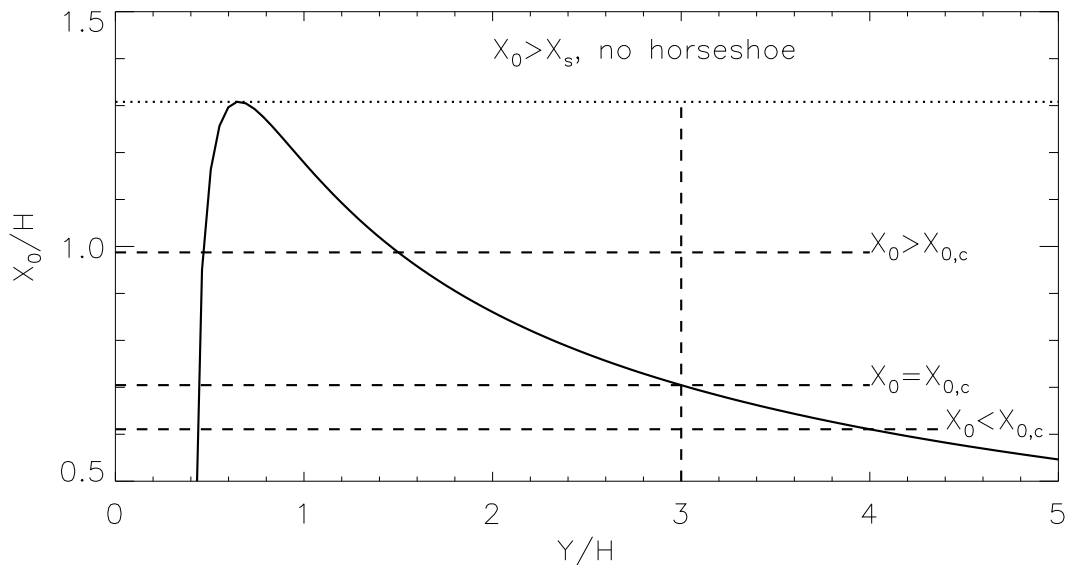


Figure 3.23: Merging conditions based on shearing sheet dynamics. For a given X_0 , the allowed values of minimum separation Y lies between the intersection of the horizontal line $X_0/H = \text{constant}$ and the solid curve. The vertical dashed line is a hypothetical critical separation, below which merging occurs.

These simple models give qualitatively the same results as vortex pair and disc-planet simulations, and serve to explain the effect of self-gravity delaying merging by interpreting vortices as co-orbital planets that execute mutual horseshoe turns, thereby imposing a minimum inter-vortex separation (mainly in the azimuthal direction). If this minimum separation is still larger than a critical separation known to exist for vortex merging, then merging cannot occur. Hence, multiple-vortex configurations can

be sustained longer as the strength of self-gravity is increased.

3.6 Implications on vortex-induced migration

Chapter 2 established that vortices at gap edges can lead to brief phases of rapid inwards migration of giant planets. In summary, gap edges become unstable to vortex modes leading to vortex mergers on dynamical timescales, in turn resulting in a large-scale vortex circulating at either gap edge. Upon approaching the planet, the inner vortex can execute a horseshoe turn and move outwards across the gap, gaining angular momentum. Thus, the planet loses angular momentum and is scattered inwards.

For comparison purposes, the non-self-gravitating simulations performed in Chapter 2 were repeated with disc self-gravity. Fig. 3.24 shows the orbital migration of a Saturn-mass planet in discs with total mass $M_d = 0.035M_*$ and $0.025M_*$. For $M_d = 0.025M_*$, the surface density perturbations are also compared in Fig. 3.25. The $M_d = 0.035M_*$ non-self-gravitating disc was the fiducial run in Chapter 2. Including self-gravity delays vortex-induced migration from $t = 60P_0$ to $t = 85P_0$. The delay is consistent with both the stabilisation of low m modes, and the slower vortex merging induced by self-gravity.

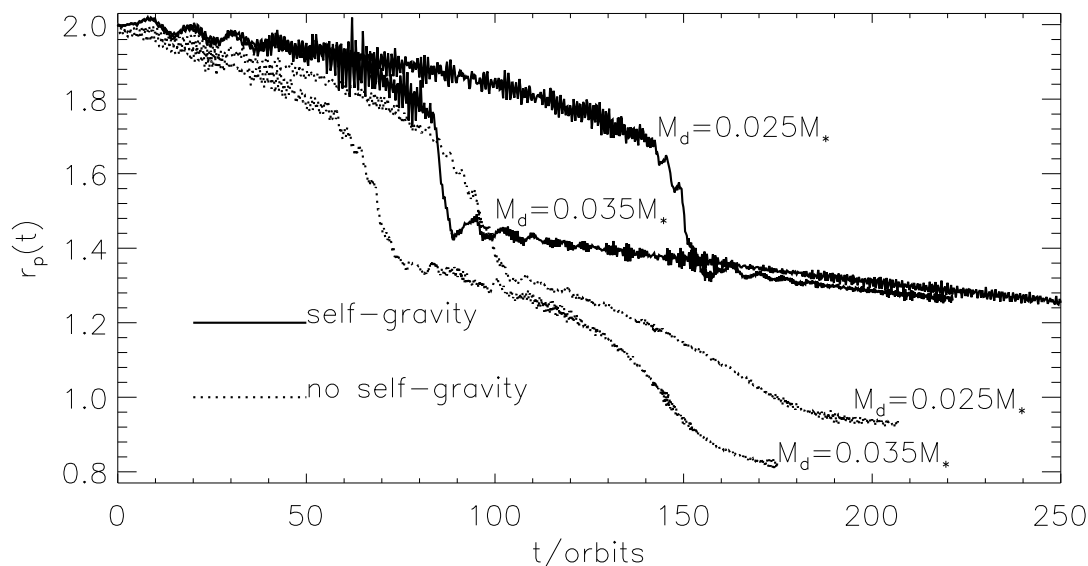


Figure 3.24: *Vortex-induced migration with (solid) and without (dotted) disc self-gravity. The discs are inviscid and initially with uniform surface density.*

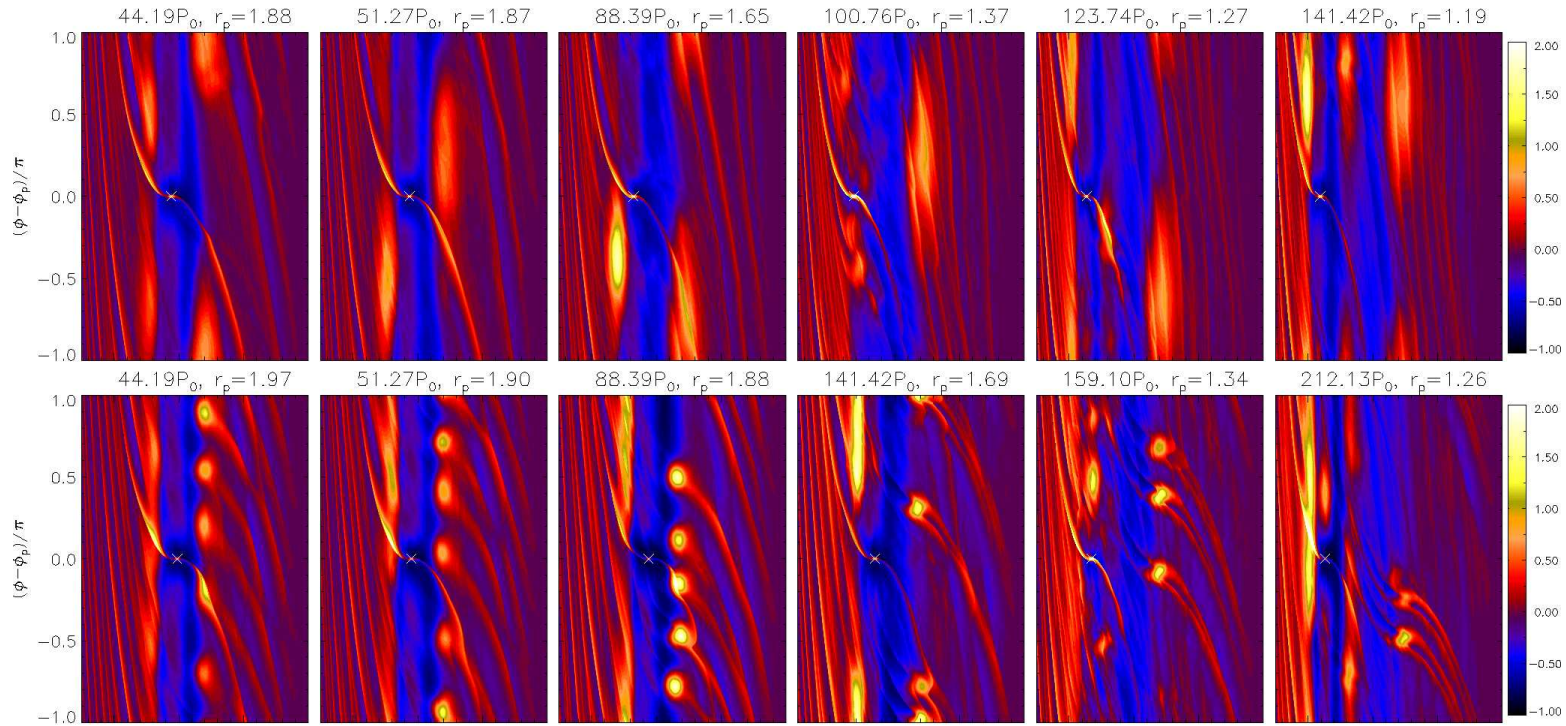


Figure 3.25: Vortex-induced migration in the $M_d = 0.025M_*$ disc (see Fig. 3.10), without self-gravity (top) and with self-gravity (bottom). The single vortex at the inner gap edge causes the planet to jump inwards when it scatters the planet. The formation of this vortex is slower in the self-gravitating case.

Consider $M_d = 0.025M_*$. In the self-gravitating disc, it takes longer for the inner vortex to build up (e.g. comparing $t = 88.4P_0$ in Fig. 3.25), disrupt the coorbital region and flow across the gap. This delays vortex-induced rapid migration, by about $50P_0$ compared to the non-self-gravitating case.

The extent of rapid migration is unaffected by self-gravity. Notice also the increased oscillations in $r_p(t)$ when self-gravity is included. This is because of the sustained multi-vortex configuration at the outer gap edge causing large oscillations in disc-planet torques, whereas without self-gravity these vortices would have merged.

After the first scattering event, migration stalls while the planet opens a gap at its new orbital radius and vortex formation recurs. With self-gravity included, the re-formation of a single vortex takes longer and it is narrower in radial extent than the non self-gravitating case. When the thinner self-gravitating vortex passes by the planet, little vortex material splits off from the main vortex and flows across the gap, unlike the non-self-gravitating vortex where material breaks off more easily (Chapter 2). This is probably due to the self-gravity of the vortex. Hence, there is a longer stalling period when self-gravity is included. Thus the net effect of self-gravity is to slow the migration in this example.

For the setup used in Chapter 2, a second fast migration episode was not seen within the simulated time when self-gravity was included, but it may eventually occur. The total practical simulation time of a few hundred orbits is still very short compared to disc lifetimes. However, for the $Q_o = 4$ disc model used in previous parts of this Chapter, two episodes of rapid migration occurs (Fig. 3.26). Self-gravity does not change the physical nature of vortex-induced, non-smooth migration.

3.7 Summary and discussion

The effect of disc self-gravity on vortex-forming instabilities associated with planetary gaps were explored. Analysis shows vortex modes are stabilised by self-gravity through its effect on the linear mode when the background remains fixed. This aspect has been confirmed by linear calculations. Linear calculations showed that the vortex forming modes with the highest growth rate shift to higher m with increasing disc mass. This is due to the combined effect of self-gravity through the response and through the background state.

Hydrodynamic simulations with and without self-gravity for a range of disc masses were performed. More vortices form as the disc mass increased in accordance with linear calculations. However, for sufficiently strong self-gravity, the vortex modes are

3. VORTEX MODES IN SELF-GRAVITATING DISCS

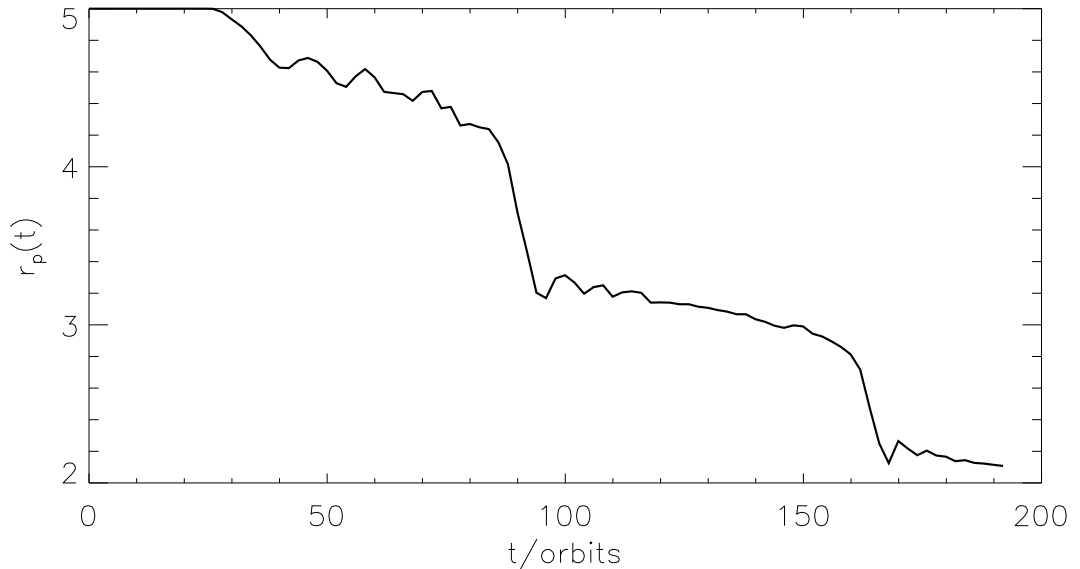


Figure 3.26: *Repeated episodes of vortex-induced migration in a self-gravitating disc, for the $Q_o = 4$ disc model used in previous sections.*

suppressed and global spiral modes develop instead. Self-gravity delays vortex merging and multi-vortex configurations can be sustained for longer with increasing disc mass, allowing vortices to evolve individually.

The nature of post-merging vortices is also affected by self-gravity. With weak self-gravity (i.e. $M_d \leq 0.024M_*$), a single vortex, extended in azimuth forms and circulates at the outer gap edge. For $M_d = 0.031M_*$ the final configuration is a vortex pair at the outer gap edge, each localised in azimuth. In this case a vortex, containing on the order of $20M_\oplus$, is gravitationally bound. The effect of their gravitational influence on the rest of the disc is to redistribute mass radially whereas for lower disc masses, redistribution is restricted to being azimuthal.

The internal flow in these self-gravitating, localised vortices adjusts so that they are not destroyed by the background shear, taking on a structure similar to that of a vortex resulting from perturbing the disc with the Kida solution (Kida, 1981). Such vortices form in discs much less massive than that required for direct disc fragmentation. Vortices can trap dust particles due to their association with pressure maxima (Bracco et al., 1999; Chavanis, 2000). Increasing self-gravity leads to vortices of stronger density contrast (Fig. 3.16), so self-gravitating vortices may be more effective at collecting solid particles, and assist planetesimal formation.

Supplementary simulations of Kida-like vortices were performed to understand the effect of self-gravity on inter-vortex interactions. In a self-gravitating disc, coorbital vortex pairs behave like coorbital planets and can execute mutual horseshoe turns. As a consequence there exists a minimum inter-vortex distance. Merging is then avoided if this minimal separation is still larger than a critical separation below which vortex merging occurs.

The effect of self-gravity found here, should also be applicable to other types of structured features in a protoplanetary disc that could support vortex forming instabilities. Such a possibility is the boundary region between a dead zone and active region of the disc (Lyra et al., 2009).

The basics of vortex-induced migration remains unchanged by self-gravity. With self-gravity, the resistance to forming a single large vortex results in such migration being delayed. Self-gravitating vortices are less effective in scattering the planet because they do not disrupt the co-orbital region as significantly as their non-self-gravitating counterparts. It can be said that in the regime of disc masses where vortices form and are significantly affected by self-gravity, vortex-induced migration is slowed down.

3. VORTEX MODES IN SELF-GRAVITATING DISCS

Chapter 4

Edge modes in self-gravitating discs

Numerical experiments Chapter 3 showed that global spiral disturbances associated with gap edges can develop in massive discs. They arise from a linear instability associated with both the self-gravity of the disc and local vortensity maxima which coincide with gap edges. They can be physically interpreted as disturbances localised around a vortensity maximum that further perturb gravitationally the smooth parts of the disc by exciting waves at Lindblad resonances. The angular momentum carried away reacts back on the edge disturbance so as to destabilise it.

These edge modes are studied in this Chapter. The disc-planet models are identical to that used in Chapter 3, but briefly reproduced in §4.1 for ease of reference. In §4.2, the existence of edge modes and its basic features are illustrated with a numerical simulation. An analytic discussion and interpretation of edge modes is then presented in §4.3—§4.4. In §4.5, linear calculations for various disc models confirm the existence of edge dominated modes for low values of the azimuthal mode number. The physical picture of edge modes presented in the analytical discussion is consistent with these numerical solutions.

In §4.6 results from hydrodynamic simulations for a range of disc masses are discussed. These are all stable in the absence of the planet. However, they exhibit low m edge modes once a planet-induced gap is present. The form and behaviour of these is found to be in accord with linear theory. The spiral arms associated with the edge instabilities are shown to produce fluctuating torques acting on the planet. Fast inwards type III migration may occur, but *outwards* migration due to interaction with spiral arms is also observed. §4.8 concludes this Chapter.

4.1 Disc model

To apply angular momentum conservation in the analytical description of edge modes, three-dimensional disc models are required. This is because although the disc material may be considered to be confined to a thin sheet, the gravitational potential it generates is still three-dimensional. Thus, it is important to consider the self-gravity part as fully three-dimensional. This is governed by the Poisson equation

$$\nabla^2\Phi = 4\pi G\rho,$$

where ρ is the three dimensional mass density and Φ is the associated gravitational potential which can be found via

$$\Phi(\mathbf{r}) = -G \int_{\mathcal{D}} \frac{\rho(\mathbf{r}')d^3\mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|}, \quad (4.1)$$

where \mathcal{D} is the domain where ρ is non zero. This coincides with the disc domain when this is not separated from external material. The full 3D hydrodynamic equations are listed in §1.4 and is the basis for analytical work. In analytic discussions, fluids with a general barotropic equation of state are considered, for which $p = p(\rho)$ and $dp/d\rho = c_s^2$.

4.1.1 Razor thin discs

For numerical work, the disc is considered two-dimensional and only its gravitational potential in the plane is used. The governing equations are therefore those listed in §1.4.6, which were used throughout Chapter 3. In numerical experiments, the locally isothermal equation of state is used with fixed aspect-ratio $h = 0.05$. The physical setup and notation is the same as Chapter 3, but simulations will be presented for a range of viscosity and softening parameters.

The discs occupy $r = [r_i, r_o] = [1, 10]$ and are parametrised by Q_o , the Keplerian Toomre Q value at the outer boundary. The initial surface density profile is given by Eq. 3.1 (Chapter 3). Recall that 2D self-gravity requires the use of a gravitational softening length (see §1.4.6, Eq. 1.21) to prevent a singularity and approximately account for the vertical dimension. The softening length prescription $\epsilon_g(r) = \epsilon_{g0}hr$ is used, and $\epsilon_{g0} = 0.3$ set as a fiducial value, but values up to 1.0 will be considered.

The initial azimuthal velocity is found by assuming the centrifugal force balances forces due to stellar gravity, the disc's self-gravity and pressure gradient. Thus the

initial azimuthal velocity is given by

$$u_\phi^2 = \frac{r}{\Sigma} \frac{dp}{dr} + \frac{GM_*}{r} + r \frac{d\Phi}{dr}, \quad (4.2)$$

For the local isothermal equation of state adopted, the contribution due to the pressure is given by

$$\frac{r}{\Sigma} \frac{dp}{dr} = c_s^2 \left\{ -\frac{5}{2} + \frac{r\sqrt{r_i}(r+H_i)^{-3/2}}{2\left[1 - \sqrt{r_i/(r+H_i)}\right]} \right\}. \quad (4.3)$$

At $r = r_i$, for $h = 0.05$, this is $\sim 20c_s^2$ which is approximately 5% of the square of the local Keplerian speed so that the contribution of the pressure force is indeed minor when compared to that arising from the gravity of the central star.

The initial radial velocity is set to $u_r = 3\nu/r$, corresponding to the initial radial velocity of an one-dimensional Keplerian accretion disc with $\Sigma \propto r^{-3/2}$ and uniform kinematic viscosity. The fiducial value $\nu = 10^{-5}$ in dimensionless units is adopted, but results for $\nu = 10^{-7}$ to $\nu = 5 \times 10^{-5}$ will be presented. The disc is evolved for ~ 280 Keplerian orbits at r_i before introducing the planet.

When a planet of mass M_p is introduced, it is inserted in a circular orbit, under the gravity of both the central star and disc, at a distance $r_p = 5$ from the central star. Its gravitational potential is then ramped up over a time interval $10P_0$. Recall P_0 is the circular Keplerian orbital period at the planet's initial radius.

4.2 Properties of edge modes

The existence of spiral modes associated with planetary gaps is first demonstrated through numerical simulations. Numerical details are given in §4.6 where additional hydrodynamic simulations of this type as well as others, where the planet is allowed to migrate, will be investigated more fully.

4.2.1 Example simulation

Described here are results obtained for the $Q_o = 1.5$ disc model, corresponding to a total disc mass of $M_d = 0.063M_*$, in which there is an embedded planet with mass $M_p = 3 \times 10^{-4}M_*$. The uniform kinematic viscosity was taken to be $\nu = 10^{-5}$ in dimensionless units. For this simulation the planet was held on a fixed circular orbit. Without a perturber the disc is gravitationally stable because the $Q_{\text{Kep}} = 2.62, 1.5$ near the

4. EDGE MODES IN SELF-GRAVITATING DISCS

planet and the outer boundary, respectively. General non-axisymmetric gravitational instabilities occur in smooth discs when $Q_{\text{Kep}} \lesssim 1.5$.

The profile of the gap opened by the planet is shown in Fig. 4.1 in terms of the azimuthally averaged surface density $\langle \Sigma \rangle_\phi$, vortensity $\langle \eta \rangle_\phi$ and Toomre Q . The latter is calculated using the azimuthally averaged epicycle frequency. The surface density gap corresponds to a decrease of surface density by $\simeq 20\%$ relative to the initial profile.

Notice that the vortensity and Toomre Q profiles are qualitatively similar. The neighbourhoods of the inner and outer gap edges both contain maxima and minima of Q and η . For typical disc models structured by disc-planet interactions, such as those illustrated in Fig. 4.1, local maxima/minima in Q and η approximately coincide. Vortensity generation as material flows through planet-induced shocks leads to vortensity maxima. Since the axisymmetric Toomre parameter can be directly related to vortensity, $Q = (c_s/\pi G)(2\Omega\eta/\Sigma)^{1/2}$, the resulting Q and η profiles are very similar in their extrema locations.

Fig. 4.1 shows that in the neighbourhood of the outer gap edge, the maximum and minimum values of Q occur at $r = 5.45$ and $r = 5.75$, and are $Q = 4.2$ and $Q = 1.75$ respectively. In the region of the inner gap edge, the maximum and minimum values of Q occur at $r = 4.55$ and $r = 4.25$. These are $Q = 4.75$ and $Q = 2.55$ respectively. The characteristic width of the gap edge is the local scale-height. On average $Q \sim 2$ for $r > 6$. Since $Q > 1$ everywhere, the disc is stable against local axisymmetric perturbations. Furthermore, since $Q > 1.5$ in the gap region, gravitational instabilities should not be directly associated with the local value of Q . Instead, instability is associated with the global Q -profile.

The simulation shows that both gap edges become unstable. The surface density contours at $t = 50P_0$ are shown in Fig. 4.2. A mode with azimuthal wavenumber $m = 3$ and relative surface density perturbation $\Delta\Sigma/\Sigma \simeq 0.4$ is associated with the inner edge and a mode with $m = 2$ and $\Delta\Sigma/\Sigma \simeq 0.9$ associated with the outer edge. The modes have become nonlinear, with the development of shocks, on dynamical time-scales. Spirals do not extend across the gap indicating effective gap opening by the planet and disc self-gravity is not strong enough to connect the spiral modes on either side of r_p .

Plots of the radial dependence of the pattern rotation speed¹, Ω_{pat} , obtained from the $m = 2$ component of the Fourier transform of the surface density, together with the azimuthally averaged values of Ω and $\Omega \pm \kappa/2$ are presented in Fig. 4.3. The mode

¹The pattern speed is $(1/m)\partial\varphi/\partial t$, where $\tan\varphi = -\text{Im}(a_m)/\text{Re}(a_m)$. Recall $a_m = \int \Sigma e^{-im\phi} d\phi$ is the Fourier amplitude. Calculating the pattern speed thus require $\partial\Sigma/\partial t$. However, $\partial\Sigma/\partial t = -\nabla \cdot (\Sigma\mathbf{u})$ from the continuity equation. Instantaneous time-derivatives can therefore be calculated from spatial derivatives. This gives the pattern speed from a single data output.

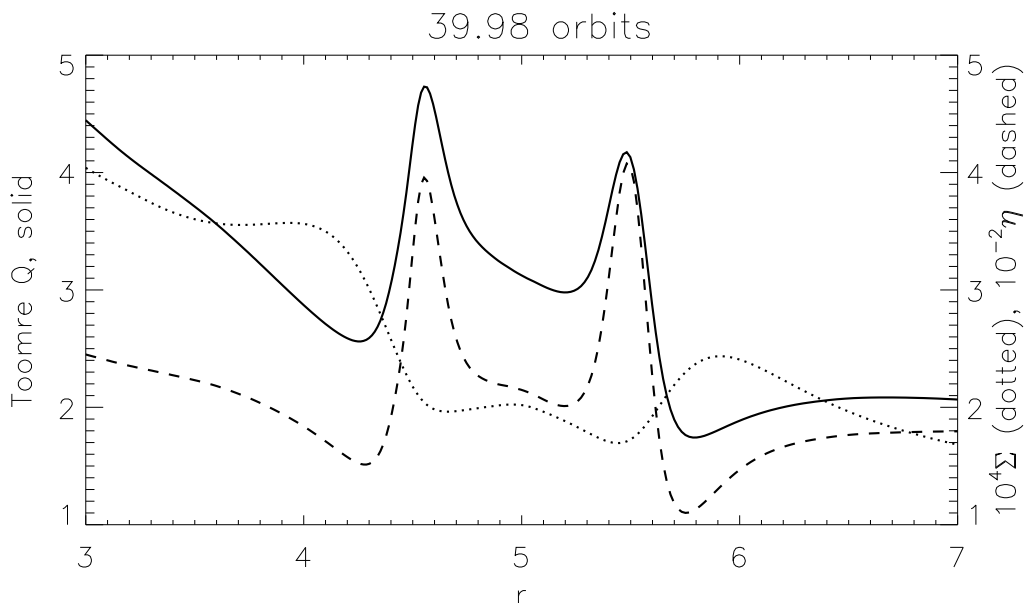


Figure 4.1: Gap profile produced by a Saturn-mass planet embedded in a disc with $Q_o = 1.5$. The azimuthally averaged surface density (dotted line), vortensity (dashed line) and Toomre Q parameter (solid line) are plotted. Note that the Q profile shows a maximum and minimum associated with each gap edge. The planet is in a fixed circular orbit located at $r = 5$.

associated with the outer gap edge is of particular interest, because the $m = 2$ spiral mode has more than twice the relative surface density amplitude compared to the inner spiral mode.

By necessity, the Fourier transform includes features due to the planetary wakes and inner disc spiral modes as well as the dominant outer disc mode. On average, $\Omega_{\text{pat}} \sim 0.08$. This value was also explicitly checked by measuring the angle through which the spiral pattern rotates in a given time interval. There is a co-rotation point in the outer disc at $r = 5.5$ where $\Omega = \Omega_{\text{pat}} \sim 0.078$, i.e. at the local vortensity maximum (surface density minimum) of the basic state, which is just within the gap (see Fig. 4.1). This is consistent with pattern speeds obtained from linear calculations presented later. Fig. 4.3 indicates another co-rotation point at $r \simeq 4.7$ but this is due to inclusion of features associated with the planetary wakes.

Unstable spiral modes identified here are referred to as *edge modes*. Their properties can be compared to those of *groove modes* in particle discs (Sellwood & Kahn, 1991) or fluid discs (Meschiari & Laughlin, 2008). Sellwood & Kahn describe groove modes as gravitational instabilities associated with the coupling of disturbances associated with

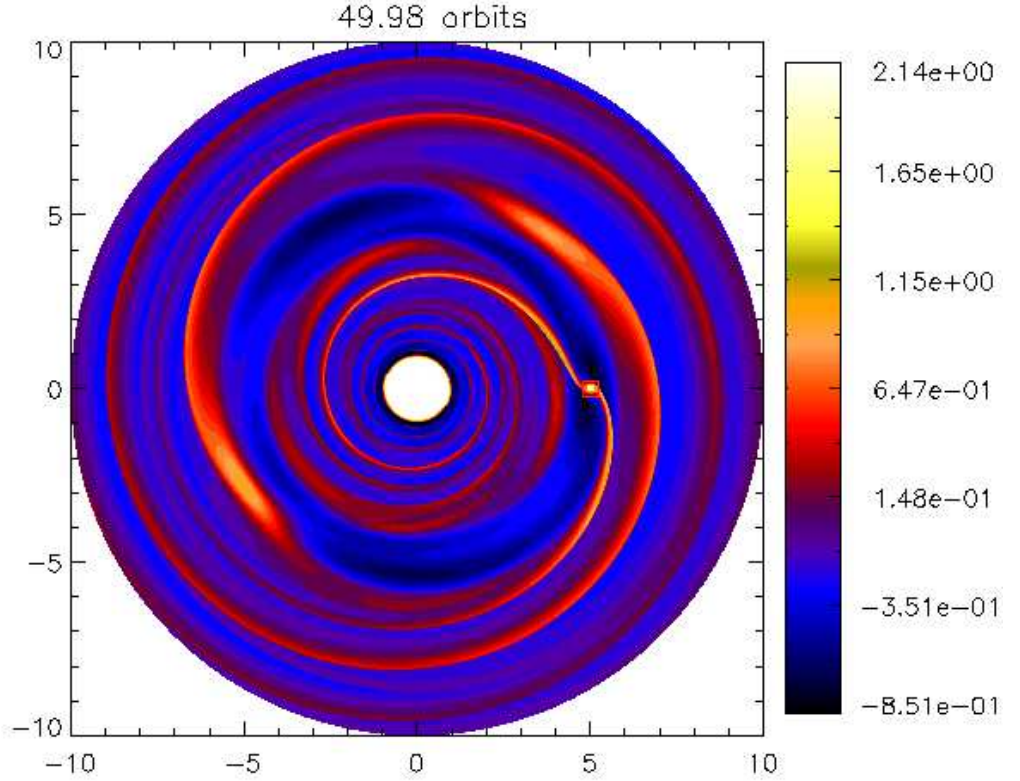


Figure 4.2: *Relative surface density perturbation for $Q_o = 1.5$ at $t = 50P_0$.*

two edges across the gap between them. However, the analytic description of the edge mode instability, presented later, is a coupling between a disturbance at a single gap edge and the disturbance it excites in the adjoining smooth disc away from co-rotation. The other gap edge plays no role in this description. Although both types of mode are associated with vortensity maxima, the groove modes described above would have co-rotation at the gap centre midway between the edges, whereas for the planetary gap co-rotation is at the gap edge.

This is significant because it implies differential rotation between the spiral pattern and the planet. In addition, unstable modes on the inner and outer gap edges need not have the same m (Fig. 4.2 shows $m = 3$ on the inner edge and $m = 2$ on the outer edge). If there was a groove mode with co-rotation at the gap centre, then the spiral pattern would co-rotate with the planet and the coupled edges must have disturbances dominated by the same value of m .

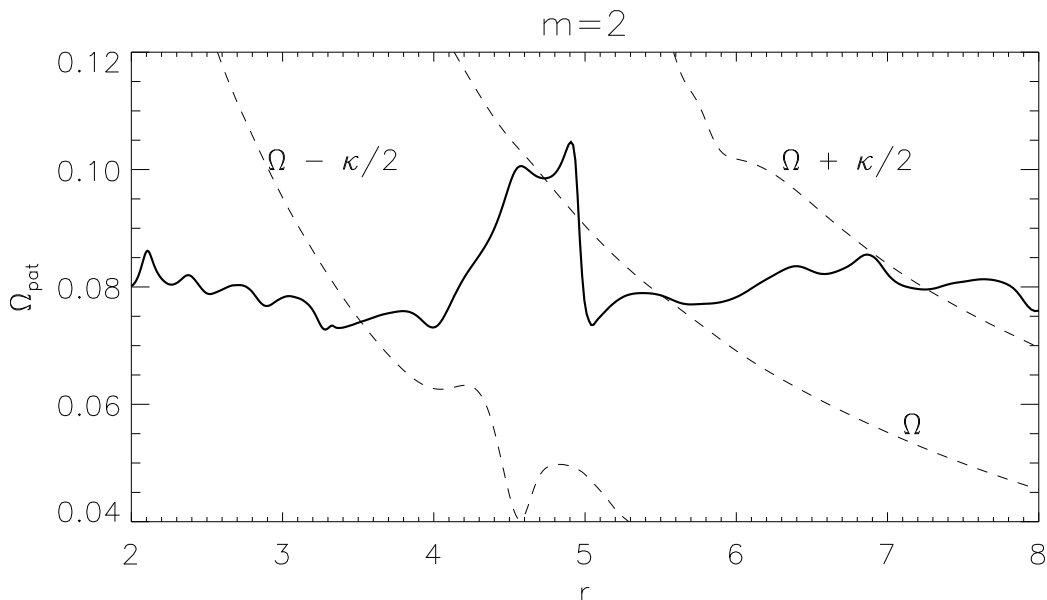


Figure 4.3: Same case as Fig. 4.2, but shown here is the radial form of the pattern speed found from the $m = 2$ component of the Fourier transform of the surface density (solid). Note that this mode is associated with the outer disc and, apart from within the gap, this mode appears to have a stable pattern speed of $\Omega_{pat} \sim 0.078$ corresponding to a corotation point at $r = 5.5$, corresponding to vortensity maximum in the gap profile. Plots of Ω and $\Omega \pm \kappa/2$ are also given (dashed).

4.2.2 Comparison to vortex instabilities

It is interesting to compare edge modes to the localised vortex-forming instabilities studied in Chapter 2–3, which develop in non-self-gravitating or weakly self-gravitating discs near gap edges. However, the vortex instability requires low viscosity. For comparison purposes, here a kinematic viscosity of $\nu = 10^{-6}$ is imposed. The behaviour of disc models with $Q_o = 1.5$ and $Q_o = 4.0$ are compared below, the former having strong self-gravity and the latter weak self-gravity.

Fig. 4.4 shows two types of instabilities depending on disc mass. The lower mass disc with $Q_o = 4$ develops 6 vortices localised in the vicinity of the outer gap edge. The more massive disc with $Q_o = 1.5$ develops edge modes of the type identified in the earlier run with $\nu = 10^{-5}$. Here, the $m = 3$ spiral mode is favoured because of the smaller viscosity coefficient used¹

Vortex modes are said to be local because instability is associated with flow in the vicinity of co-rotation (Lovelace et al., 1999) and does not require the excitation

¹This results in a sharper gap profile, which generally supports higher m .

4. EDGE MODES IN SELF-GRAVITATING DISCS

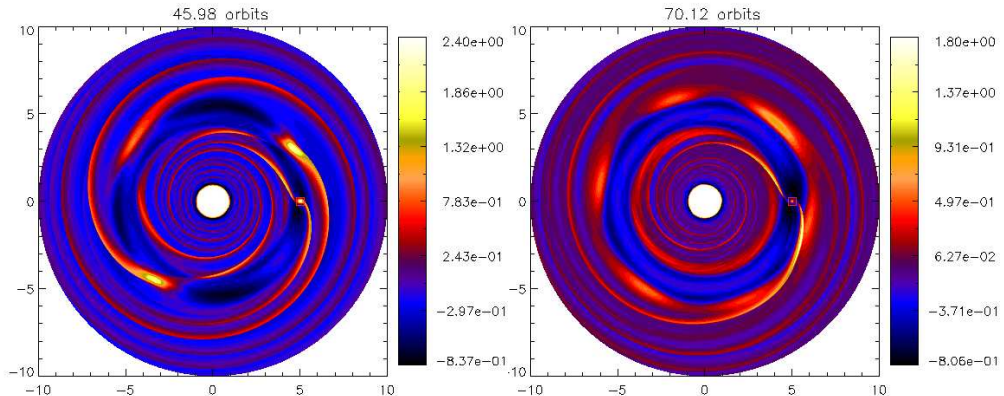


Figure 4.4: *Relative surface density perturbation for the $Q_o = 1.5$ disc (left) and the $Q_o = 4.0$ disc (right) with $\nu = 10^{-6}$. These plots show that the gap opened by a Saturn-mass planet supports vortex instabilities in low mass discs and spiral instabilities in sufficiently massive discs.*

of waves. Edge modes are said to be global because instability requires interaction between an edge disturbance and waves launched at Lindblad resonances, away from co-rotation in the smooth part of the disc. Vortices perturb the disc even without self-gravity (Paardekooper et al., 2010a). Although the vortex mode in Fig. 4.4 shows significant wave perturbations in the outer parts of the disc, these should be seen as a consequence of unstable vortices forming around co-rotation, rather than the cause of a linear instability.

Vortex modes have co-rotation close to local vortensity *minima* in the undisturbed gap profile. The self-gravitating vortex modes are localised with higher m being dominant ($m > 5$). On the other hand the spiral modes found here are global with $m \leq 3$. Edge modes make the gap less identifiable than vortex modes do because the former, with corotation closer to r_p , protrudes the gap edge more. While the edge mode only takes a few P_0 to become non-linear with spiral shocks attaining comparable amplitudes to the planetary wakes, the vortex mode takes a significantly longer time (the plot for $Q_o = 4$ is $30P_0$ after gap formation). This can be understood as waves emitted by the vortex modes tends to stabilise them, whereas wave emission drives the edge mode instability.

4.3 Analytical discussion of edge modes

The fiducial disc simulated above is massive with $M_d \sim 0.06M_*$ but it is still stable against gravitational instabilities in the sense that $Q \geq 1.5$ everywhere and the disc without an embedded planet, which has a smoothly varying surface density profile, does not exhibit the spontaneous growth of spiral instabilities. A planet of Saturn's mass ($M_p = 3 \times 10^{-4}M_*$) is expected to only open a partial gap ($\sim 30\%$ deficit in surface density), but even this is sufficient to trigger a $m = 2$ spiral instability.

The case above thus suggests spiral modes may develop under conditions that are not as extreme as those considered by Meschiari & Laughlin (2008). They used a prescribed gap profile with a gap depth of 90% relative to the background, which is three times deeper than the simulation above. Their gap corresponds to $M_p = 0.002M_*$, although a planet potential was not explicitly included. In both their model and the fiducial case, the gap width is $\simeq 2r_h$, but since Meschiari & Laughlin effectively used a two Jupiter mass planet, at the same r_p their gap is about 1.9 times wider than for a Saturn mass planet.

This section and the next are devoted to a theoretical discussion of spiral modes associated with planetary gap edges. This work is based on the analysis of the governing equations for linear perturbations. As angular momentum balance is important for enabling small perturbations to grow unstably, one begins by formulating the conservation of angular momentum for linear perturbations.

4.3.1 The conservation of angular momentum for a perturbed disc

A conservation law can be derived for angular momentum associated with the perturbations of a disc. This enables the angular momentum density and flux to be identified within the framework of linear perturbation theory. The behaviour of these quantities is found to be important for indicating the nature of the angular momentum balance in a system with a neutral or weakly growing normal mode and how positive (negative) angular momentum fluxes associated with wave losses may drive the instability of a disturbance that decreases (increases) the local angular momentum density.

4.3.2 Barotropic discs

The linearised equation of motion in three dimensions for the Lagrangian displacement ξ for a differentially rotating fluid with a barotropic equation of state and with self-

4. EDGE MODES IN SELF-GRAVITATING DISCS

gravity included is given by

$$\frac{D^2 \boldsymbol{\xi}}{Dt^2} + 2\Omega \hat{\mathbf{k}} \wedge \frac{D\boldsymbol{\xi}}{Dt} + r\hat{\mathbf{r}} (\boldsymbol{\xi} \cdot \nabla \Omega^2) = -\nabla S - \nabla \Phi_{\text{ext}}, \quad (4.4)$$

where

$$S = c_s^2 \frac{\delta \rho}{\rho} + \delta \Phi \quad (4.5)$$

(see eg. Lynden-Bell & Ostriker, 1967; Lin et al., 1993). A cylindrical polar coordinate system (r, ϕ, z) is adopted here. Perturbations to quantities are denoted with δ . The operator $D/Dt \equiv \partial/\partial t + \Omega \partial/\partial \phi$ is the convective derivative following the unperturbed motion and $\hat{\mathbf{k}}$ is the unit vector in the z direction which is normal to the disc mid-plane. For perturbations that depend on ϕ through a factor $\exp(im\phi)$, m being the azimuthal mode number, assumed positive, the operator D/Dt reduces to the operator $(\partial/\partial t + im\Omega)$. In addition, for a barotropic equation of state, $\Omega = \Omega(r)$ depends only on the cylindrical radius. Then Eq. 4.4 becomes

$$\frac{\partial^2 \boldsymbol{\xi}}{\partial t^2} + 2\Omega \hat{\mathbf{k}} \wedge \frac{\partial \boldsymbol{\xi}}{\partial t} + 2im\Omega \frac{\partial \boldsymbol{\xi}}{\partial t} + 2im\Omega^2 \hat{\mathbf{k}} \wedge \boldsymbol{\xi} + r\hat{\mathbf{r}} (\boldsymbol{\xi} \cdot \nabla \Omega^2) - m^2 \Omega^2 \boldsymbol{\xi} = -\nabla S - \nabla \Phi_{\text{ext}}. \quad (4.6)$$

The Eulerian density perturbation is given by

$$\delta \rho = -\nabla \cdot (\rho \boldsymbol{\xi}). \quad (4.7)$$

The perturbation to the disc's gravitational potential is given by linearising Poisson's equation to give

$$\nabla^2 \delta \Phi = 4\pi G \delta \rho. \quad (4.8)$$

An external potential perturbation Φ_{ext} has been added to the equations of motion to assist with the identification of angular momentum flux later on.

By taking the scalar product of Eq. 4.6 with $\rho \boldsymbol{\xi}^*$, and taking the imaginary part, after making use of Eq. 4.7 and Eq. 4.8, a conservation law that expresses the conservation of angular momentum for the perturbations can be derived:

$$\frac{\partial \rho_J}{\partial t} + \nabla \cdot (\mathbf{F}_A + \mathbf{F}_G + \mathbf{F}_{\text{ext}}) = T, \quad (4.9)$$

where

$$\rho_J \equiv -\frac{m\rho}{2}\text{Im}\left(\boldsymbol{\xi}^* \cdot \frac{\partial \boldsymbol{\xi}}{\partial t} + \Omega \hat{\mathbf{k}} \cdot \boldsymbol{\xi} \wedge \boldsymbol{\xi}^* + im\Omega|\boldsymbol{\xi}|^2\right) \quad (4.10)$$

$$\mathbf{F}_A = -\frac{m\rho}{2}\text{Im}(\boldsymbol{\xi}^* S) \quad (4.11)$$

$$\mathbf{F}_G = -\frac{m}{2}\text{Im}\left(\frac{1}{4\pi G}\delta\Phi\nabla\delta\Phi^*\right) \quad (4.12)$$

$$\mathbf{F}_{\text{ext}} = -\frac{m\rho}{2}\text{Im}(\boldsymbol{\xi}^* \Phi_{\text{ext}}) \quad (4.13)$$

and

$$T = \frac{m}{2}\text{Im}(\Phi_{\text{ext}}\delta\rho^*). \quad (4.14)$$

Here ρ_J is the angular momentum density and the angular momentum flux is split into three contributions, \mathbf{F}_A being proportional to the Lagrangian displacement, is the advective angular momentum flux, \mathbf{F}_G is the flux associated with the perturbed gravitational stresses and \mathbf{F}_{ext} is the flux associated with the external potential which is inactive for free perturbations.

The quantity T is a torque density associated with the external potential as can be seen from that fact that the real torque integrated over azimuth and divided by 2π is

$$-\frac{1}{2\pi}\int_0^{2\pi}\text{Re}(\delta\rho e^{im\phi})\text{Re}\left(\frac{\partial(\Phi_{\text{ext}}e^{im\phi})}{\partial\phi}\right)d\phi = T = \frac{m}{2}\text{Im}(\Phi_{\text{ext}}\delta\rho^*). \quad (4.15)$$

The $e^{im\phi}$ dependence has been reinstated on the LHS for clarity. This justifies the scalings used for the angular momentum density and fluxes.

4.3.3 Two dimensional discs

The form of the conservation law for a razor thin disc is obtained by integrating Eq. 4.9 over the vertical coordinate. For terms $\propto \rho$ the integrand is non zero only within the disc. $\boldsymbol{\xi}$ may be assumed to depend only on r and ϕ . The effect of the integration is simply to replace ρ and $\delta\rho$ by the surface density Σ and its perturbation $\delta\Sigma$ respectively. The fluxes become vertically integrated fluxes. Vertically integrating the term associated with the gravitational stresses, assuming the potential perturbation vanishes at large distances, gives

$$\int_{-\infty}^{\infty}\nabla \cdot \mathbf{F}_G dz = -\frac{m}{2}\text{Im}(\delta\Phi(r, \phi, 0)\delta\Sigma^*) = \frac{1}{r}\frac{\partial(rF_{Gr})}{\partial r}, \quad (4.16)$$

4. EDGE MODES IN SELF-GRAVITATING DISCS

where $\nabla^2 \delta\Phi = 4\pi G \delta\Sigma \delta(z)$ has been used. The angular momentum flux associated with self-gravity is

$$F_{Gr} = -\frac{m}{8\pi G} \text{Im} \int_{-\infty}^{\infty} \delta\Phi \frac{\partial \delta\Phi^*}{\partial r} dz. \quad (4.17)$$

Thus for a two dimensional razor thin disc the conservation law is of the same form as Eq. 4.9 but with ρ_J , \mathbf{F}_A , and \mathbf{F}_{ext} given by Eq. 4.10, Eq. 4.11 and Eq. 4.13 with ρ replaced by Σ respectively. As the vertical direction has been integrated over only the radial components of the fluxes contribute. From the above analysis the vertically integrated angular momentum flux due to gravitational stresses becomes

$$\mathbf{F}_G \rightarrow F_{Gr} \hat{\mathbf{r}}, \quad (4.18)$$

with F_{Gr} given by Eq. 4.17. Accordingly this term, unlike the others involves an integration over the vertical direction. Explicit expressions for the fluxes in terms of the perturbation $S = \delta\Sigma c_s^2 / \Sigma + \delta\Phi$, appropriate to the razor thin disc, shall be derived below.

4.3.4 Locally isothermal equation of state

It is possible to repeat the analysis of §4.3.1 when the disc has a locally isothermal equation of state. In this case $p = \rho c_s^2$, where c_s is a prescribed function of position and the linearised equation of motion (Eq. 4.4) is modified to read

$$\frac{D^2 \boldsymbol{\xi}}{Dt^2} + 2\Omega \hat{\mathbf{k}} \wedge \frac{D\boldsymbol{\xi}}{Dt} + r \hat{\mathbf{r}} (\boldsymbol{\xi} \cdot \nabla \Omega^2) = -\nabla S + \frac{\delta\rho}{\rho} \nabla(c_s^2) - \nabla \Phi_{\text{ext}}. \quad (4.19)$$

An expression of the form of Eq. 4.9 may be derived but now the torque density T takes the form

$$T = \frac{m}{2} [\text{Im}(\Phi_{\text{ext}} \delta\rho^*) - \text{Im}(\delta\rho \boldsymbol{\xi}^* \cdot \nabla(c_s^2))]. \quad (4.20)$$

When c_s is constant corresponding to a strictly isothermal equation of state, the fluid is again barotropic and previous expressions recovered. Eq. 4.20 contains terms in addition to external contributions. In general these imply the possibility of an exchange of angular momentum between perturbations and the background. A very similar discussion applies to the razor thin limit. The analysis above can be applied to determine the angular momentum balance for the linear perturbations of razor thin discs and thus

determine properties of and conditions for unstable normal modes.

4.3.5 Properties of the linear modes of a razor thin disc

Linear perturbations are assumed to have ϕ and t dependence through a factor of the form $e^{i(\sigma t + m\phi)}$ where σ is the complex eigenfrequency and m is the azimuthal mode-number. From now on this factor is taken as read. Linearising the hydrodynamic equations for the inviscid case and ignoring external potentials, gives

$$i\bar{\sigma}\delta\Sigma = -\frac{1}{r}\frac{d(\Sigma r\delta u_r)}{dr} - \frac{im\Sigma\delta u_\phi}{r} \quad (4.21)$$

$$i\bar{\sigma}\delta u_r - 2\Omega\delta u_\phi = -\frac{dS}{dr} + \frac{\delta\Sigma}{\Sigma}\frac{dc_s^2}{dr} \quad (4.22)$$

$$i\bar{\sigma}\delta u_\phi + \frac{\kappa^2}{2\Omega}\delta u_r = -\frac{imS}{r}, \quad (4.23)$$

Recall $\bar{\sigma} = \sigma + m\Omega$ is the shifted mode frequency. These equations were also used in Chapter 2 and Chapter 3 in terms of $W \equiv \delta\Sigma/\Sigma$. However, it is convenient for the analytical discussion later on, to derive a governing equation in a slightly different form than previous Chapters. Using Eq. 4.22 and Eq. 4.23 to eliminate δu_r and δu_ϕ in Eq. 4.21 gives

$$\begin{aligned} r\delta\Sigma = & \frac{d}{dr} \left[\frac{r\Sigma}{D} \left(\frac{dS}{dr} + \frac{2m\Omega\bar{\sigma}S}{r\kappa^2} \right) \right] - \frac{2m\Omega\bar{\sigma}\Sigma}{\kappa^2 D} \left(\frac{dS}{dr} + \frac{2m\Omega\bar{\sigma}S}{r\kappa^2} \right) + \frac{mS}{\bar{\sigma}} \frac{d}{dr} \left(\frac{1}{\eta} \right) \\ & - \frac{4m^2\Omega^2\Sigma S}{r\kappa^4} - \frac{d}{dr} \left[\frac{r\delta\Sigma}{D} \frac{dc_s^2}{dr} \right] + \frac{2m\Omega\delta\Sigma}{D\bar{\sigma}} \frac{dc_s^2}{dr}. \end{aligned} \quad (4.24)$$

Recall $D \equiv \kappa^2 - \bar{\sigma}^2$ and $\eta = \kappa^2/2\Omega\Sigma$. This form is similar to that used by Papaloizou & Savonije (1991) and explicitly shows that a co-rotation singularity, where $\bar{\sigma} = 0$, can be avoided if co-rotation corresponds to a vortensity extremum ($d\eta/dr = 0$). The potential singularity at co-rotation associated with the term involving dc_s^2/dr has negligible effect in practice, explained below.

The above expressions may be applied to a disc, either with a locally isothermal equation of state, or with a barotropic equation of state, where the integrated pressure $p = p(\Sigma)$ and the sound-speed $c_s^2 \equiv dp/d\Sigma$. To obtain the latter case, the quantity dc_s^2/dr is simply replaced by zero. Inclusion of additional terms dependent on this has been found numerically to produce only a slight modification of the discussion that applies when they are neglected. This is because the locally isothermal c_s^2 varies slowly compared to either the linear perturbations or the surface density in the vicinity of an

4. EDGE MODES IN SELF-GRAVITATING DISCS

edge, thus $(r/c_s^2)dc_s^2/dr$ shall be neglected in the analysis from now on. Also recall that $S = \delta\Sigma c_s^2/\Sigma + \delta\Phi$ and the gravitational potential is given by the vertically integrated Poisson integral as

$$\delta\Phi = S - \frac{\delta\Sigma c_s^2}{\Sigma} = -G \int_{\mathcal{D}} K_m(r, r') \delta\Sigma(r') r' dr'$$

where

$$K_m(r, r') = \int_0^{2\pi} \frac{\cos(m\phi) d\phi}{\sqrt{r^2 + r'^2 - 2rr' \cos(\phi) + \epsilon_g^2}}.$$

Here the domain of integration \mathcal{D} is that of the disc provided there is no external material. However, in a situation where density waves propagate beyond the disc boundaries \mathcal{D} either should be formally extended or the form of K_m changed to properly reflect the boundary conditions.

The governing equation above, together with angular momentum conservation, can be used to model angular momentum flux balance for normal modes. For simplicity only barotropic case are considered in the analysis. The analytic discussion concerns a disc model with one boundary being a sharp edge with arbitrary propagation of density waves directed away from this boundary being allowed. When applied to a gap opened by a planet in a global disc, the analytic models corresponds to the section of the disc from the inner disc boundary to the inner gap edge or the section from the outer gap edge to the outer disc boundary. These two sections are assumed to be decoupled in the analytic discussion of stability but not in the numerical one. In addition, although it produces the background profile, the planet is assumed to have no effect on the linear stability analysis discussed here, but it is fully incorporated in nonlinear simulations. The correspondence between the various approaches adopted indicates the above approximations are reasonable. The discussion of the outer and inner disc sections is very similar, accordingly these are considered below together.

4.3.6 Angular momentum flux balance for normal modes

The vertically integrated angular momentum flux associated with perturbations can be related to the background vortensity profile. A barotropic disc model is assumed, so

4.3 Analytical discussion of edge modes

terms involving dc_s/dr are neglected. Multiplying Eq. 4.24 by S^* and integrating:

$$\begin{aligned} \mathcal{F}(S, \sigma) \equiv & - \int_{r_i}^{r_o} r \delta \Sigma S^* dr \pm \left[\frac{r \Sigma}{D} \left(\frac{dS}{dr} + \frac{2m\Omega \bar{\sigma}}{r \kappa^2} S \right) S^* \right] \Big|_{r_b} \\ & - \int_{r_i}^{r_o} \frac{r \Sigma}{D} \left(\frac{dS}{dr} + \frac{2m\Omega \bar{\sigma}}{r \kappa^2} S \right) \left(\frac{dS^*}{dr} + \frac{2m\Omega \bar{\sigma}}{r \kappa^2} S^* \right) dr \\ & - \int_{r_i}^{r_o} \frac{4m^2 \Omega^2 \Sigma |S|^2}{r \kappa^4} dr + \int_{r_i}^{r_o} \frac{m |S|^2}{\bar{\sigma}} \frac{d}{dr} \left(\frac{1}{\eta} \right) dr = 0, \end{aligned} \quad (4.25)$$

where, here and below, the positive sign alternative applies to the case where the disc domain has an inner sharp edge where $r = r_i$ and then $r_b = r_o$, the outer boundary radius. The negative sign alternative applies to the case where the disc domain has an outer sharp edge where $r = r_o$ and then $r_b = r_i$, the inner boundary radius. In both cases waves may propagate through $r = r_b$. Note also that there is no contribution from edge boundary terms as the surface density is assumed to be negligible there.

One may now express the terms in the above integral relation, which for convenience has been taken to define a functional of S that depends on σ , in terms of quantities related to the transport of angular momentum by taking its imaginary part. One begins by assuming σ is real or, more precisely, the limit that marginal stability is approached.

By making use of Eq. 4.16, one finds that

$$\begin{aligned} \text{Im} \left[\int_{r_i}^{r_o} r \delta \Sigma S^* dr \right] &= \text{Im} \left[\int_{r_i}^{r_o} r \delta \Sigma \left(\frac{\delta \Sigma^* c_s^2}{\Sigma} + \delta \Phi^* \right) dr \right] = \text{Im} \left[\int_{r_i}^{r_o} r \delta \Sigma \delta \Phi^* dr \right] \\ &= \pm \frac{2}{m} [r F_{Gr}] \Big|_{r_b}. \end{aligned} \quad (4.26)$$

Recall that F_{Gr} is the vertically integrated angular momentum flux transported by gravitational stresses. Note that F_{Gr} at the sharp edge is ignored. Eq. 4.16, together with the requirement that F_{Gr} is regular for $r \rightarrow 0$ and vanishes more rapidly than $1/r$ for $r \rightarrow \infty$, implies that F_{Gr} is zero at the sharp edge separating the disc domain and the exterior (assumed) vacuum or very low density region.

In addition, from Eq. 4.22 and Eq. 4.23, the radial Lagrangian displacement is given by

$$\xi_r = \frac{\delta u_r}{i \bar{\sigma}} = -\frac{1}{D} \left(\frac{dS}{dr} + \frac{2mS\Omega}{r \bar{\sigma}} \right). \quad (4.27)$$

From this it follows that the imaginary part of the second term on the RHS of Eq.

4. EDGE MODES IN SELF-GRAVITATING DISCS

4.25) is

$$\pm \operatorname{Im} \left[\frac{r\Sigma}{D} \left(\frac{dS}{dr} + \frac{2m\Omega\bar{\sigma}}{r\kappa^2} S \right) S^* \right] \Big|_{r_b} = - \pm \operatorname{Im} [r\Sigma\xi_r S^*] \Big|_{r_b} = - \pm (2/m)[rF_{Ar}] \Big|_{r_b}.$$

Using the above and taking the imaginary part of Eq. 4.25 yields the remarkably simple expression

$$\pm [r(F_{Gr} + F_{Ar})] \Big|_{r_b} = \frac{m}{2} \operatorname{Im} \left[\int_{r_i}^{r_o} \frac{m|S|^2}{\bar{\sigma}} \frac{d}{dr} \left(\frac{1}{\eta} \right) dr \right]. \quad (4.28)$$

In the above, it has been assumed that σ is real and that the mode is at marginal stability. Then the RHS of Eq. 4.28 is apparently real. However, the integrand is potentially singular at corotation where $\bar{\sigma} = 0$. Thus the approach to marginal stability has to be taken with care.

Setting $\sigma = \sigma_R + i\gamma$, where σ_R , being the real part of σ , defines the corotation radius, r_c , through $\Omega(r_c) = -\sigma_R/m$ and γ is the imaginary part of σ , so that the mode is unstable for $\gamma < 0$ with growth rate $-\gamma$. Marginal stability can be approached by assuming that γ has a vanishingly small magnitude but is negative. Then one can apply the Landau prescription and replace $1/\bar{\sigma}$ by $\mathcal{P}(1/\bar{\sigma}) + i\pi\delta(\bar{\sigma})$, where \mathcal{P} indicates that the principal value of the integral is to be taken and δ denotes Dirac's delta function. Adopting this, Eq. 4.28 becomes

$$[r(F_{Gr} + F_{Ar})] \Big|_{r_b} = \pm \frac{\pi|m|}{2} \left[\frac{|S|^2}{|d\Omega/dr|} \frac{d}{dr} \left(\frac{1}{\eta} \right) \right] \Big|_{r_c}. \quad (4.29)$$

The above relation can be viewed as stating that at marginal stability, either corotation is at a vortensity extremum and both terms in Eq. 4.29 vanish, or angular momentum losses as a result of waves passing through $r = r_b$ (the smooth boundary), are balanced by torques exerted at corotation. The latter torque is proportional to the gradient of the vortensity η (Goldreich & Tremaine, 1979). Note that η^{-1} scales with Σ .

The propagating waves quite generally carry negative angular momentum when located in an inner disc ($r < r_c$) and positive angular momentum when located in an outer disc ($r > r_c$) (Goldreich & Tremaine, 1979). Thus a balance may be possible between angular momentum losses resulting from the propagation of these waves out of the system and angular momentum gained or lost by an edge disturbance (near co-rotation radius) as expressed by Eq. 4.29. For an edge disturbance in a region of increasing (decreasing) surface density, such as an inner (outer) edge to an outer (inner) disc, the edge disturbance loses (gains) angular momentum.

4.4 Description of the edge mode instability

In practice, an inner (outer) edge disturbance may be associated with a positive (negative) surface density slope near a vortensity maximum or near a vortensity minimum occurring as a result of the variation of both κ^2 and Σ (e.g. Lovelace et al., 1999; Papaloizou & Lin, 1989). The former case is associated with discs for which self-gravity is important and spiral density waves are readily excited. The latter is associated with weakly or non self-gravitating discs and is associated with vortex formation at gap edges as has already been discussed in Chapters 2–3.

The boundary condition at $r = r_b$ (the non-sharp boundary) may correspond to stipulating outgoing waves or a radiation condition, so that the fluxes in Eq. 4.29 are non zero. Another possibility is to have a surface density taper to zero which would mean wave reflection and removal of the boundary flux terms. In fact, the issue of stability, for the particular modes of interest, is not sensitive to precise boundary conditions at the smooth boundary. In the case of non self-gravitating discs and the low m vortex modes, the regions away from the edge are evanescent and amplitudes are small there, the disturbance being localised in the vicinity of the edge. For self-gravitating discs, as indicated below, instability can appear because of an unstable interaction between *outwardly propagating positive* (*inwardly propagating negative*) angular momentum density waves and a *negative* (*positive*) angular momentum edge disturbance localised near an *inner* (*outer*) disc gap edge. As long as these waves are excited it should not matter whether they are reflected or transmitted at large distance, as long as their angular momentum density remains in the wider disc and is not fed back to the exciting edge disturbance.

For these reasons and also simplicity it shall be assumed that at the boundary where $r = r_b$, either the surface density has tapered to zero, or there is reflection such that the boundary terms in expressions like Eq. 4.25 may be dropped. Eq. 4.29 then simply states that marginal stability occurs when corotation is at a vortensity extremum. It is found that the vortensity *maximum* case is associated with an instability in self-gravitating discs that is associated with strong spiral waves. On the other hand the vortensity *minimum* case is associated with the vortex-forming instability in non-self-gravitating discs that has been previously studied.

4.4 Description of the edge mode instability

The edge mode instability in a self-gravitating disc can be interpreted as due to a disturbance associated with the edge causing the excitation of spiral waves that propagate away from it resulting in destabilisation. This occurs because the emitted waves carry

4. EDGE MODES IN SELF-GRAVITATING DISCS

away angular momentum which has the opposite sign to that associated with the edge disturbance (see §4.3.6 and Eq. 4.29). Accordingly the excitation process is expected to lead to the growth of this disturbance.

Thus one can consider a disturbance localised in the vicinity of the edge and the potential perturbation it produces to excite spiral density waves in the bulk of the disc. Numerical calculations show that edge dominated modes of this type occur for low m and are dominant in the nonlinear regime. The analysis below assumes the mode is weakly growing corresponding to the back reaction of the waves on the edge disturbance being weak. Thus in the first instance one calculates a neutral edge disturbance (zero growth rate) and then calculate the wave emission as a perturbation. It should be emphasised again that an important feature is that corotation is at a vortensity *maximum* (in contrast to the non-self-gravitating case for which corotation is located at a vortensity *minimum*). Such modes require self-gravity. In an average sense they require a sufficiently small Q value and so cannot occur in the non self-gravitating limit ($Q \rightarrow \infty$).

4.4.1 Neutral edge disturbances with corotation at a vortensity maximum

Consider the expression for $\delta\Sigma$ expressed by Eq. 4.24 and assume that only the vortensity gradient term (third term on the RHS) need be retained. This is because this term should dominate near corotation in the presence of large vortensity gradients that are presumed to occur near the edge and the disturbance is assumed localised there ¹. Thus

$$\delta\Sigma = \frac{S}{r\bar{\omega}} \frac{d}{dr} \left(\frac{1}{\eta} \right), \quad (4.30)$$

where $\bar{\omega} = \bar{\sigma}/m$. The associated gravitational potential is given by

$$\delta\Phi = -G \int_{r_i}^{r_o} K_m(r, r') \frac{S(r')}{\bar{\omega}(r')} \frac{d}{dr'} \left[\frac{1}{\eta(r')} \right] dr'. \quad (4.31)$$

The relation $S = \delta\Sigma c_s^2 / \Sigma + \delta\Phi$ gives the integral equation

$$S(r) \left[1 - \frac{c_s^2}{r\Sigma\bar{\omega}} \frac{d}{dr} \left(\frac{1}{\eta} \right) \right] = - \int_{r_i}^{r_o} GK_m(r, r') \frac{S(r')}{\bar{\omega}(r')} \frac{d}{dr'} \left[\frac{1}{\eta(r')} \right] dr'. \quad (4.32)$$

¹See also the discussion of groove modes in collisionless particle discs of Sellwood & Kahn (1991), where similar assumptions are made.

4.4 Description of the edge mode instability

For a disturbance dominated by self-gravity, pressure should be negligible. Under such an approximation, $S \sim \delta\Phi$. Eq. 4.30 then implies $\text{sgn}(\bar{\omega}) = \text{sgn}(d\eta/dr)$ in order to obtain a negative (positive) potential perturbation for a positive (negative) surface density perturbation. At co-rotation where $\bar{\omega} = 0$ and there is a vortensity extremum, this requirement becomes $\text{sgn}(d\bar{\omega}/dr) = \text{sgn}(d\Omega/dr) = \text{sgn}(d^2\eta/dr^2)$. Since typical rotation profiles have $d\Omega/dr \equiv \Omega' < 0$, the physical requirement that potential and surface density perturbations have opposite signs implies that $d^2\eta/dr^2 < 0$ at co-rotation, i.e. the vortensity is a maximum. The discussion thus applies to vortensity maxima only. For the class of vortensity profiles for which $\text{sgn}(d\eta/dr) = \text{sgn}(\bar{\omega})$, the integral Eq. 4.32 may be transformed into a Fredholm integral equation with symmetric kernel. This is demonstrated below.

Noting that corotation is located at a vortensity maximum, the requirement above implies the vortensity increases (decreases) interior (exterior) to corotation¹. As the disturbance is expected to be localised around corotation, it is reasonable to assume this holds and that one may contract the integration domain (r_i, r_o) to exclude regions which do not conform without significantly affecting the problem.

Next, introduce the function \mathcal{H} such that

$$S(r) = \mathcal{Z}(r)\mathcal{H}(r), \quad (4.33)$$

where

$$\mathcal{Z}(r) = \left[-\frac{d}{dr} \left(\frac{1}{\eta} \right) \frac{1}{\bar{\omega}} \right]^{-1/2} \left[1 - \frac{c_s^2}{r\Sigma\bar{\omega}} \frac{d}{dr} \left(\frac{1}{\eta} \right) \right]^{-1/2}. \quad (4.34)$$

Note that for $\text{sgn}(d\eta/dr) = \text{sgn}(\bar{\omega})$, \mathcal{Z} is real. Defining the new symmetric kernel $\mathcal{R}(r, r')$ as

$$\mathcal{R}(r, r') \equiv GK_m(r, r')\mathcal{Y}(r)\mathcal{Y}(r') \quad (4.35)$$

where

$$\mathcal{Y}(r) = \left[-\frac{d}{dr} \left(\frac{1}{\eta} \right) \frac{1}{\bar{\omega}} \right]^{1/2} \left[1 - \frac{c_s^2}{r\Sigma\bar{\omega}} \frac{d}{dr} \left(\frac{1}{\eta} \right) \right]^{-1/2} \quad (4.36)$$

¹The discussion may also be extended to the case when that is true with $d^2\eta/dr^2$ vanishing at corotation

4. EDGE MODES IN SELF-GRAVITATING DISCS

one obtains the integral equation

$$\mathcal{H} = \int_{r_i}^{r_o} \mathcal{R}(r, r') \mathcal{H}(r') dr'. \quad (4.37)$$

In other words, \mathcal{H} is required to be the solution to

$$\lambda \mathcal{H} = \int_{r_i}^{r_o} \mathcal{R}(r, r') \mathcal{H}(r') dr', \quad (4.38)$$

with $\lambda = 1$. However, Eq. 4.38 is just a standard homogeneous Fredholm integral equation of the second kind. Thus the existence of a nontrivial solution ($\mathcal{H} \neq 0$) implies \mathcal{H} is an eigenfunction of the kernel \mathcal{R} with unit eigenvalue.

The kernel \mathcal{R} is positive and symmetric. Accordingly the Fredholm equation above has the property that the maximum eigenvalue $\lambda > 0$ and is the maximum of the quantity Λ given by

$$\Lambda = \frac{\int_{r_i}^{r_o} \int_{r_i}^{r_o} \mathcal{R}(r, r') \mathcal{H}(r') \mathcal{H}^*(r) dr dr'}{\int_{r_i}^{r_o} |\mathcal{H}(r)|^2 dr} \quad (4.39)$$

over all \mathcal{H} . Thus, if it is possible to show that $\max(\Lambda) < 1$, then it is not possible to have a non-trivial solution corresponding to a neutral mode. Upon application of the Cauchy-Schwarz inequality (twice), one can deduce

$$\Lambda^2 \leq \int_{r_i}^{r_o} \int_{r_i}^{r_o} |\mathcal{R}(r, r')|^2 dr dr' = \Lambda_{est}^2 \geq \max(\Lambda^2). \quad (4.40)$$

Thus the existence of a neutral mode of this type is not possible if Λ_{est}^2 is less than unity.

To relate the necessary condition for mode existence to physical quantities, one can estimate Λ_{est}^2 . The Poisson kernel $K_m(r, r')$ is largest at $r = r'$. Other factors in the numerator of \mathcal{R} involve the factor $1/\bar{\omega}$. Assume these factors have their largest contribution at corotation where $r = r_c$ and $\bar{\omega} = 0$. Then $\Lambda_{est}^2 \sim \int_{r_i}^{r_o} \int_{r_i}^{r_o} |\mathcal{R}(r_c, r_c)|^2 dr dr'$. Assuming the edge region has width of order L_c , taken to be much less than the local radius but not less than the local scale-height, gives the estimate

$$\Lambda_{est} \sim |\mathcal{R}(r_c, r_c)| L_c = GK_m(r_c, r_c) \left| \frac{1}{\Omega'} \left(\frac{1}{\eta} \right)'' \right| L_c \left/ \left| 1 - \frac{c_s^2}{r_c \Sigma \Omega'} \left(\frac{1}{\eta} \right)'' \right| \right|. \quad (4.41)$$

All quantities are evaluated at co-rotation and the double prime denotes d^2/dr^2 . Because the edge is thin by assumption, the Poisson kernel can be further approximated

4.4 Description of the edge mode instability

by

$$K_m(r, r') \sim \frac{2}{\sqrt{rr'}} K_0 \left(\frac{m}{\sqrt{rr'}} \sqrt{|r - r'|^2 + \epsilon_g^2} \right), \quad (4.42)$$

where K_0 is the modified Bessel function of the second kind of order zero and recall that ϵ_g is the softening length. If pressure effects are negligible (c_s^2 being small) then

$$\Lambda_{\text{est}} \sim \frac{2GK_0(m\epsilon_g/r_c)}{r_c} \left| \frac{1}{\Omega'} \left(\frac{1}{\eta} \right)'' \right|_{r_c} L_c. \quad (4.43)$$

Since $\eta = \kappa^2/2\Omega\Sigma$, the requirement that Λ_{est}^2 exceeds unity, leads to the expectation that modes of the type under discussion can exist only for sufficiently large surface density scales. Setting $\Omega' \sim \Omega/r_c$ and suppose the vortensity varies on a length scale L_c implies

$$\Lambda_{\text{est}} \sim \frac{4G\Sigma K_0(m\epsilon_g/r_c)}{\Omega^2 L_c}. \quad (4.44)$$

Note $\kappa \sim \Omega$ has been used to obtain the final expression. Thus taking $L_c \sim H$ and ϵ_g being a fraction ϵ_{g0} of the local scale height, gives $\Lambda_{\text{est}} \sim 4 \ln(1/(m\epsilon_{g0}h))/(\pi Q)$. On account of the logarithmic factor, the condition $\Lambda_{\text{est}} > 1$ suggests that edge disturbances which lead to instabilities can be present for Q significantly larger than unity, as is confirmed numerically. However, a surface density threshold must be exceeded. Thus such modes associated with vortensity *maxima* do not occur in a non self-gravitating disc. The fundamental quantity is the vortensity profile, Eq. 4.43 indicates that if it is not sufficiently peaked, there can be no mode. Similarly, mode existence becomes less favoured for increasing softening and/or increasing m .

It is important to note an analysis of the type discussed above does not work for vortensity minima. An equation similar to Eq. 4.38 could be derived but in this case the corresponding kernel \mathcal{R} would be negative, implying inconsistent negative eigenvalues λ . This situation results from the surface density perturbation leading to a gravitational potential perturbation with the wrong sign to satisfy the condition Eq. 4.30. Thus the type of gravitationally-dominated edge disturbance considered in this section is not expected to be associated with vortensity minima. The fact that edge modes associated with vortensity maxima require a threshold value of Q^{-1} to be exceeded separates them from vortex forming modes which are modes associated with vortensity minima and require Q^{-1} not to be above a threshold in order to be effective. For most disc models considered in this Chapter, with minimum $Q \leq 2$, vortex modes do not occur. Discs

4. EDGE MODES IN SELF-GRAVITATING DISCS

with minimum $Q \gtrsim 2$ were considered in Chapter 3.

4.4.1.1 Implications for disc-planet systems

For the locally isothermal disc models adopted in this Chapter, with a small aspect-ratio $h = 0.05$, edge modes could appear more easily as the disc is cold. This is because if the disc thickness sets the length scale of the edge, Eq. 4.41 and Eq. 4.43—4.44 indicate that, for fixed Q , lowering sound-speed and hence the disc aspect ratio increases Λ_{est} .

Equations above can be applied to gaps opened by a Saturn mass planet, which is considered in linear and nonlinear numerical calculations later on. Without instabilities, these gaps deepen with time. There is also vortensity mixing in the co-orbital region. In a fixed orbit, these two effects respectively reduce the gap surface density and the magnitude of the edge vortensity peaks. In this way the conditions for edge modes become less favourable with time. In this case edge modes are expected to develop early on during gap formation, if at all. However, this effect may be less pronounced if the orbit evolves because the planet migrates.

For larger planetary masses such as Jupiter, a deeper gap will be opened but stronger shocks are also induced, which may lead to stronger vortensity peaks. Thus, in view of potentially competing effects, the conditions for edge modes as a function of planet migration and planetary mass must be investigated numerically.

4.4.2 Launching of spiral density waves

Although localised at the gap edge, the edge disturbance perturbs the bulk of the disc through its gravitational potential exciting density waves. This is expected to be through torques exerted at Lindblad resonances (Goldreich & Tremaine, 1979). When the disc is exterior (interior) to the edge the wave excitation can be viewed as occurring at the the outer (inner) Lindblad resonance respectively. These resonances occur where $\sigma/m = -\Omega \mp \kappa/m$. Here the negative (positive) sign alternative applies to the outer (inner) Lindblad resonance respectively. The perturbing potential is given by Eq. 4.31 and below this potential is given the symbol Φ_{edge} .

The total conserved angular momentum flux associated with the launched waves, when they are assumed to propagate out of the system is given by Goldreich & Tremaine (1979) as

$$F_T = \frac{\pi^2 m r \Sigma}{\beta} \left| \frac{d\Phi_{\text{edge}}}{dr} + \frac{2m\Omega\bar{\sigma}\Phi_{\text{edge}}}{r\kappa^2} \right|^2. \quad (4.45)$$

4.4 Description of the edge mode instability

where $\beta = 2\kappa(\mp\kappa' - m\Omega')$. These waves carry positive angular momentum outwards when excited by an inner edge, or equivalently negative angular momentum inwards when excited by an outer edge. Thus they will destabilise negative angular momentum disturbances at an inner disc edge or positive angular momentum disturbances at an outer disc edge that cause their emission. They will lead to an angular momentum flux at the boundary where $r = r_b$ given by $(2\pi)[r(F_{Gr} + F_{Ar})]|_{r_b} = F_T$

4.4.3 Spiral density waves and the growth of edge modes

For weakly growing modes, the effects of wave losses at the non-sharp boundary can be considered as a perturbation on the edge disturbance. To do this, the assumption that the surface density tapers to zero at the boundary where $r = r_b$ is relaxed. Instead, a small value is adopted there such that self-gravity is not important for the density waves. Thus the domain \mathcal{D} is retained for evaluating the gravitational potential ($r \in [r_i, r_o]$). Then, for real frequencies, all the terms in Eq. 4.25 apart from the term inversely proportional to $\bar{\sigma}$ and the boundary term associated with the advective angular momentum flux at $r = r_b$ are real. Assume that the mode is close to marginal stability (subscript ‘ms’ below) with corotation at a vortensity maximum where $\sigma = \sigma_r$ and write Eq. 4.25 in the form $\mathcal{F}(S, \sigma) = \mathcal{F}(S, \sigma_r + \delta\sigma) = 0$. For small $\delta\sigma$, one can expand to first order in $\delta\sigma$, obtaining

$$\left. \frac{\partial \mathcal{F}}{\partial \sigma} \right|_{ms} \delta\sigma + \mathcal{F}(S, \sigma_r) \equiv (D_r + iD_i)\delta\sigma + \mathcal{F}(S, \sigma_r) = 0. \quad (4.46)$$

Differentiating Eq. 4.25 gives

$$D_r = \left(\mathcal{P} \int_{r_i}^{r_o} \frac{m|S|^2}{\eta^2 \bar{\sigma}^2} \frac{d\eta}{dr} dr \right) \Big|_{\sigma=\sigma_r} - \frac{\partial}{\partial \sigma} \left(\int_{r_i}^{r_o} \frac{r\Sigma}{D} \left| \frac{dS}{dr} + \frac{2m\Omega\bar{\sigma}}{rk^2} S \right|^2 dr \right) \Big|_{\sigma=\sigma_r} \quad (4.47)$$

and

$$D_i = -\frac{\pi|S|^2}{m\Omega'|\Omega'|} \frac{d^2}{dr^2} \left(\frac{1}{\eta} \right) \Big|_{ms}. \quad (4.48)$$

Setting $\delta\sigma = \delta\sigma_r + i\gamma$, where $-\gamma$ is the growth rate and taking the imaginary part of Eq. 4.46, noting that the imaginary part of the first two terms on the RHS of Eq. 4.25 contribute to the imaginary part of \mathcal{F} giving this as $\text{Im}(\mathcal{F}) = -\pm(2/m)[r(F_{Gr} + F_{Ar})]|_{r_b} =$

4. EDGE MODES IN SELF-GRAVITATING DISCS

$\mp(1/(m\pi))F_T$ with F_T given above,

$$-\gamma = \mp \frac{F_T}{m\pi D_r} + \delta\sigma_r \frac{D_i}{D_r}. \quad (4.49)$$

Similarly the real part of (4.46) gives

$$D_r \delta\sigma_r = \gamma D_i - \text{Re}(\mathcal{F}), \quad (4.50)$$

where $\text{Re}(\mathcal{F})$ is the real part of \mathcal{F} . Suppose that the waves emitted by the edge disturbance result in $\gamma \neq 0$, but that the back reaction does not change the location of the co-rotation point from that of the marginally stable mode (as indicated by numerical results). Thus co-rotation remains at the original vortensity maximum, implying that conditions adjust so that $\delta\sigma_r = 0$. Then:

$$-\gamma = \mp \frac{F_T}{m\pi D_r}, \quad (4.51)$$

where the upper (lower) sign applies to an inner (outer) edge. The inner edge case is the one considered in linear and nonlinear calculations. Instability occurs when wave emission causes negative (positive) angular momentum to be transferred to a negative (positive) angular momentum edge disturbance located at an inner (outer) sharp edge.

According to Eq. 4.51, to obtain instability ($-\gamma > 0$) for an inner edge disturbance one requires $D_r < 0$ (since $F_T > 0$). For a disturbance concentrated at corotation near an inner sharp edge, with the disc lying beyond, the first term on the RHS of Eq. 4.47 dominates. As corotation is at a vortensity maximum, one expects the contribution to the first integral of D_r from the region just beyond corotation, where $d\eta/dr < 0$, to be negative and the contribution from the region just interior to corotation, where $d\eta/dr > 0$, to be positive. Because the region exterior to an inner edge has higher surface density than that interior to it, and the integrand for the first term in D_r is proportional to Σ , one may expect the region exterior to co-rotation dominates the contribution to D_r , making it negative and therefore unstable. Indeed, it was found that $D_r < 0$ for the numerical fiducial case presented in §4.5.1. In this case there is instability due to the reaction of the emitted outwardly propagating positive angular momentum wave on the negative angular momentum inner gap edge disturbance. A corresponding discussion applies to a disc with an outer sharp edge.

The above considerations depend on the excitation of waves at a Lindblad resonance that were transported with a conserved action or angular momentum flux towards a

boundary where they were lost. However, linear calculations and simulations described below indicate lack of sensitivity to such a boundary condition. This can be understood if it is emphasised that the significant issue is that after emission, the wave action density should not return to the edge disturbance responsible for it. This is possible for example if the waves become trapped in a cavity in the wider disc in which case a radiative boundary would not be needed.

4.5 Linear calculations

Numerical solutions to the linear problem are now presented. These calculations are a continuation of those presented in Chapter 3 to massive disc models where edge modes are the dominant disturbance. The basic state is obtained from hydrodynamic disc-planet simulations by azimuthally averaging the surface density and azimuthal velocity component fields to obtain one dimensional profiles. The background is thus axisymmetric. The radial velocity is assumed to be zero. For consistency with nonlinear simulations, the local isothermal equation of state $p = h^2 GM_* \Sigma / r$ with $h = 0.05$ is adopted, and the self-gravity softening prescription used is that presented in §4.1. This prescription makes the Poisson kernel asymmetric, but this is of no practical significance since the softening length is small (compared to the local radius and characteristic width of the edge region).

Rather than solve in terms of the quantity $S = \delta \Sigma c_s^2 / \Sigma + \delta \Phi$, which was convenient for analytic discussion, it is more convenient here to work in terms of the relative surface density perturbation, $W = \delta \Sigma / \Sigma$ for which a single governing equation may be written down explicitly:

$$\begin{aligned} \frac{d}{dr} \left[\frac{r \Sigma}{D} \left(c_s^2 \frac{dW}{dr} + \frac{d\delta\Phi}{dr} \right) \right] + \left[\frac{2m}{\bar{\sigma}} \left(\frac{\Sigma \Omega}{D} \right) \frac{dc_s^2}{dr} - r \Sigma \right] W \\ + \left[\frac{2m}{\bar{\sigma}} \frac{d}{dr} \left(\frac{\Sigma \Omega}{D} \right) - \frac{m^2 \Sigma}{r D} \right] (c_s^2 W + \delta\Phi) \equiv \mathcal{L}(W) = 0. \end{aligned} \quad (4.52)$$

This is the same linear equation as that used in Chapter 3. Since this is a homogeneous equation, one can multiply through by, e.g. $\bar{\sigma}$, without changing the problem. However, numerical convergence was more easily attained with this additional factor. The term in dc_s^2/dr has been kept for consistency with simulations used to setup the basic state. By solving the linear problem with $dc_s^2/dr = 0$, this term can be verified to have negligible effect on the results obtained below. This is because c_s varies on a global scale, whereas the edge disturbance is associated with strong local gradients.

4. EDGE MODES IN SELF-GRAVITATING DISCS

Eq. 4.52 is converted to a matrix equation $\sum_{j=1}^{N_r} \mathcal{L}_{ij}(\sigma)W_j = 0$ by discretising it on $N_r = 1025$ equally spaced grid points in the domain $r = [1, 10]$. The same iterative method of solution was used as in Chapter 3 (§3.3.3), except that the quadratic approximation used to study localised vortex modes in Chapter 3, which gave trial eigenvalues for the full problem, is not appropriate here since edge modes are intrinsically global.

However, because simulations suggest the important disturbances are associated with the outer gap edge, it is reasonable to search for values of σ such that co-rotation lies near the outer gap edge. The reciprocal of the final matrix condition number was verified to be small (at the level of machine precision) in order for results to be accepted. For simplicity, the boundary condition $dW/dr = 0$ was applied at $r = r_i = 1$ and $r = r_o = 10$. As discussed in §4.3.6, modes are found to be driven by the back reaction of emitted density waves on a disturbance located at an outer gap edge, a process expected to be insensitive to boundary conditions. Indeed, results are insensitive as to whether the above or other boundary conditions are used (see §4.5.5).

4.5.1 Fiducial case

As a fiducial case, consider the disc model with $Q_o = 1.5$ and with a gap opened by a Saturn mass planet. This is adopted in §4.6. However, simulations described in §4.6 from which the model was extracted had a viscosity $\nu = 10^{-5}$, whereas the linear calculations described below are for inviscid discs.

The basic state gap profile is illustrated in Fig. 4.5 where the surface density Σ , the relative deviation of the angular velocity from the Keplerian value $\Omega/\Omega_k - 1$ and the vortensity η are plotted. As expected from analytic discussions of modes near to marginal stability, local extrema in η in the vicinity of gap edges are closely associated with instability. This is manifest through mode co-rotation points being very close to them.

The magnitude of relative surface density perturbation, $|W|$, for self-gravitating (SG) and non-self-gravitating (NSG) responses are shown in Fig. 4.6. NSG means that $\delta\Phi = 0$. However, the background profile is identical for both calculations. The eigenfrequencies for SG (NSG) modes are given by $-\sigma = 0.1587 + i0.4515 \times 10^{-2}$ ($-\sigma = 0.1458 + i0.2041 \times 10^{-2}$). These correspond to co-rotation points at $r_c = 5.4626, 5.7959$ for the SG and NSG modes, respectively. The SG growth rate corresponds to ~ 3 times the local orbital period. Thus although $|\gamma|/\sigma_R \sim 0.03$, is relatively small, the instability grows on a dynamical timescale.

The co-rotation points of SG and NSG modes are very close to a local maximum

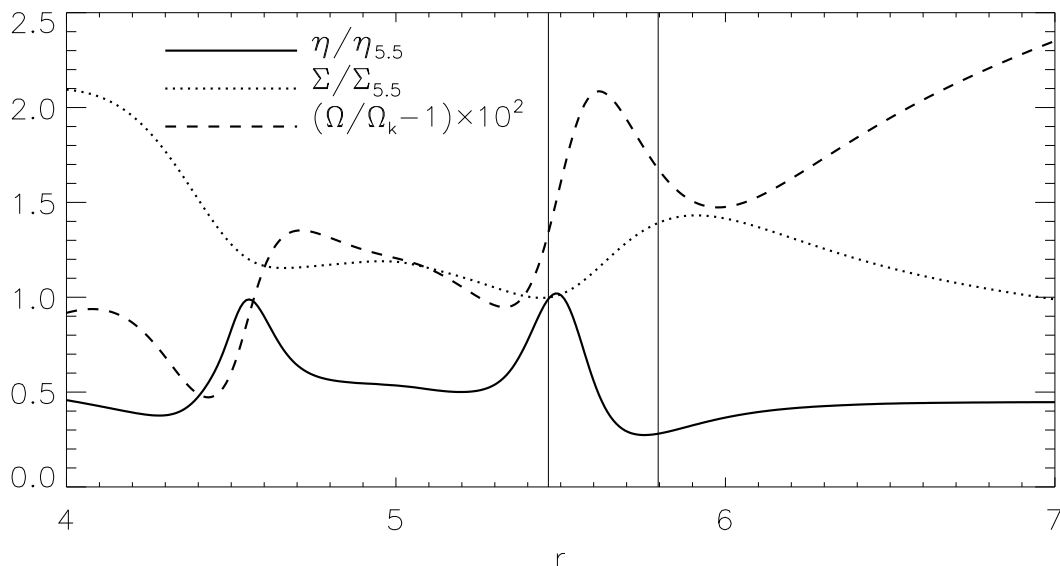


Figure 4.5: Gap profile produced by a Saturn-mass planet in the $Q_o = 1.5$ disc with viscosity $\nu = 10^{-5}$. The surface density Σ and vortensity η are scaled by their values at $r = 5.5$. The relative deviation of the angular velocity from the Keplerian value is also plotted. Vertical lines indicate co-rotation radii r_c (where $\text{Re}(\bar{\sigma}) = 0$) for the $m = 2$ mode. With self-gravity, $r_c = 5.5$, close to a vortensity maximum. When self-gravity is neglected $r_c = 5.8$, close to a vortensity minimum.

and minimum of η , respectively (Fig. 4.5). The SG mode grows twice as fast as the NSG mode, consistent with the observation made when comparing edge modes and vortex modes in §4.2, where the former became non-linear sooner. The SG and NSG eigenfunctions are similar around co-rotation ($r \in [5, 6]$), although the SG mode has a larger width and is shifted slightly to the left (Fig. 4.6). As expected from the discussion in §4.4.2, the SG mode has significant wave-like region interior to the inner Lindblad resonance ($r \leq 3.4$) and exterior to the outer Lindblad resonance at ($r \geq 7.2$). By contrast, the NSG mode has negligible amplitude outside $r \in [5, 7]$ compared to that at co-rotation, whereas the SG amplitude for $r \in [8, 10]$ can be up to $\simeq 56\%$ of the peak amplitude near co-rotation. Increasing m in the NSG calculation increases the amplitude in the wave regions, but even for $m = 6$, it was found that waves in $r \in [8, 10]$ for the NSG case have an amplitude of about 33% of that at co-rotation, a smaller value than that pertaining to the SG case with $m = 2$.

The behaviour in the wider disc, away from co-rotation, shows that for low m , the NSG mode is a localised vortex mode whereas the SG mode corresponds to an edge

4. EDGE MODES IN SELF-GRAVITATING DISCS

mode with global spirals. Noting that maxima in the vortensity and Toomre Q nearly coincide, it makes sense that the SG mode is global because away from co-rotation, the background Toomre Q is decreasing, which makes it easier to excite density waves, because the evanescent zones between corotation and the Lindblad resonances that are expected from WKBJ theory, narrow accordingly.

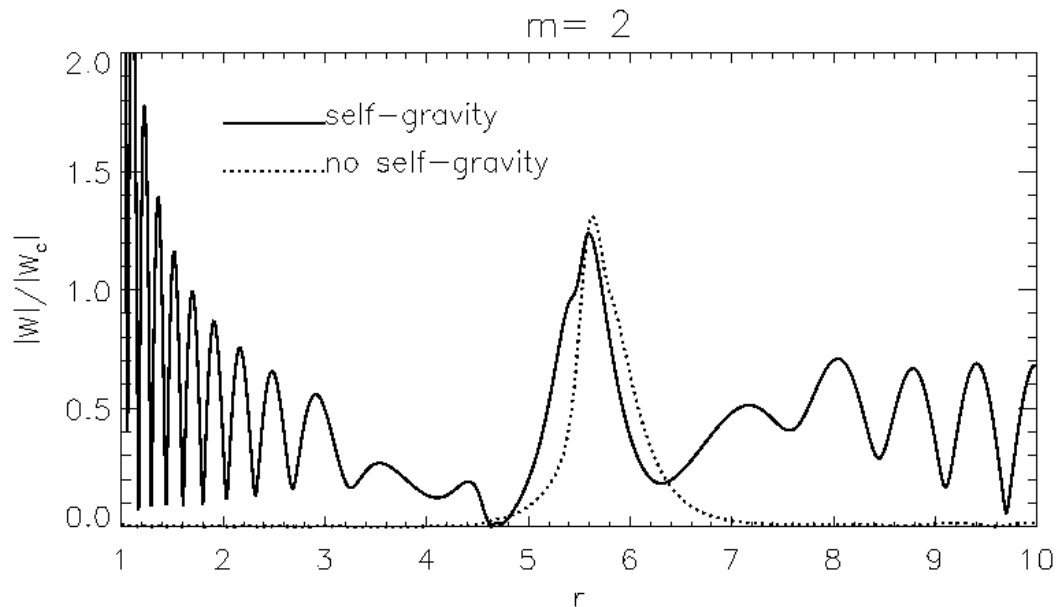


Figure 4.6: *Unstable $m = 2$ modes found from linear stability analysis for the fiducial model, with self-gravity (solid) and without self-gravity (dotted) in the linear response. $|W|$ has been scaled by its value at co-rotation.*

Comparing the SG and NSG modes in Fig. 4.6 shows that including self-gravity in the linear response enables additional waves in the disc at low m . Fig. 4.7 shows the gravitational potential perturbation for the SG mode. A comparison with $|W|$ in Fig. 4.6 shows that the surface density perturbation around r_c has an associated potential perturbation that varies on a more global scale. The peak in $|W|$ about r_c is confined to $r \in [4.6, 6.2]$ whereas that for $|\delta\Phi|$ is confined to $r \in [3.2, 7.5]$, overlapping Lindblad resonances (see Fig. 4.2) in the latter case. Thus $\delta\Phi$ is less localised around co-rotation than $|W|$. This is qualitatively similar to a planet — a localised mass with globally varying potential.

Rapid oscillations seen in $|W|$ for $r > 7$ are not observed for $|\delta\Phi|$ in the same region. Furthermore, $|\delta\Phi|(r > 7)$ is at most $\simeq 20\%$ of the co-rotation amplitude, which is a smaller ratio than that for $|W|$. This leads to the notion that the disturbance about the

outer gap edge is driving the disturbance in the outer disc as in the analytical discussion given above. In effect, the disturbance at co-rotation acts like an external perturber (e.g. a planet) to drive density waves in the wider disc, through its gravitational field. Clearly, this is only possible in a self-gravitating disc.

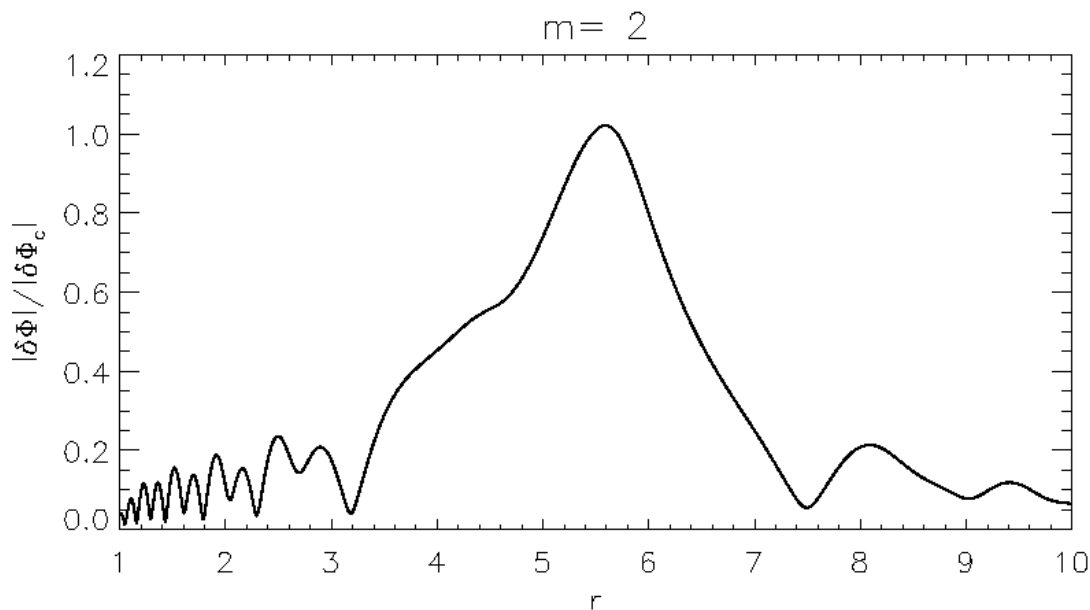


Figure 4.7: Gravitational potential perturbation $\delta\Phi$ ($\equiv \delta\Phi$ in the plot), scaled by its value at $r = 5.5$ for the $m = 2$ mode with self-gravity.

4.5.2 Energy balance of edge and wavelike disturbances

The analysis in §4.4 describes the edge mode as being associated with a coupling between disturbances associated with vortensity extrema near an edge and the smooth regions interior or exterior to the edge. This interpretation can be validated by applying energy balance to the reference case with self-gravity. Specifically, the region $r \in [5, 10]$ is considered because hydrodynamic simulations indicate the spiral arms are more prominent in the outer disc (§4.2 and §4.6). Multiplying the governing Eq. 4.52 by S^* , integrating over $[r_1, r_2]$ and then taking the real part gives

$$\operatorname{Re} \int_{r_1}^{r_2} \varrho dr = \operatorname{Re} \int_{r_1}^{r_2} (\varrho_{\text{corot}} + \varrho_{\text{wave}}) dr, \quad (4.53)$$

4. EDGE MODES IN SELF-GRAVITATING DISCS

where

$$\varrho \equiv r\Sigma WS^* = r(c_s^2|\delta\Sigma|^2/\Sigma + \delta\Sigma\delta\Phi^*), \quad (4.54)$$

$$\varrho_{\text{corot}} \equiv \frac{2m}{\bar{\sigma}} \frac{d}{dr} \left(\frac{\Sigma\Omega}{D} \right) |S|^2, \quad (4.55)$$

$$\varrho_{\text{wave}} \equiv S^* \frac{d}{dr} \left[\frac{r\Sigma}{D} \left(c_s^2 \frac{dW}{dr} + \frac{d\delta\Phi}{dr} \right) \right] \left(\frac{\Sigma\Omega}{D} \right) \frac{dc_s^2}{dr} WS^* - \frac{m^2\Sigma}{rD} |S|^2. \quad (4.56)$$

ϱ has dimensions of energy per unit length and, for a normal mode, its real part is four times the energy per unit length associated with pressure perturbations (the term $\propto c_s^2$) and gravitational potential perturbations (the term $\propto \delta\Phi^*$). As the factor of four is immaterial to the discussion, the integral of ϱ over the region concerned will simply be referred to as the thermal-gravitational energy (TGE). Note that when self-gravity dominates, the TGE is negative and when pressure dominates, it is positive.

When the integral is performed, the TGE is balanced by various terms on the RHS of Eq. 4.53. For simplicity, terms on the RHS are split into just two parts that are integrals of ϱ_{corot} and ϱ_{wave} over the region of interest. This is of course not a unique procedure. The vortensity term ϱ_{corot} has been isolated because it contains the potential co-rotation singularity which can be amplified by large vortensity gradients at the gap edge. The rest of the RHS is collected into ϱ_{wave} . Note that the term in dc_s^2/dr is included in ϱ_{wave} , despite being proportional to $1/\bar{\sigma}$. This is motivated by trying to keep ϱ_{corot} as close to the vortensity term identified in the analytical formalism as possible. However, it has been numerically checked that attributing the dc_s^2/dr term to ϱ_{corot} or neglecting it altogether makes negligible difference to results below. Again, this is because $c_s^2(r) \propto 1/r$ varies on a much larger scale than vortensity gradients.

It is important to note that while the real part of the *combination* $\varrho_{\text{corot}} + \varrho_{\text{wave}}$ gives rise to the TGE, one cannot interpret $\text{Re}(\varrho_{\text{corot}})$ as an energy density of the co-rotation region. It contains a term contributing to the TGE that is proportional to the vortensity gradient (see below) and potentially associated with a corotation singularity. Similar arguments apply to ϱ_{wave} which, in the strictly isothermal case, can be seen to be associated with density waves (see §4.3.6 above). This splitting is used to show that for the modes of interest, the vortensity term contributes most to the TGE.

Numerically however, quantities defined above are inconvenient because of the vanishing of D for neutral modes at Lindblad resonances. To circumvent this, one can work with $D^2\varrho$, $D^2\varrho_{\text{corot}}$, and $D^2\varrho_{\text{wave}}$. These are called modified energy densities. This change does not disrupt the purpose of this section, because the most important balance turns out to be between the terms involving $\text{Re}(D^2\varrho)$ and $\text{Re}(D^2\varrho_{\text{corot}})$, both

focused near co-rotation, where $D \sim \kappa^2$. In the region near Lindblad resonances and beyond, $\text{Re}(\varrho)$ and $\text{Re}(\varrho_{\text{corot}})$ are small. Thus the incorporation of the D^2 factor does not influence conclusions about the TGE balance. The modified energy densities are plotted in Fig. 4.8 for the $m = 2$ mode in the fiducial case. The curves share essential features with those obtained using the original definitions without the additional factor of D^2 (these are presented and discussed in Appendix F).

Fig. 4.8 shows that $\text{Re}(D^2\varrho)$ is negative around co-rotation which is located near the outer gap edge, $r \in [5, 6]$. Accordingly $\text{Re}(\varrho) < 0$ here, and must be dominated by the gravitational energy contribution since the pressure contribution is positive definite. This supports the interpretation of the disturbance as an edge mode where self-gravity is important. If it were the vortex mode then $\text{Re}(D^2\varrho) > 0$ near corotation, because in that case, self-gravity is unimportant compared to pressure.

For $r \geq 8.4$, $\text{Re}(D^2\rho)$ becomes positive due to pressure effects and oscillates towards the outer boundary as the perturbation becomes wave-like. Similarly, the pressure perturbation dominates towards the inner boundary (not shown). This signifies that self-gravity becomes unimportant relative to pressure within these regions. This is consistent with the behaviour of the gravitational potential perturbations, which are largest near co-rotation.

Fig. 4.8 shows that $\text{Re}(D^2\rho_{\text{corot}})$ and $\text{Re}(D^2\rho_{\text{wave}})$ have their largest amplitudes for $r \in [5, 6]$. Integrating over $r \in [5, 10]$ gives $\tilde{U} < 0$, where

$$\tilde{U} \equiv \text{Re} \int_5^{10} D^2 \varrho dr.$$

Using a normalisation such that $\tilde{U} = -|\tilde{U}|$ gives

$$\begin{aligned} \tilde{U}_{\text{corot}} &\equiv \text{Re} \int_5^{10} D^2 \varrho_{\text{corot}} dr \simeq -0.822|\tilde{U}|, \\ \tilde{U}_{\text{wave}} &\equiv \text{Re} \int_5^{10} D^2 \varrho_{\text{wave}} dr \simeq -0.179|\tilde{U}|. \end{aligned}$$

This implies the TGE is negative, and hence a gravitationally-dominated disturbance. As $|\tilde{U}_{\text{corot}}/\tilde{U}| \sim 0.8$, this suggests the TGE is predominantly balanced by the vortensity term which from Fig. 4.8, is localised in the gap edge region. It balances the gravitational energy of the mode. Because vortensity gradients are largest near the gap edge, gravitational energy is most negative here overtaking pressure, resulting in a negative TGE.

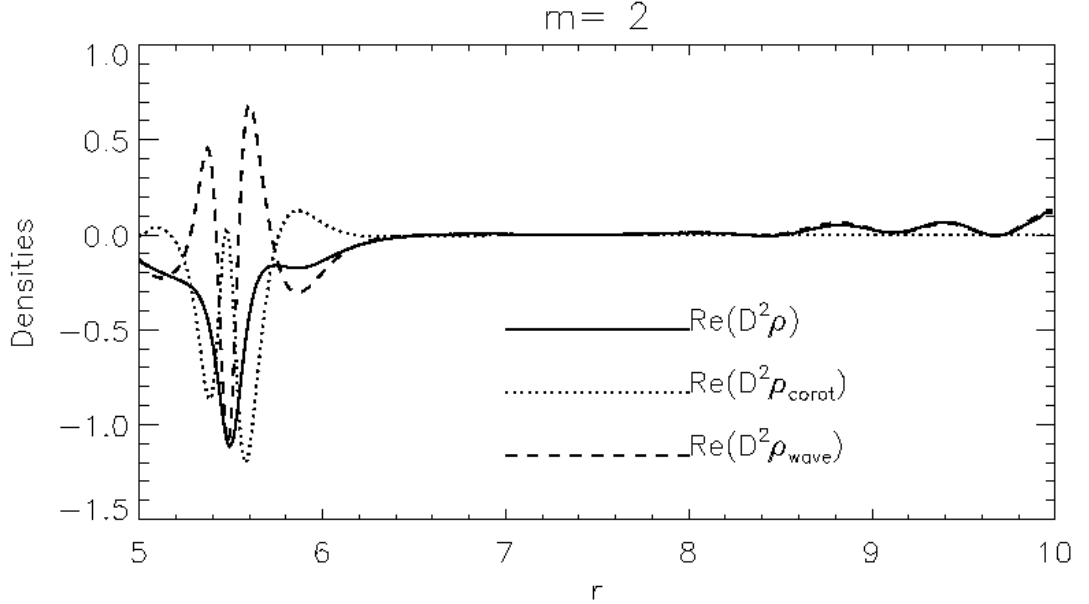


Figure 4.8: One-dimensional modified energy densities computed from the eigenfunctions W , $\delta\Phi$ for the fiducial case $Q_o = 1.5$.

4.5.2.1 Further analysis of the vortensity term

The vortensity term ρ_{corot} defined above differs from that naturally identified in the alternative form of the governing equation in §4.3. However, this difference is not significant. The vortensity term in the linear calculations can be further decomposed as

$$\begin{aligned}\varrho_{\text{corot}} &= \varrho_{\text{corot},1} + \varrho_{\text{corot},2} \\ \varrho_{\text{corot},1} &= \frac{m}{\bar{\sigma}(1-\bar{\nu}^2)} \frac{d}{dr} \left(\frac{1}{\eta} \right) |S|^2, \\ \varrho_{\text{corot},2} &= \frac{2m}{\kappa\eta(1-\bar{\nu}^2)^2} \frac{d\bar{\nu}}{dr} |S|^2,\end{aligned}$$

where $\bar{\nu} = \bar{\sigma}/\kappa$. For the time being, consider $\varrho_{\text{corot},1}$ as the new vortensity term, so that $\varrho_{\text{corot}} \rightarrow \varrho_{\text{corot},1}$ and attribute $\varrho_{\text{corot},2}$ to the wave term so that $\varrho_{\text{wave}} \rightarrow \varrho_{\text{wave}} + \varrho_{\text{corot},2}$. This new vortensity term is proportional to the vortensity gradient explicitly. The new definitions give

$$|\tilde{U}_{\text{corot}}/\tilde{U}| \sim 0.690.$$

Thus the new co-rotation term alone accounts for $\sim 70\%$ of the (modified) TGE and giving almost the same result as the previous one. In addition as most of the contribution comes from around co-rotation, where $|\bar{\nu}| \ll 1$. Replacing the factor $(1 - \bar{\nu}^2)$ by unity has a negligible effect on the results. However, making this replacement gives

$$\rho_{\text{corot},1} \rightarrow \frac{m|S|^2}{\bar{\sigma}} \frac{d}{dr} \left(\frac{1}{\eta} \right),$$

which is exactly the vortensity term used in the analysis given in §4.3. Whether this term contributes positively or negatively to the TGE depends on the sign of $\int_{r_1}^{r_2} \varrho_{\text{corot},1} dr$. Most of the contribution to this integral is expected from the co-rotation region where $\bar{\sigma} \simeq 0$. Then $\text{sgn} \left(\int_{r_1}^{r_2} \varrho_{\text{corot},1} dr \right) = \text{sgn}(\varrho_{\text{corot},1}|_{r_c})$. Evaluating this for a neutral mode with co-rotation at a vortensity extremum gives $\varrho_{\text{corot},1}|_{r_c} = -\eta^{-2}|S|^2(d^2\eta/dr^2)/\Omega'|_{r_c}$. Since quite generally, $\Omega' \equiv d\Omega/dr < 0$, if co-rotation occurs at $\max(\eta)$ then the vortensity term contributes negatively to the TGE. As stated above, when self-gravity dominates pressure, the TGE is negative, so it can then be mostly accounted for by the vortensity term. This is precisely the type of balance assumed for the neutral edge mode modelled in §4.4.1.

However, a consequence of the above discussion is that if co-rotation occurs at a minimum value of η , then the vortensity term contributes positively to the TGE. This can only give the dominant contribution when the TGE is positive, which can only be the case when the pressure contribution dominates over that from self-gravity (see also §3.2.5). In the limit of weak self-gravity, the TGE will be > 0 and it can be balanced by a localised disturbance with co-rotation at a vortensity minimum as happens for the vortex modes.

4.5.3 Dependence on m and disc mass

Solutions to the linear eigenmode problem for discs with $1.2 \leq Q_o \leq 1.6$ are presented below. The basic state is set up from hydrodynamic simulations as described in §4.2, with $\nu = 10^{-5}$. The local $\max(Q)$ near the outer gap edge ranges between $Q = 3.3$ and $Q = 4.4$. Self-gravity is included in the response ($\delta\Phi \neq 0$).

Fig 4.9(a) compares eigenfunctions, $|W|$, for different m for the $Q_o = 1.2$ disc. Increasing m increases the amplitude in the wave region relative to that around co-rotation ($r_c \simeq 5.5$), resulting in $m = 6$ being more global than $m = 3$. High m modes here do not fit the description above of edge modes in the region of interest ($r > 5$), because the disturbance at co-rotation becomes comparable to or even smaller than

4. EDGE MODES IN SELF-GRAVITATING DISCS

that in the wave region beyond the outer Lindblad resonance. It is then questionable to interpret the waves as a secondary phenomenon induced by the edge disturbance, even though the physics of the modes, as being entities driven by an unstable interaction between edge and wave disturbances, may be more or less the same. Hence, the edge-dominated modes described above were only found with $m \lesssim 3$ in simulations where the surface density perturbation is maximal near the gap edge. For fixed $m = 2$, hence in that regime, Fig. 4.9(b) shows that lowering Q_o makes the co-rotation amplitude larger relative to that in the wave region. This can be expected because increasing the level of self-gravity means the necessary condition for edge disturbance is more easily satisfied (§4.4.1).

Fig. 4.10 shows the corotation radii and growth rates for the unstable edge modes as a function of the azimuthal mode number, m and a range of disc masses (parametrised by Q_o). Note that changing Q_o affects both the background state and the linear response, while m acts as a parameter of the linear response only. The plots in Fig. 4.10(a) show that the co-rotation radius tends to move outwards with increasing m and/or Q_o . However, corotation is always located in the edge region where the vortensity is either decreasing or maximal. The largest growth rates are found for low m ($\lesssim 3$) edge-dominated modes.

Referring to the basic state (Fig. 4.5), one sees that increasing m and/or Q_o , which weakens self-gravity, shifts co-rotation towards $\min(\eta)$, which disfavours edge dominated modes. For sufficiently low mass or non-self-gravitating discs, co-rotation is associated with vortensity minima so vortex modes dominate (e.g. $Q_o = 4$ in §4.2). Thus one only expects edge dominated modes for low m and sufficiently large disc mass. An exception to the general trend above is that the $Q_o = 1.6$, $m = 6$ mode, has a smaller co-rotation radius than the corresponding mode with $Q_o < 1.6$. This may be a boundary effect because for this value of m and Q_o the modes are distributed through the outer disc and require the implementation of accurate radiative boundary conditions, rather than the simplified boundary conditions actually applied. However, such high m modes typically have growth rates smaller than those for lower m , accordingly they are not expected to dominate, nor are they observed in the nonlinear simulations discussed later.

Fig.4.10(b) shows that growth rates increase as Q_o is lowered (increasing surface density scale). Note that decreasing Q_o results in deeper gaps due to increased effective planetary mass (see §3.4.3). Steeper gaps are expected to be more unstable, therefore there is a contribution to the increased growth rate from changes to the background profile as the disc mass is increased. This also contributes to the larger disturbance

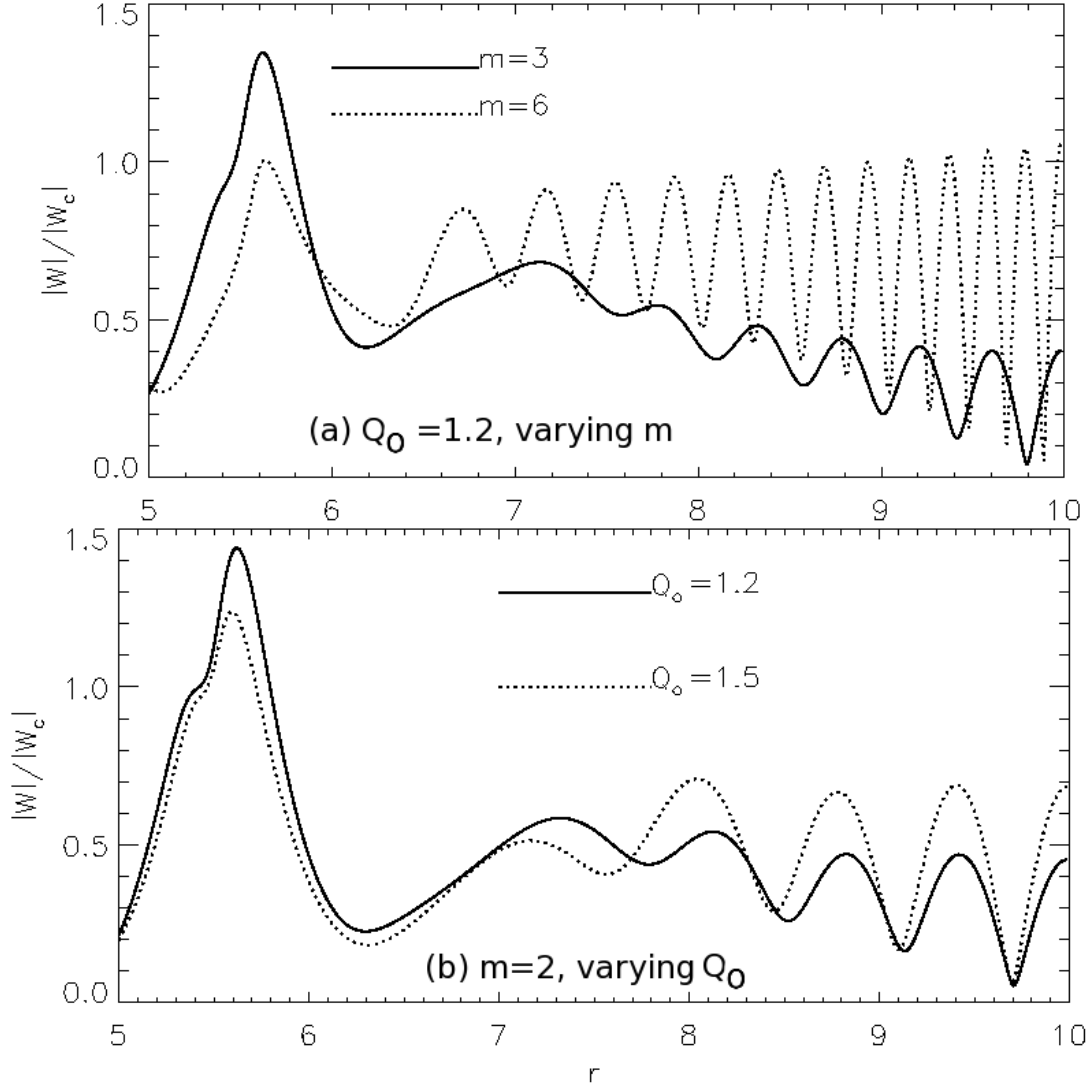


Figure 4.9: The effect of azimuthal wave-number m (a) and the effect of disc mass, parametrised by Q_0 (b) on linear edge modes.

amplitude around co-rotation relative to the wave region as Q_0 is lowered, as observed in Fig. 4.9(b).

One expects that the effect of lowering Q_0 , through the linear response $\delta\Phi$, is to destabilise edge dominated modes as they rely on self-gravity. However, the effect of self-gravity through the response is weaker for larger m values because the size of the Poisson kernel decreases. Hence, the increase of the growth rates with disc mass for

4. EDGE MODES IN SELF-GRAVITATING DISCS

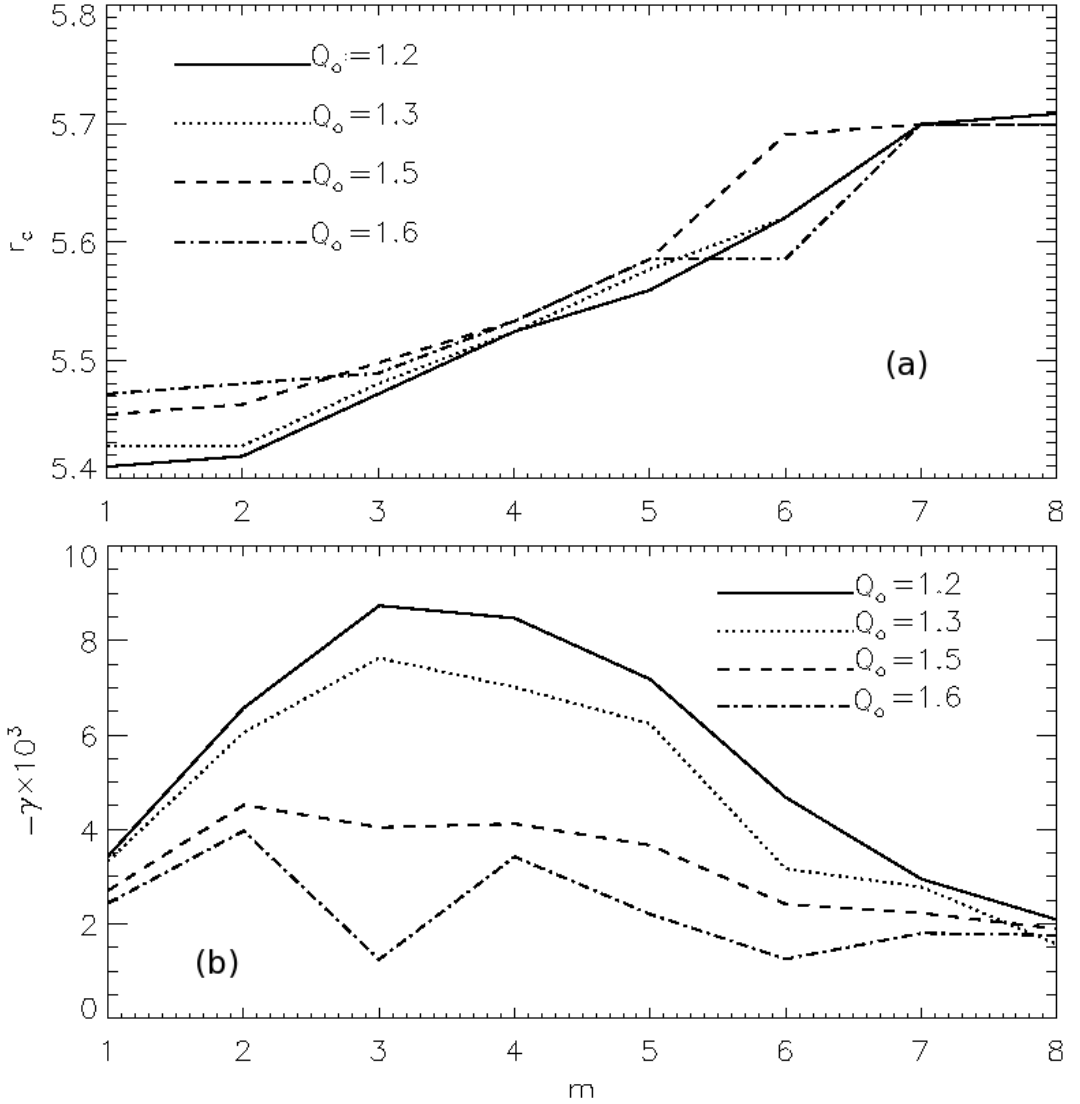


Figure 4.10: Co-rotation radii (a) and growth rates (b) as a function of azimuthal wave-number m , for a range of Q_o values.

$m = 4-6$ is likely more attributable to the effect of self-gravity on the basic state. Growth rates for the most unstable mode approximately doubles as the disc mass is increased by 25% ($Q_o = 1.5 \rightarrow Q_o = 1.2$). Notice for fixed $Q_o \leq 1.5$, the $m = 3-5$ growth rates are similar but $m > 3$ are not edge modes.

For $Q_o = 1.2, 1.3$ the most unstable mode has $m = 3$, whereas for $Q_o = 1.5$, the $m = 2$ mode has the largest growth rate. As discussed in §4.4, edge dominated

modes require sufficient disc mass and/or low m . For fixed m , they are stabilised with increasing Q_o . Hence, for edge modes to exist with decreasing disc mass, they must shift to smaller m . The higher m modes are less affected by self-gravity through $\delta\Phi$. This may explain the double peak in growth rate plotted as a function of m for $Q_o = 1.6$ in moving from $Q_o = 1.5$ (for $Q_o = 1.5$ the $m = 3$ growth rate is also slightly smaller than $m = 2, 4$).

4.5.3.1 Energy partition

The integrals of the modified energy densities for each azimuthal wavenumber are given for the fiducial case with $Q_o = 1.5$ in Table 4.1. Only the modes with $m \leq 2$ are edge dominated modes because for these sum of the integrals contributing to the (modified) TGE is negative, corresponding to a gravitationally dominated disturbance and over 50% of the energy is accounted by the vortensity edge term. The $m = 2$ mode had the largest growth rate with corotation radius at $r = 5.46$. This is fully consistent with the nonlinear simulation of the fiducial case performed in §4.2.1. There the dominant mode in the outer disc had $m = 2$ and corotation radius at $r \simeq 5.5$.

For $m > 2$, the total (modified) energy becomes positive, which must be due to the pressure term ($c_s^2|\delta\Sigma|^2/\Sigma$) becoming dominant. The vortensity term alone cannot balance this because it contributes negatively. Increasing m shifts co-rotation to increased radius, away from the local vortensity maxima, and its contribution becomes smaller in magnitude. However, it does not change sign because r_c never reaches the local vortensity minimum. For $m \geq 4$, the vortensity term contribution is small consistent with high m solutions being predominantly wave-like (Fig. 4.9). However, note that in spite of this the vortensity term may still play some role in driving the modes through corotation torques. These trends have also been found for other models so they appear to be general. However, high m global modes are unimportant as they do not develop in nonlinear simulations.

4.5.4 Softening length

Gravitational softening has to be used in a two-dimensional calculation. It prevents a singularity at $r = r'$ in the Poisson kernel and approximately accounts for the disc's vertical dimension but is associated with some uncertainty. Apart from the linear response, softening also has an effect through the gap profile set up by nonlinear simulations. A fully self-consistent treatment requires a new simulation with each new softening considered.

4. EDGE MODES IN SELF-GRAVITATING DISCS

Table 4.1: *Eigenfrequencies, corotation radii and modified energy integrals for different azimuthal mode numbers, m , for the fiducial disc with $Q_o = 1.5$. The modified energy integrals are taken over the outer disc $r \in [5, 10]$.*

m	$-\sigma_R$	$-\gamma \times 10^2$	r_c	$-\tilde{U}_{\text{corot}}/ \tilde{U} $	$-\tilde{U}_{\text{wave}}/ \tilde{U} $
1	0.079	0.270	5.454	0.65	0.35
2	0.159	0.452	5.463	0.82	0.18
3	0.237	0.404	5.498	0.34	-1.24
4	0.313	0.412	5.533	0.026	-1.03
5	0.386	0.366	5.585	0.0048	-1.004
6	0.462	0.242	5.603	0.0015	-0.998
7	0.524	0.223	5.699	0.0010	-0.997
8	0.599	0.189	5.699	0.00099	-0.995

For reasons of numerical tractability though, experiments were only performed using two values of softening parameter $\epsilon_{g0,e}$ in simulations to setup the gap profile and range of values $\epsilon_{g0,1}$ used in the linear response. In nonlinear simulations a single disc potential softening parameter ϵ_{g0} is used. Results are summarised in Table 4.2. The integrated total modified energy values indicate edge modes since the major contribution is due to the co-rotation/vortensity term.

Comparing growth rates for fully self-consistent cases $\epsilon_{g0,e} = \epsilon_{g0,1} = 0.3, 0.6$ shows that increasing the softening length stabilises edge modes, which is expected since they are driven by self-gravity. In addition all growth rates for $\epsilon_{g0,e} = 0.6$ are smaller than those for $\epsilon_{g0,e} = 0.3$ independently of $\epsilon_{g0,1}$. This indicates stabilisation of edge modes via the basic state when softening is increased (since the effective planet mass decreases with increasing $\epsilon_{g0,e}$, resulting in a shallower profile). For fixed $\epsilon_{g0,e}$, $|\gamma|$ decreases as $\epsilon_{g0,1}$ increases. Eventually, edge modes for $\epsilon_{g0,1} \geq 0.6$ with $\epsilon_{g0,e} = 0.3$ were not found, nor for $\epsilon_{g0,1} \gtrsim 0.83$ when $\epsilon_{g0,e} = 0.6$. An upper limited is expected from analytical considerations (see §4.4.1) because if self-gravity in the linear response is too weak, a self-gravitating vortensity/co-rotation disturbance cannot be maintained. And because the edge mode requires gravitational interaction between gap edge and the smooth disc, if self-gravity is too weak, then this interaction is also weak so edge dominated modes are suppressed for large softening lengths.

4.5.4.1 Convergence issues

For fixed $\epsilon_{g0,e} = 0.3$, the energy ratios in Table 4.2 indicate convergence as $\epsilon_{g0,1} \rightarrow 0$. Perhaps surprisingly, $|\tilde{U}_{\text{corot}}/\tilde{U}|$ decreases with softening. This is because ϱ_{wave} , as defined by equation (4.56), includes a term proportional to $d^2\delta\Phi/dr^2$ which was found

Table 4.2: As for Table 4.1 but for linear calculations with different softening parameters, for the $Q_o = 1.5$ fiducial case and azimuthal mode number $m = 2$. Subscripts ‘l’ denote the softening parameter used in the linear response, and ‘e’ denotes the softening parameter used in setting up the basic state. Note that although growth rates vary by about a factor of two, the location of the corotation radius does not vary significantly.

$\epsilon_{g0,l}$	$\epsilon_{g0,e}$	$-\sigma_R$	$-\gamma \times 10^2$	$-\tilde{U}_{\text{corot}}/ \tilde{U} $	$-\tilde{U}_{\text{wave}}/ \tilde{U} $
0.03	0.3	0.159	0.560	0.66	0.33
0.05	0.3	0.159	0.544	0.67	0.32
0.1	0.3	0.159	0.508	0.69	0.30
0.2	0.3	0.159	0.461	0.74	0.23
0.3	0.3	0.159	0.452	0.82	0.18
0.5	0.3	0.158	0.435	0.87	0.14
0.3	0.6	0.159	0.420	0.91	0.09
0.5	0.6	0.158	0.410	0.95	0.064
0.6	0.6	0.158	0.375	0.99	0.0017
0.8	0.6	0.158	0.276	1.15	-0.14

to be the dominant contribution to $|\tilde{U}_{\text{wave}}|$, and its contribution mostly comes from the co-rotation region (as the potential perturbation is largest there). As the strength of self-gravity is increased by decreasing softening, the edge disturbance is modified, but the subsequent effect of the $d^2\delta\Phi/dr^2$ term on the energies cannot be anticipated a priori.

For $\epsilon_{g0,l} = 0.03$, the vortensity term does not dominate the modified total energy as much. However, this case is in fact not self-consistent because the softening used to setup the gap profile is an order of magnitude larger than that used in linear calculations. The limit of zero softening is also physically irrelevant because the disc has finite thickness. For the self-consistent cases with reasonable softening values (0.3, 0.6), the vortensity term dominates the energy balance, so the analytic description developed above works reasonably well for more realistic setups.

4.5.5 Edge mode boundary issues

The boundary condition $W' = 0$ was applied for simplicity. Vortex modes in low mass or non-self-gravitating discs are localised and insensitive to boundary conditions (de Val-Borro et al., 2007). This is explicitly demonstrated in Table 4.3 where the fiducial problem was solved with self-gravity switched off in the linear response, for various boundary effects described below. The edge dominated modes rely on self-gravity and are therefore intrinsically global, boundary effects cannot be assumed unimportant a

4. EDGE MODES IN SELF-GRAVITATING DISCS

Table 4.3: *Eigenfrequencies for miscellaneous linear calculations with different boundary conditions. These calculations neglect self-gravity and are therefore vortex modes. In the ‘none’ BC case, derivatives at the boundaries are approximated with interior grid points. Solutions are essentially independent of boundary effects.*

r_0	r_m	BC	$-\sigma_R$	$-\gamma \times 10^2$	r_c
1	10	$W' = 0$	0.1458	0.2041	5.7959
2.5	10	$W' = 0$	0.1458	0.2041	5.7930
1	10	none	0.1458	0.2041	5.7959
1.1	9.8	$W' = 0$	0.1458	0.2041	5.7941

Table 4.4: *Same as Table 4.3 but now with self-gravity, so these are the edge modes.*

r_0	r_m	BC	$-\sigma_R$	$-\gamma \times 10^2$	r_c
1	10	$W' = 0$	0.1587	0.4515	5.4626
2.5	10	$W' = 0$	0.1587	0.4689	5.4640
1	10	none	0.1591	0.4904	5.4538
1.1	9.8	$W' = 0$	0.1591	0.4794	5.4542

priori.

Additional experiments were performed with varying boundary conditions for the fiducial case ($Q_o = 1.5$). These include re-locating the inner boundary to $r = 1.1$, 2.5 or the outer boundary to $r = 9.8$ and approximating/extrapolating boundary derivatives using interior grid points (Adams et al., 1989). In the last case, the linear ODE is applied at endpoints of the grid so that boundary conditions are not explicitly set. This is strictly a numerical procedure to generate a linear system to solve, which may or may not correspond to physical solutions. However, this numerical experiment gave similar results to cases where boundary conditions are explicitly imposed.

For self-gravitating disc calculations giving edge dominated modes, Table 4.4 show these various conditions gave co-rotation radii varying by about 0.2% and growth rates varying by at most 10%. Overall the eigenfunctions W are similar, e.g. Fig. 4.11, showing the essential features of the edge mode, including relative amplitude of outer disc wave region and the co-rotation region. The existence of edge dominated modes is therefore not too sensitive to boundary effects. Physically this is because edge modes are mainly driven by the local vortensity maximum or edge, which is an internal feature away from boundaries.

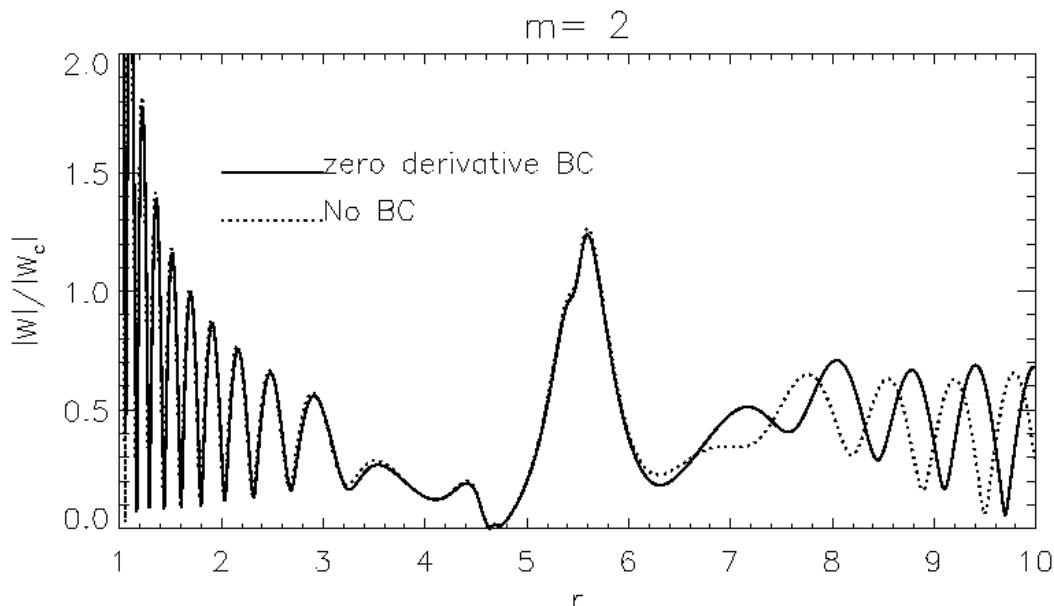


Figure 4.11: *Effect of boundary conditions on edge modes. The solid curve denote the reference solution with $W' = 0$ applied at boundaries, the dotted line indicate evaluation of linearised equations at boundaries by approximating boundary derivatives using interior points.*

4.6 Nonlinear hydrodynamic simulations

Having understood the basic physics of linear edge modes, their nonlinear evolution can be studied with hydrodynamic simulations. Disc models with $1.2 \leq Q_o \leq 2$, corresponding to disc masses $0.079M_* \geq M_d \geq 0.047M_*$, are simulated. Most simulations use a viscosity of $\nu = 10^{-5}$ which has been typically adopted for protoplanetary discs. This value has also been found to suppress vortex instabilities (de Val-Borro et al., 2007). The fiducial self-gravitational softening parameter is set to $\epsilon_{g0} = 0.3$. The planet has mass $M_p/M_* = 3 \times 10^{-4}$ and is set on a fixed circular orbit at $r_p = 5$ in order to focus on gap stability. The planet potential is introduced at $25P_0$ and its gravitational potential ramped up over $10P_0$, where P_0 is the Keplerian period at $r = 5$. The effects of varying viscosity and softening lengths are explored later in this section. Planetary migration is briefly considered in the next section and explored in more detail in Chapter 5.

4. EDGE MODES IN SELF-GRAVITATING DISCS

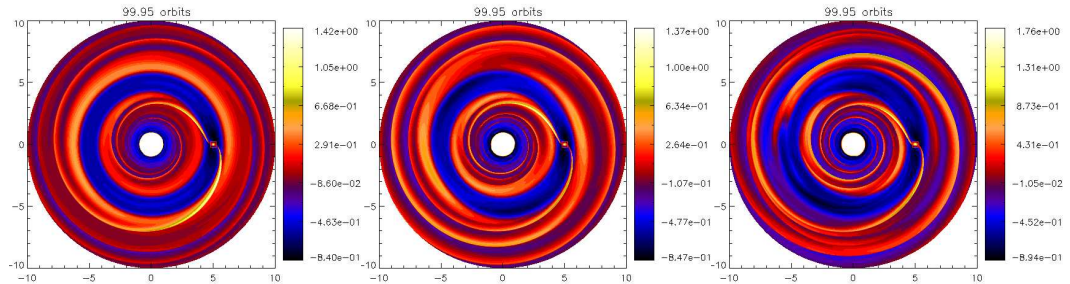


Figure 4.12: Surface density perturbations (relative to $t = 0$) at $t = 100P_0$ as a function the minimum value of Toomre Q in the disc, Q_o . From left to right: $Q_o = 2.0, 1.5, 1.2$.

4.6.1 Numerics

The numerical setup and methods are the same as those employed in Chapter 3. The hydrodynamic equations are evolved with the `FARGO` code including self-gravity (Masset, 2000a; Baruteau & Masset, 2008). The disc is divided into $N_r \times N_\phi = 768 \times 2304$ grid points in radius and azimuth giving a resolution of $\Delta r/H = 16.7$ cells per scale-height. The grid cells are nearly square, with $\Delta r/r\Delta\phi = 1.1$. Open boundaries are imposed at r_i and non-reflecting boundaries at r_o (Godon, 1996)¹. The latter has also been used in the self-gravitating disc-planet simulations of Zhang et al. (2008). As argued above, because edge modes are physically driven by an internal edge, boundary conditions are not expected to significantly affect whether or not they exist. Indeed, simulations with open outer boundaries, or damping boundaries (de Val-Borro, 2006) all show development of edge modes.

4.6.2 Overview

Fig. 4.12 shows the relative surface density perturbation ($\Delta\Sigma/\Sigma$) for $Q_o = 1.2, 1.5$ and 2.0 . The gap profile formed in these cases have local $\max(Q) = 3.3, 4.2, 5.3$ near the outer gap edge and the average Q for $r \in [6, r_o]$ is $1.6, 2.0, 2.6$. Although the edge mode is associated with the local vortensity maximum (located close to $\max(Q)$ for gap profiles), it requires coupling to the external smooth disc via self-gravity. Hence, edge modes only develop if Q in these smooth regions is sufficiently small, otherwise global disturbances cannot be constructed.

¹Here, open boundary means the flow out of the domain is allowed while flow into the domain is inhibited. The non-reflecting boundary applies a similar condition, but to flow characteristics (Riemann invariants), rather than to the primitive flow variables directly. Waves then propagate out of the domain but not into it.

The case $Q_o = 2.0$ gives a standard result for gap-opening planets, typically requiring planetary masses comparable or above that of Saturn. For that mass a partial gap of depth $\sim 50\%$ is formed, a steady state attained and remains stable to the end of the simulation. This serves as a stable case for comparison with more massive discs. This simulation was repeated with $\nu = 10^{-6}$, in which case we identified both vortex and edge modes. The standard viscosity $\nu = 10^{-5}$ thus suppress both types of instabilities for $Q_o = 2$. Although $\nu = 10^{-5}$ was adopted to avoid complications from vortex modes, note that they are stabilised by sufficiently massive discs anyway (e.g. $Q_o = 1.5$, $\nu = 10^{-6}$ do not show vortices, see §4.2 and Chapter 3).

Gap edges become unstable for the fiducial case $Q_o \leq 1.5$, $\nu = 10^{-5}$. As implied by linear analysis, the $m = 2$ edge mode dominates when $Q_o = 1.5$ and it saturates by $t = 100P_0$. It is more prominent in the outer disc, and over-densities deform the gap edge into an eccentric ring. The inner disc remains fairly circular, though weak non-axisymmetric disturbance ($m = 3$) can be identified. Note the global nature of the edge modes in the outer disc: spiral arms can be traced back to disturbances at the gap edge rather than the planet. Consider the spiral disturbance at the outer gap edge just upstream of the planet where $\Delta\Sigma/\Sigma$ is locally maximum. Moving along this spiral outwards, $\Delta\Sigma/\Sigma$ reaches a local minimum at $(x, y) = (6, -4)$ before increasing again. This is suggestive of the outer disc spirals being induced by the edge disturbance.

For the $Q_o = 1.2$ case, there is significant disruption to the outer gap edge, it is no longer clearly identifiable. The $m = 1$ edge mode becomes dominant in the outer disc, although overall it still appears to carry an $m = 2$ disturbance because of the perturbation due to the planet. The inner disc becomes visibly eccentric with an $m = 2$ edge mode disturbance. The shocks due to edge modes can comparable strength to those induced by the planet.

4.6.3 Development of edge modes for $Q_o = 1.5$

Fig. 4.13 shows the development of the edge mode in the fiducial disc $Q_o = 1.5$. The final dominance of $m = 2$ is consistent with linear calculations. The planet opens a well-defined gap by $t \sim 40P_0$, or $\sim 5P_0$ after the planet potential has been fully ramped up. At $t \sim 46P_0$, two over-dense blobs, associated with the edge mode can be identified at the outer gap edge. At this time the gap only has $\Delta\Sigma/\Sigma \simeq -0.3$, implying edge modes can develop *during* gap formation, since the steady state stable gap in $Q_o = 2$ reaches $\Delta\Sigma/\Sigma \simeq -0.6$. This can be expected from the discussion in §4.4.1.1. Edge modes are global and are associated with long trailing spiral waves with

4. EDGE MODES IN SELF-GRAVITATING DISCS

density perturbation amplitudes that are not small compared to that at the gap edge. This is clearly seen at $t \sim 50P_0$ when spiral shocks have already developed. The edge modes have a growth rate consistent with predictions from linear theory and become nonlinear within the time frame $46P_0 < t < 50P_0$, i.e. dynamical time-scales.

The edge perturbations penetrate the outer gap edge and trail an angle similar to the planetary wake. Notice the $m = 3$ disturbance near $r = 4$ in the snapshots taken at $t = 50P_0$ and $t = 56P_0$. Two of the three over-densities, seen as a function of azimuth, correlate to the $m = 2$ mode in the outer disc. The third over-density adjacent to the planet is probably caused by the outer planetary wake because of the disc self-gravity. It is also possible that different m modes have developed for the inner and outer discs, since the analytic description of edge modes instability only involves one edge.

The presence of both planetary wakes (with pattern speed $\simeq \Omega_k(5)$) and edge modes (with pattern speed $\simeq \Omega_k(5.5)$) of comparable amplitude enable direct interaction between them. Passage of edge mode spirals through the planetary wake may disrupt the former, explaining some of the apparently split-spirals in the plots. At $t \sim 64P_0$ Fig. 4.13 indicates that a spiral density wave feature coincides with the (outer) planetary wake. The enhanced outer planetary wake implies an increased negative torque exerted on the planet at this instant. This effect occurs during the overlap of positive density perturbations. Assuming that at a fixed radius the edge mode spiral has azimuthal thickness nh where n is a dimensionless number, then the time taken for this spiral to cross the planetary wake is $\Delta T = nh/|\Omega_k(5.5) - \Omega_k(5)| \simeq 0.06nP_0$. This is $\ll P_0$ for $n = O(1)$. Such crossing spiral waves will induce sharply peaked oscillations in the disc-planet torque.

The nonlinear evolution of edge modes in disc-planet systems can give complicated surface density fields. The gap becomes highly deformed. Fig. 4.13 shows that large voids develop to compensate for the over-densities in spiral arms, leading to azimuthal gaps ($t = 64P_0$). At $t = 72P_0$, the gap width ahead of the planet is narrowed by the spiral disturbance at the outer edge trying to connect to that at the inner edge. However, this gap-closing effect is opposed by the planetary torques which tend to open it.

A quasi-steady state is reached by $t \geq 76P_0$ showing an eccentric gap. For $t = 99.5P_0$ additional contour lines are plotted to indicate two local surface density maxima along the edge mode spiral adjacent to the planet. This spiral is disjointed around $r \sim 7.5$ where on traversing an arm, there is a minimum in $\Delta\Sigma/\Sigma$. Taking co-rotation of the spiral at $r_c = 5.5$ gives the outer Lindblad resonance of an $m = 2$ disturbance at $r = 7.2$, assuming a Keplerian disc. This is within the disjointed region of the spiral

4.6 Nonlinear hydrodynamic simulations

arm, which supports the interpretation that the disturbance at the edge is driving activity in the outer disc by perturbing it gravitationally and launching waves at outer Lindblad resonances.

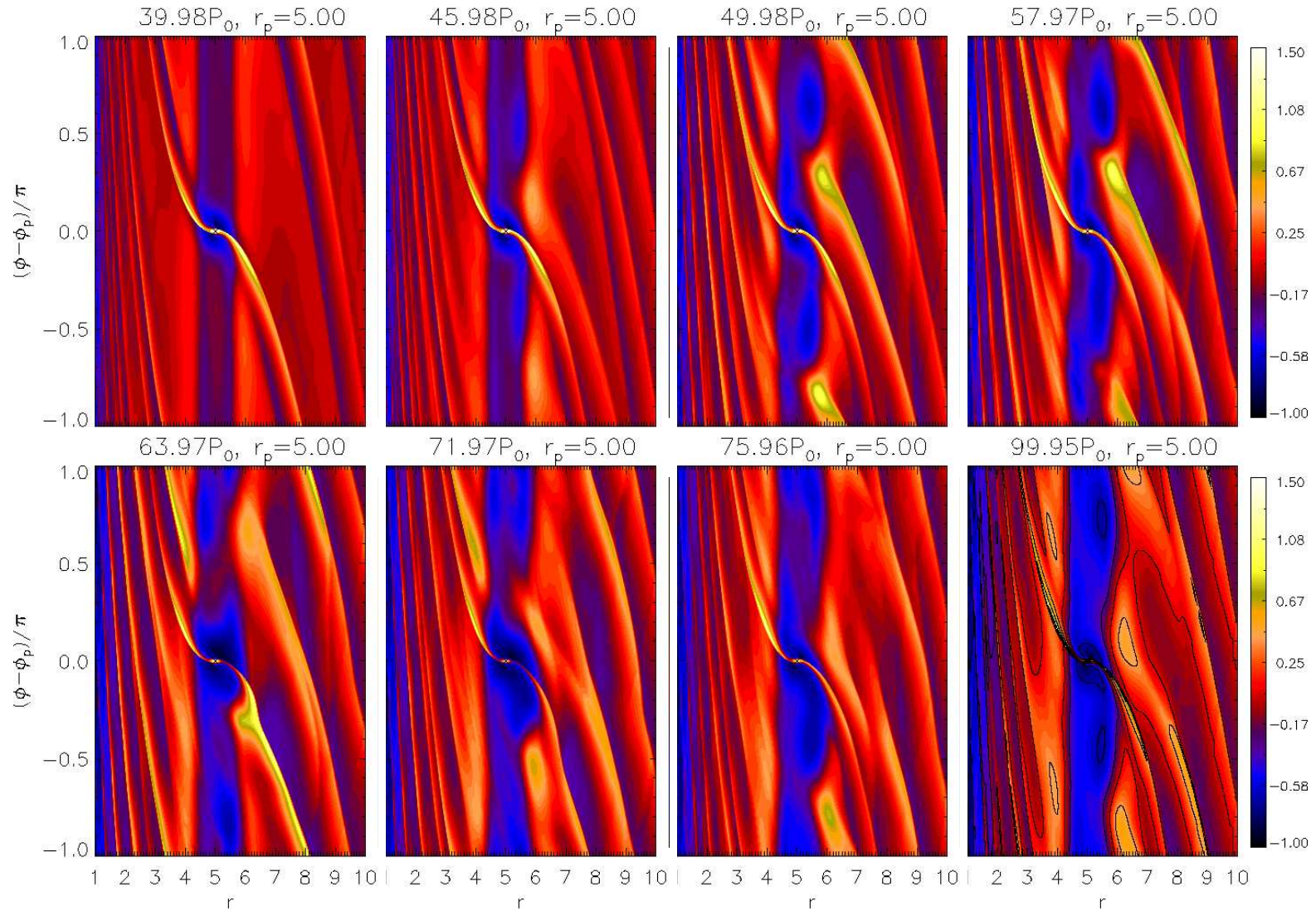


Figure 4.13: Evolution of the surface density perturbation (relative to $t = 0$) for the $Q_o = 1.5$ case from $t \sim 40P_0$ to $t \sim 100P_0$.

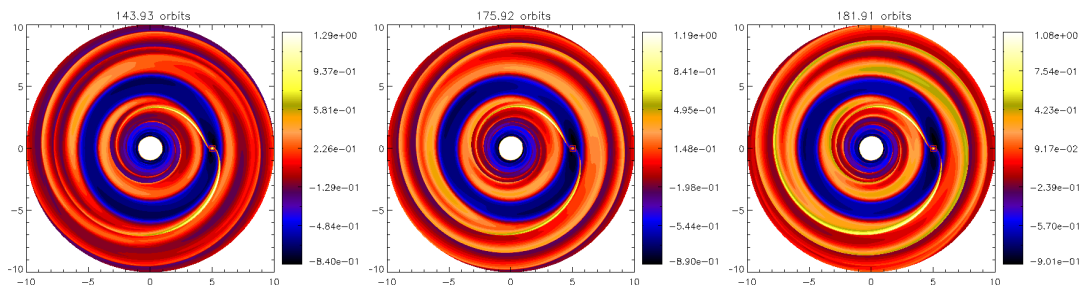


Figure 4.14: *Precession of an eccentric gap in the presence of an $m = 2$ edge mode for a disc with $Q_o = 1.5$.*

4.6.3.1 Eccentric gaps

Several snapshots of the $Q_o = 1.5$ run taken during the second half of the simulation are shown in Fig. 4.14, which display the precession of an eccentric outer gap edge. Deformation of a circular gap into an eccentric one requires an $m = 1$ disturbance. By inspection of the form of the relative surface density perturbation for $Q_o = 1.5$ (see Fig. 4.12), edge modes are seen associated with an eccentric outer gap edge.

In their non-self-gravitating disc-planet simulations, Kley & Dirksen (2006) found eccentric discs can result for planetary masses $M_p > 0.003M_*$ fixed on circular orbit. Simulations here show that edge modes in a self-gravitating disc can provide an additional perturbation to deform the gap into an eccentric shape for lower mass planets.

4.6.4 Evolution of the gap structure for $Q_o = 1.2$

With increasing disc mass, the gap profile evolves away from its original form. Fig. 4.15 illustrates the emergence of the $m = 3$ edge mode, expected from linear calculations, and subsequent evolution of the gap profile. By $t \sim 42P_0$ the $m = 3$ edge mode has already become nonlinear, whereas for $Q_o = 1.5$ the $m = 2$ mode only begins to emerge at $t \sim 44P_0$, being consistent with linear calculations that the former has almost twice the growth rate of the latter. In Fig. 4.15 there is a $m = 3$ spiral at $r < r_p$ in phase with the mode for $r > r_p$. The gap narrows where they almost touch. These are part of a single mode. By $t = 44P_0$, shocks have formed and extend continuously across the gap. The self-gravity of the disc has overcome the planet's gravity which is responsible for gap-opening. Edge mode spirals can be more prominent than planetary wakes.

The evolution of the form of the gap is shown in the one-dimensional averaged profiles plotted in Fig. 4.16. The quasi-steady profile before the onset of the edge mode instability is $t \sim 40P_0$. By $t = 44P_0$, the original bump at $r = 6$ has diminished. This

4. EDGE MODES IN SELF-GRAVITATING DISCS

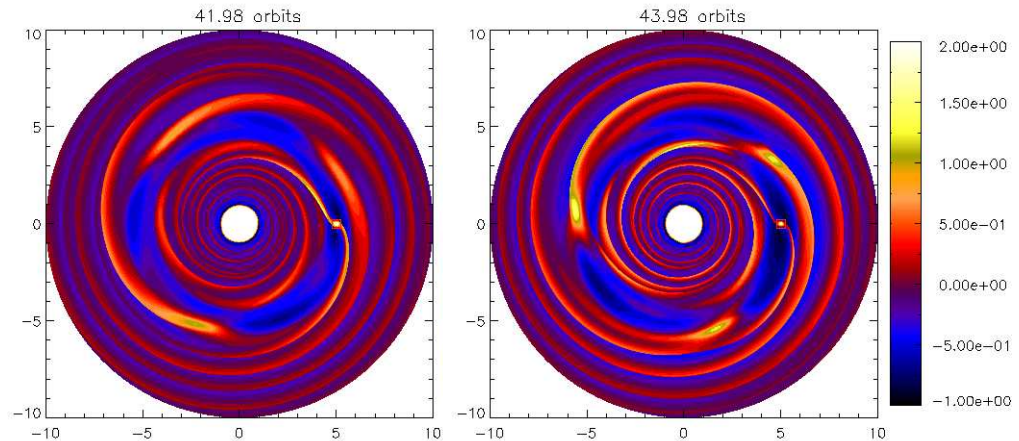


Figure 4.15: *The disc model with $Q_o = 1.2$: emergence of the $m = 3$ edge mode.*

bump originates from the planet expelling material as it opens a gap but is subsequently ‘undone’ by the edge mode (with associated over-density around co-rotation at $r \simeq 5.5$) as it grows and attempts to connect across the gap. This gap filling effect is possible as non-axisymmetric perturbation of the inner disc may be induced via the outer disc self-gravity. Notice also in Fig. 4.16 at $t = 44P_0$ the increased surface density at $r \simeq 7.5$ implying radial re-distribution of material. By $t = 120P_0$, the bump at the inner gap edge ($r = 4$) is similar to that produced by a standard gap-opening planet. By contrast, the outer bump cannot be maintained. The gap widens and there is an overall increase in surface density for $r > 6.5$

4.6.4.1 Vortensity field

Fig. 4.17 illustrates the evolution of the vortensity profile within the gap. The development of edge modes temporarily reduces the amplitude of the vortensity maxima set up by the perturbation of the planet. This is particularly noticeable when the profiles at $t = 40P_0$ and $t = 44P_0$ are compared. However, vortensity is generated through material passing through shocks induced by the planet. This provides a vortensity source that enables the amplitudes of vortensity maxima to increase again, as can be seen from the profile at $t = 60P_0$. The maxima remain roughly symmetric about the planet’s location at $r_p \pm 2r_h$. Note a general increase in the co-orbital vortensity level caused by fluid elements repeatedly passing through shocks.

The corresponding non-axisymmetric vortensity field is shown in Fig. 4.18. The important point here is that, edge modes destroy the axisymmetric rings of vortensity

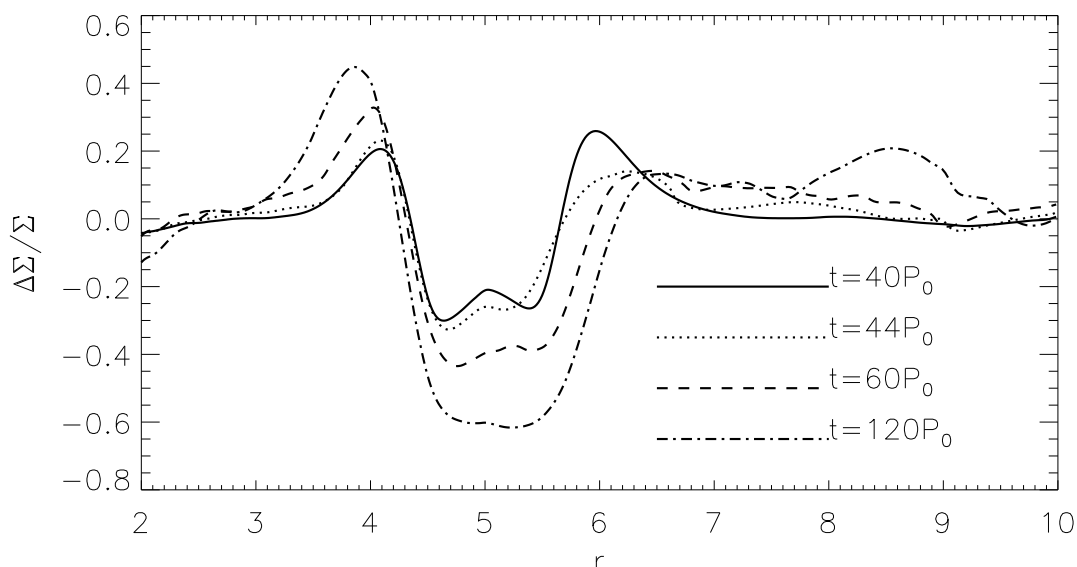


Figure 4.16: *The disc model with $Q_o = 1.2$: evolution of azimuthally averaged surface density perturbation.*

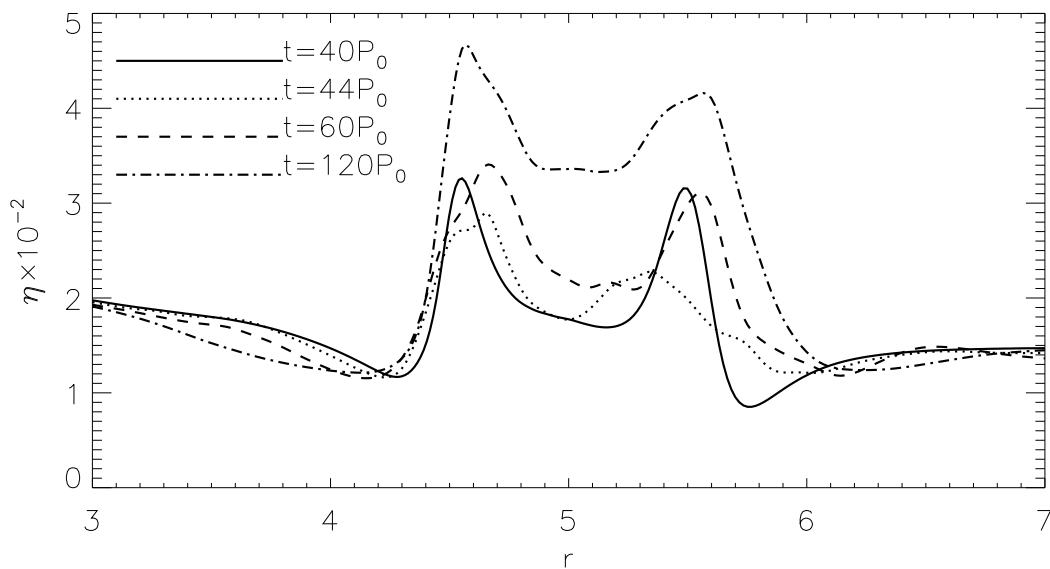


Figure 4.17: *The disc model with $Q_o = 1.2$: evolution of the azimuthally averaged vortensity η profile in the region of the gap.*

maxima, because they are associated with $\max(\eta)$. However, for the outer gap edge, local vortensity minima resides exterior to maxima. Hence, global edge modes in the outer disc affect vortensity minima too. This contrasts to vortex modes, which are localised and associated with vortensity minima and their development do not affect

4. EDGE MODES IN SELF-GRAVITATING DISCS

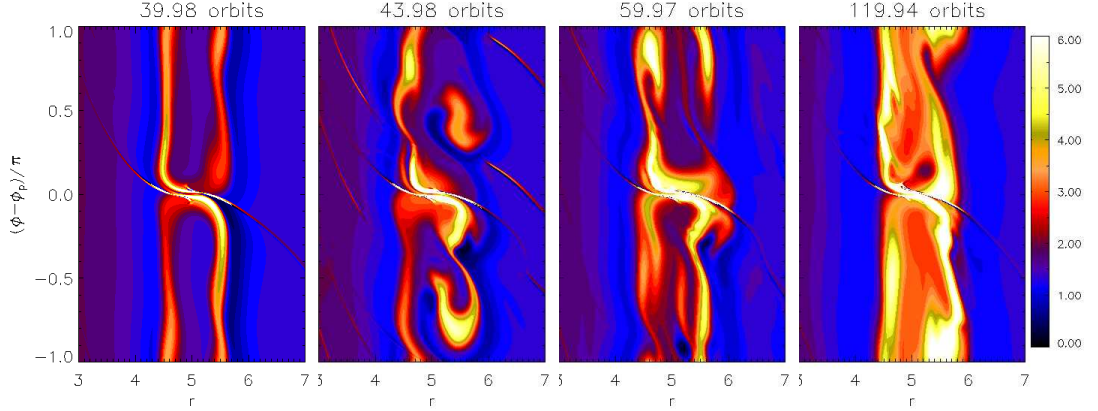


Figure 4.18: *Evolution of the non-axisymmetric vortensity field for the $Q_o = 1.2$ disc model.*

the vortensity maxima rings.

4.6.5 Similarity between an edge mode and a planet

The similarity between the effects induced by an edge mode and an external perturber can be further illustrated by the induced mass transport. Just as when a planet opens a gap, the development of an edge disturbance results in a radial re-distribution of mass. Angular momentum transport can be expressed in terms of an α viscosity parameter, defined through $\alpha = \Delta u_r \Delta u_\phi / c_s^2$, where Δ here denotes deviation from the azimuthal average.

Angular momentum is also transmitted through gravitational torques. However, in practice this was found to be a small effect compared to transport due to Reynolds stresses. The discs here are not gravito-turbulent. The azimuthally averaged α parameters $\langle \alpha \rangle$ are shown in Fig. 4.19. The stable disc with $Q_o = 2$ has non-axisymmetric features entirely induced by the planetary perturbation and $\langle \alpha \rangle$ is localised about the planet's orbital radius giving a two-straw feature about r_p . This is typical of disc-planet interactions, and it provides a useful case for comparison with more massive discs to see the effect of edge modes and their spatial dependence.

Increasing self-gravity to $Q_o = 1.5$, angular momentum transport is enhanced around r_p and for $r < 4$, signifying the global nature of edge modes. Towards the inner boundary, $\alpha \simeq 10^{-3}$ being twice as large as for the case with $Q_o = 2$. The $Q_o = 1.5$ case also has enhanced wave-like behaviour for $r > 7$ as compared to $Q_o = 2.0$.

The $Q_o = 1.2$ case is more dramatic, with $\langle \alpha \rangle$ reaching a factor of 2–3 times larger than the imposed physical viscosity around the planet's location ($\nu/h^2 = 4 \times 10^{-3}$).

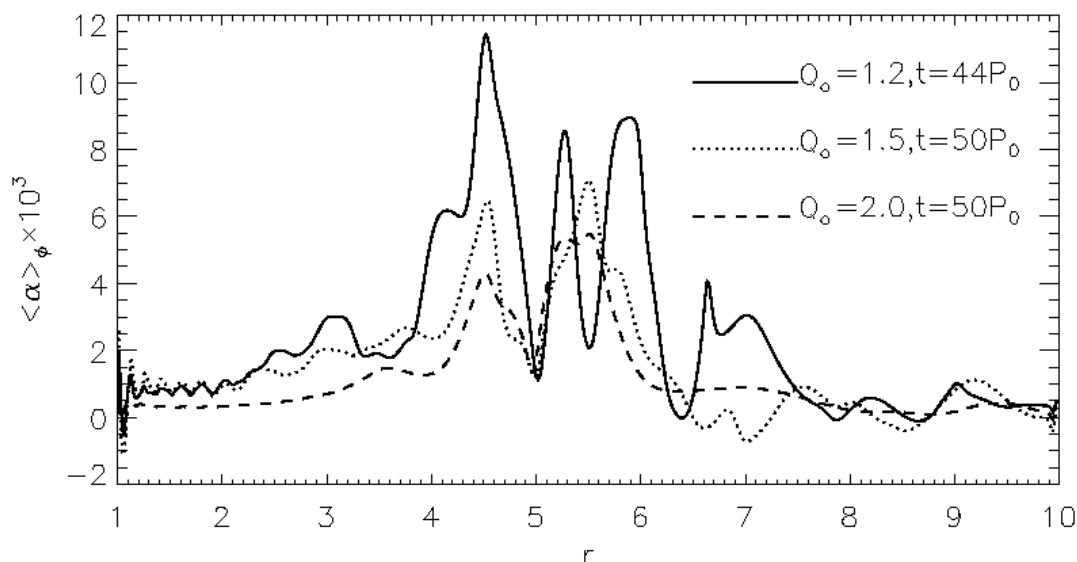


Figure 4.19: Azimuthally averaged Reynolds stresses associated with edge modes when they become nonlinear ($Q_o = 1.2, 1.5$), the $Q_o = 2$ case is shown as a stable case with no edge modes for comparison.

This case displays a three-straw feature, with an additional trough at about $r = 5.5$, i.e. close to edge mode co-rotation, as if another planet were placed there. Although the $Q_o = 1.5$ case also develops edge modes, there is no such additional trough at co-rotation. This may be because the $Q_o = 1.2$ case is more unstable, developing an $m = 3$ mode (which produces 3 additional effective coorbital planets placed at the gap edge), whereas the $Q_o = 1.5$ case develops a less prominent $m = 2$ mode. However, compared to $Q_o = 1.5$, $Q_o = 1.2$ has no enhanced transport away from the gap region.

Fig. 4.20 shows the evolution of disc mass as a function of time, $M_d(t)$. Viscous evolution leads to mass loss ($Q_o = 2.0$) and there is clearly enhanced mass loss if edge modes develop ($Q_o = 1.2, 1.5$). For $t \leq 50P_0$ the curves are identical, this interval includes gap formation and the emergence of edge modes. When edge modes become nonlinear, mass loss is enhanced ($t \simeq 50P_0$), but there is no significant difference between $Q_o = 1.2, 1.5$, which is consistent with the previous $\langle \alpha \rangle$ plots near the inner boundary. The $Q_o = 1.2$ curve is non-monotonic, though still decreasing overall, possibly because of boundary effects, and mass loss is enhanced once more around $t = 125P_0$. As Q_o is lowered, $M_d(t)$ becomes less smooth, showing the dynamical nature of edge modes.

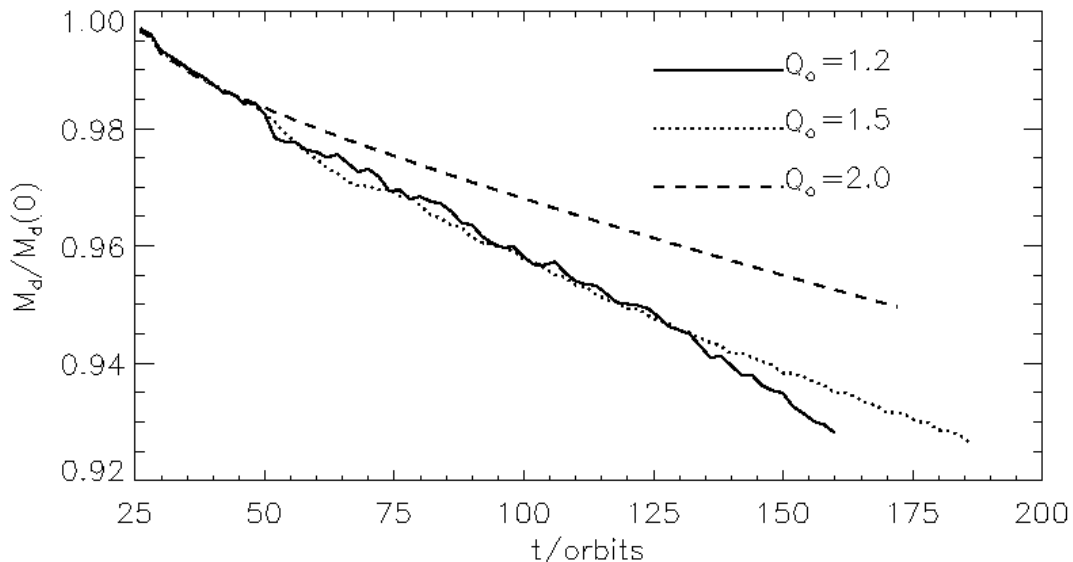


Figure 4.20: Evolution of the total disc mass for the three cases in Fig. 4.19, scaled by initial disc mass.

4.6.6 Additional effects

Edge modes are affected by softening and viscosity. To analyse mode amplitudes in these cases, the surface density field is Fourier transformed in azimuth giving $a_m(r)$ as the (complex) amplitude of the m^{th} mode. a_m is then integrated over the outer disc $r > 5$ to obtain $C_m = \int a_m dr$.

4.6.6.1 Softening

In linear theory, edge modes could not be constructed for gravitational softening parameters that are too large. The $Q_o = 1.5$ disc has been simulated with $\epsilon_{g0} = 0.6, 0.8, 1.0$ to confirm this.

Fig. 4.21 shows the evolution of the running-time averaged $m = 2$ amplitudes. The evolution for all cases is very similar for $t \leq 40P_0$. Unstable modes develop thereafter, with increasing growth rates and saturation levels as ϵ_{g0} is lowered. $\epsilon_{g0} = 1$ displays either no growth or a much reduced growth rate. The gap for $Q_o = 1.5, \epsilon_{g0} = 1.0$ reaches a steady state close to the stable case with $Q_o = 2.0, \epsilon_{g0} = 0.3$. This is consistent with the fact that as softening increases, edge disturbances can no longer perturb the remainder of the disc via self-gravity. The edge itself is also stabilised against gravitationally dominated disturbances, according to linear theory. The residual

non-axisymmetric structure is due to the planetary perturbation.

Interestingly, the decrease in saturation level in going from $\epsilon_{g0} = 0.3 \rightarrow 0.6$ is smaller than going from $\epsilon_{g0} = 0.6 \rightarrow 0.8$, which is in turn smaller than $\epsilon_{g0} = 0.8 \rightarrow 1.0$, despite a larger relative increase in softening in one pair than the next. This is suggestive of convergence for the nonlinear saturation amplitude as $\epsilon_g \rightarrow 0$. This can be expected from convergence in linear theory (§4.5.4.1). However, at $\epsilon_{g0} = 0.3$ there are only 5 cells per softening length. To probe even smaller ϵ_{g0} requires prohibitively high resolution simulations. Furthermore, since ϵ_g approximately accounts for the disc's non-zero vertical extent, very small softening lengths are not physically relevant to explore.

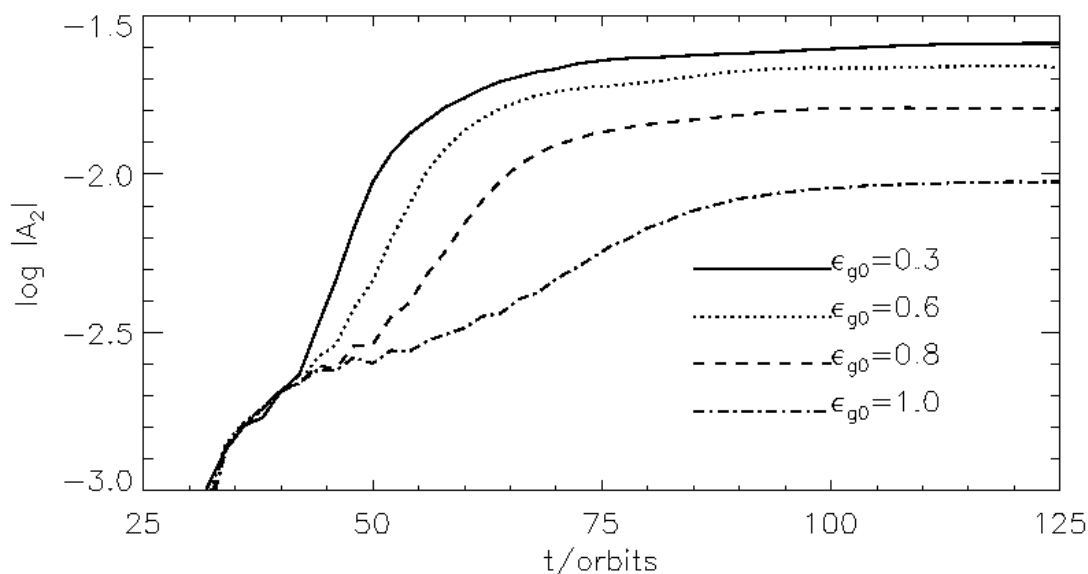


Figure 4.21: *Running-time average of the surface density $m = 2$ Fourier amplitude, integrated over $r \in [5, 10]$ and scaled by the axisymmetric component, for the $Q_o = 1.5$ disc, as a function of the gravitational softening length.*

4.6.6.2 Viscosity

An important difference between edge dominated modes and the already well-studied vortex modes in planetary gaps is the effect of viscosity on them. The standard viscosity $\nu = 10^{-5}$ prevents vortex mode development, but §4.2 showed that such viscosity values still allow the $m = 2$ edge modes to develop. Vortices are localised disturbances and more easily smeared out than global spirals in a given time interval. Hence, vortex

4. EDGE MODES IN SELF-GRAVITATING DISCS

growth is inhibited more easily by viscosity than spirals.

The $Q_o = 1.5$ run was repeated with a range of viscosities. Results are shown in Fig. 4.22. Generally, lowering viscosity increases the amplitudes of the non-axisymmetric modes. For $\nu \leq 10^{-6}$, the $m = 3$ mode emerges first (note that this is not in conflict with linear calculations which used $\nu = 10^{-5}$ in its basic state), rather than $m = 2$. This is because lowering viscosity allows a sharper vortensity peak to develop in the background model, and thus has the same effect as increasing the disc mass. The latter generally enable higher m edge modes.

Unlike vortex modes, even with twice the standard viscosity ($\nu = 2 \times 10^{-5}$) the edge mode develops and grows. It is only suppressed when $\nu \geq 5 \times 10^{-5}$. However, this is because no vortensity maxima could be setup at the gap edges due to vortensity diffusion in the co-orbital region making an almost uniform vortensity distribution in the gap (e.g. the final plot in Fig. 4.17). Since the necessary condition for edge modes is not achieved, no linear modes exist. This differs from the nature of the suppression of vortex modes, because in that case vortensity extrema can still be setup.

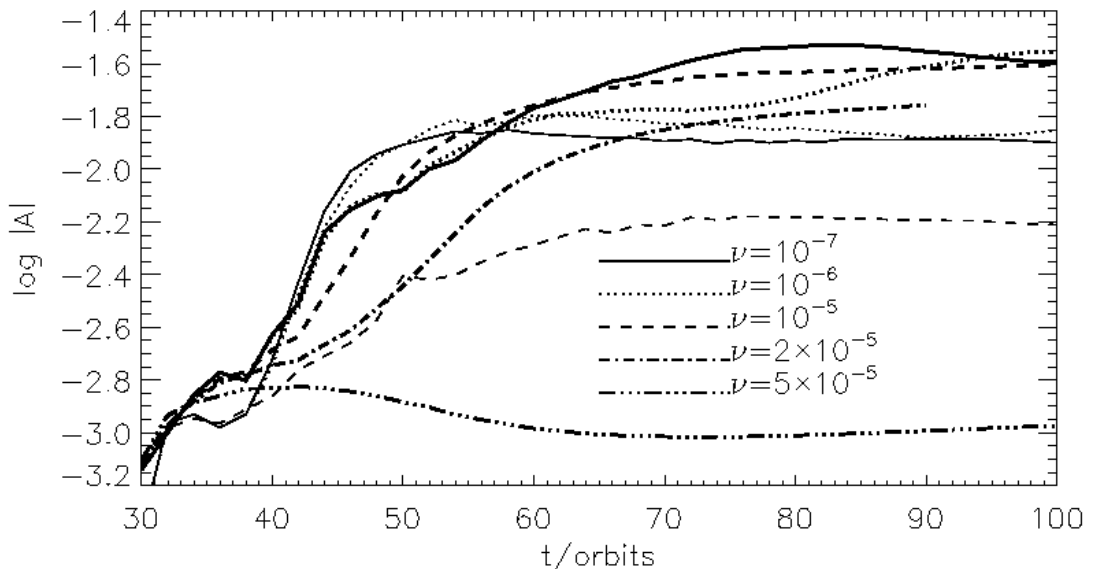


Figure 4.22: Mode amplitudes as a function of applied viscosity ν for $Q_o = 1.5$. The Fourier amplitudes are scaled by the axisymmetric component and its running-time average plotted. Thick lines indicate the $m = 2$ mode and is plotted for all ν . Thin lines indicate the $m = 3$ mode and are plotted for $\nu = 10^{-7} - 10^{-5}$ only.

4.7 Applications to disc-planet interactions

In order to focus on the issue of gap stability, the planet was held on a fixed orbit. However, torques exerted by the non-axisymmetric disc on the planet will induce migration that could occur on a short time scale (e.g. Pepliński et al., 2008b). Such torques will be affected by the presence of edge modes.

4.7.1 Disc-planet torques

The presence of large-scale edge mode spirals of comparable amplitude as the planetary wake will significantly modify disc-on-planet torques. Here, this torque is measured while still keeping the planet on fixed orbit. This eliminates torques *due* to migration.

Fig. 4.23(a) shows the evolution of the torque per unit length for the $Q_o = 1.5$ fiducial case. The torque profile at $t = 40P_0$, before the edge modes have developed significantly, shows the outer torque is larger in magnitude than inner torque, implying inward migration. This is a typical result for disc-planet interactions and serves as a reference. At $t = 60P_0$ and $t = 100P_0$, edge modes mostly modify the torque contributions exterior to the planet, though some disturbances are seen in the interior disc ($r - r_p < -5r_h$, recall r_h is the Hill radius). The original outer torque at $+r_h$, is reduced in magnitude as material re-distributed to concentrate around the co-rotation radius r_c of the edge mode (r_c is $\sim 2r_h$ away from the planet).

Edge modes can both enhance or reduce disc-planet torques associated with planetary wakes. Consider $r - r_p \in [3, 5]r_h$ in Fig. 4.23(a). Comparing the situation at $t = 60P_0$ to that at $t = 40P_0$, the torque contribution from this region is seen to be more negative. This is because an edge mode spiral overlaps the planetary wake and therefore contributes an additional negative torque on the planet. However, at $t = 100P_0$, an edge mode spiral is just upstream of the planet, exerting a positive torque on the planet, hence the torque contribution from $r - r_p \in [3, 5]r_h$ becomes positive.

Differential rotation between edge modes and the planet produces oscillatory torques, shown in Fig. 4.23(b). Unlike typical disc-planet interactions, which produce inward migration, the total instantaneous torques in the presence of edge modes can be of either sign. For a disturbance with m fold symmetry the time interval between encounters with the planet is $\Delta T = (2\pi/m)/\delta\Omega$, where $\delta\Omega$ is the difference in angular velocity of the planet and the disturbance pattern. Approximating $\delta\Omega = |\Omega_k(r_p) - \Omega_k(r_c)|$ with $r_c = 5.46$ obtained from linear theory gives $\Delta T = 4.04P_0$. Indeed, the oscillation period in Fig. 4.23 is $\simeq 4P_0$. Both inner and outer torques oscillate, but since the edge mode is more prominent in the outer disc, oscillations in the outer torque are larger,

4. EDGE MODES IN SELF-GRAVITATING DISCS

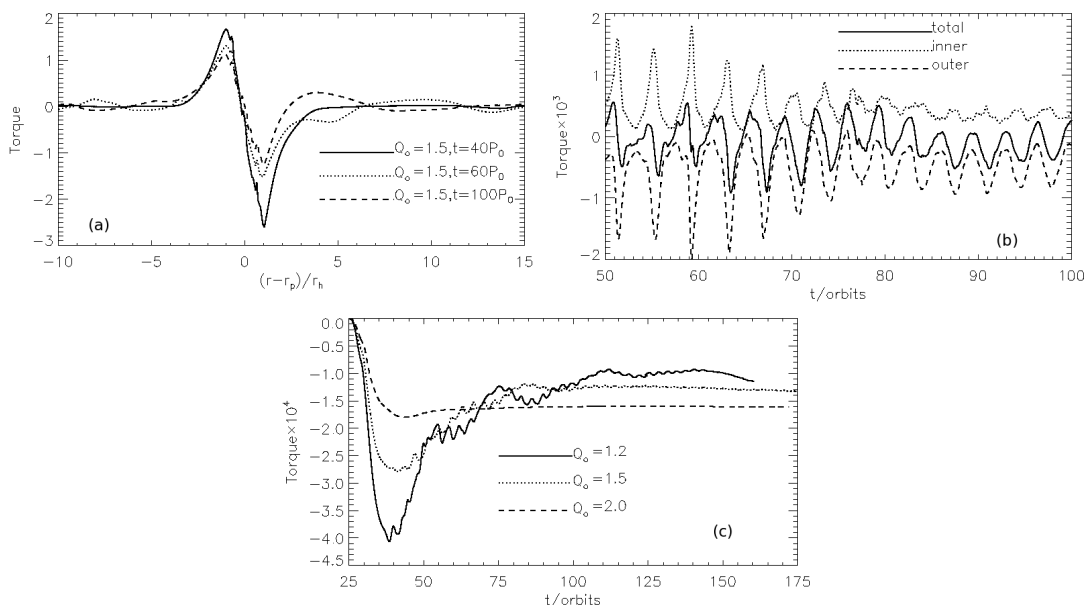


Figure 4.23: (a) Disc torque per unit radius acting on the planet. (b) Time dependent evolution of the disc planet torques, the contribution from the inner disc (dotted curve), the outer disc (dashed curve) and the total torque (solid curve). (c) The running time average of the total torque acting on the planet as a function of Q_o .

particularly after mode saturation ($t > 70P_0$).

Time-averaged torques are compared in Fig. 4.23(c). In all three disc masses, on average a negative torque acts on the planet and its magnitude largest at $t = 40P_0$. Up to this point, the torque becomes more negative as the disc mass increases, as expected. However, development of edge modes makes the averaged torques more positive, and eventually reverses the trend with Q_o . While the disc with $Q_o = 1.5$ also attains steady torque values as in $Q_o = 2$, the disc with $Q_o = 1.2$ has significant oscillations even after time-averaging and is remains non-steady at the end of the run.

Before the instability sets in, torques arise from wakes induced by the planet, and the outer (negative) torque is dominant. Edge modes concentrate material into spiral arms, leaving voids in between. Therefore, except when spiral arms cross the planet's azimuth, the surface density in the planetary wake is reduced because it resides in the void in between edge mode spirals. Hence the torque magnitude is reduced. If the reduction in surface density due to voids in between edge mode spirals is greater than the increase in surface density scale produced by decreasing Q_o , the presence of edge modes will, on average make the torque acting on the planet more positive.

4.7.2 Outward scattering by spiral arms

If edge modes develop, planetary migration may be affected by them. Their association with gap edges inevitably affects co-orbital flows. The more long term effect of edge modes on migration is explored in Chapter 5. This subsection highlights an interesting effect found when edge modes are present. This is *outward* migration induced by spiral arms analogous to vortex-induced rapid migration described in Chapter 2.

Some of the simulations described above were restarted at $t = 50P_0$ allowing the planet to move in response to the gravitational forces due to the disc. Fig. 4.24 shows the orbital radius of the planet in discs with edge modes present. No obvious trend with Q_o can be seen. This is because of oscillatory torques (positive or negative) due to edge modes and the partial gap associated with a Saturn mass planet, which is associated with type III migration (Masset & Papaloizou, 2003). The initial direction of migration can be inwards or outwards depending on the relative positioning of the edge mode spirals with respect to the planet at the time the planet was first allowed to move.

Outward migration is seen for $Q_o = 1.3$ and $Q_o = 1.5$. For $Q_o = 1.5$, the planet migrates by $\Delta r = 2$, or 8.6 times its initial Hill radius, within only $4P_0$. This is essentially the result of a scattering event. Repeating this run with tapering applied to contributions to disc torques from within $0.6r_h$ of the planet, outward scattering still occurs, but limited to $\Delta r = 1$ and the planet remains at $r \simeq 6$ until $t = 128P_0$. Hence, while the subsequent inward migration seen for $Q_o = 1.5$ may be associated with conditions close to the planet, the initial outward scattering is due to an exterior edge mode spiral.

For $Q_o = 1.3$ the planet is scattered to $r_p \simeq 8$. It is interesting to note that the subsequent inward migration for $Q_o = 1.5$ and $Q_o = 1.3$ stalls at $r \simeq 6$, i.e. the original outer gap edge. The planet remains there for sufficient time for both gap and edge mode development and for $Q_o = 1.3$ a second episode of outward scattering occurs. Interaction with edge modes can affect the disc well beyond the original co-orbital region of the planet. In the case of outward migration, it may promote gravitational activity in the outer disc. However, boundary effects may be important after significant outward scattering.

The spiral arm-planet interaction for $Q_o = 1.5$ above is detailed in Fig. 4.25. A spiral arm approaches the planet from the upstream direction ($t = 50.9P_0$) and exerts positive torque on the planet, increasing r_p . The gap is asymmetric in the azimuthal direction with the surface density upstream of the planet being larger than that downstream of the planet ($t = 51.5P_0$). As the spiral arm passes through the

4. EDGE MODES IN SELF-GRAVITATING DISCS

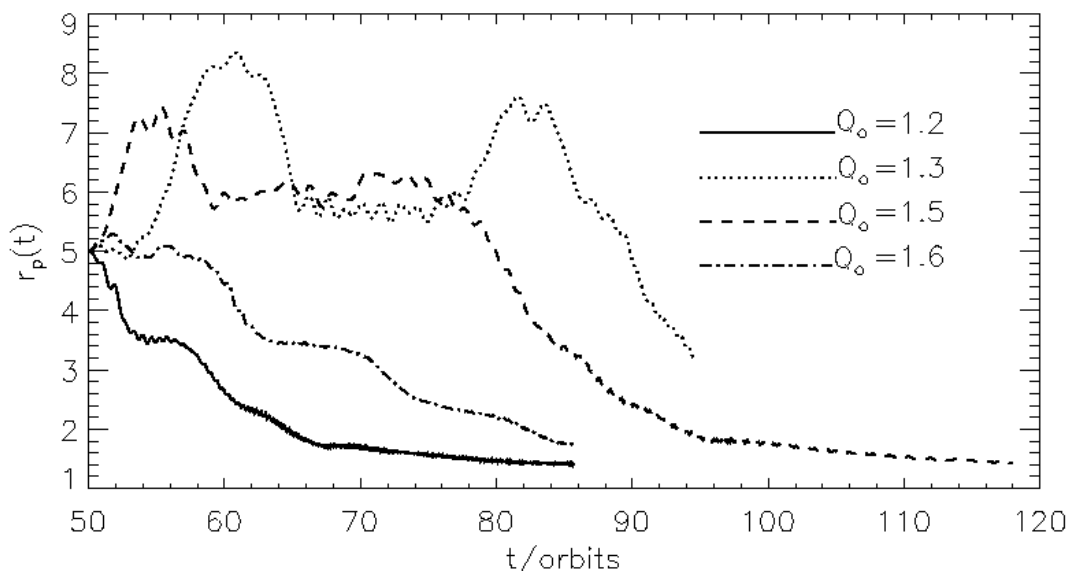


Figure 4.24: *Orbital radius evolution in discs with edge mode disturbances. The instantaneous orbital radius r_p is shown as a function of time.*

planetary wake ($t = 51.5P_0$ — $52.1P_0$) the gap surface density just ahead of the planet builds up as an edge mode is setup across the gap. Notice the fluid blob at $r = 5$, $\phi - \phi_p = 0.3\pi$. This signifies material executing horseshoe turns ahead of the planet. At the later time $t = 53.1P_0$ the planet has exited the original gap, in which the average surface density is now higher than before the scattering. Material that was originally outside the gap loses angular momentum and moves into the original gap. This material is not necessarily that composing the spiral.

Migration through scattering by spiral arms differs from vortex scattering. Vortices form about vortensity minima, which lie further from the planet than vortensity maxima. Vortensity maxima are stable in the case of vortex formation. This means the ring of vortensity maxima must be disrupted in order for vortices to flow across the planet's orbital radius for direct interaction. Edge modes themselves correspond to a disruption of vortensity maxima, hence direct interaction is less hindered than in the vortex case. Furthermore, a vortex is a material volume of fluid. Vortex-planet scattering results in an orbital radius change of both objects. However, a spiral pattern is generally not a material volume of fluid. It can be seen in Fig. 4.25 that the local $\max(\Sigma)$ in the spiral remains at approximately the same radius before and after the interaction. The spiral pattern as a whole does not move inwards. In this case, the spiral first increases the

4.7 Applications to disc-planet interactions

planet's orbital radius, thereby encouraging it to interact with the outer gap edge and scatter that material inwards.

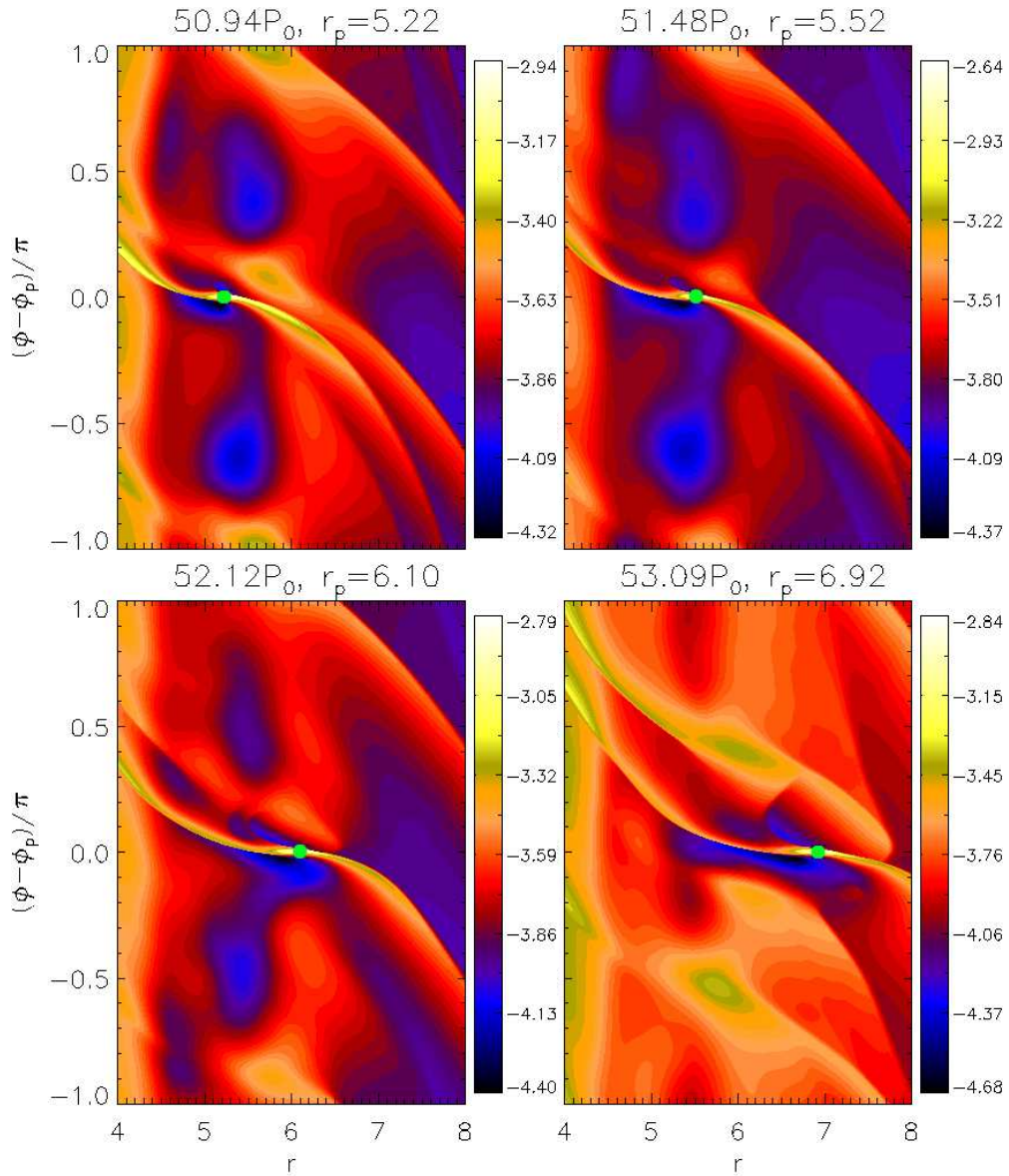


Figure 4.25: Interaction between an edge mode spiral and the planet (green dot), causing the latter to be scattered outwards. $\log \Sigma$ is shown.

4.8 Summary and discussion

The investigation of planetary gap stability has been extended to massive protoplanetary discs. Vortex instabilities, associated with local vortensity minima, have previously been found to occur in weakly or non self-gravitating discs. However, edge modes that are associated with local vortensity maxima and require sufficiently strong self-gravity were found in the more massive discs in this Chapter. These have very different properties in that they are global rather than local and are associated with large scale spiral arms as well as being less affected by viscous diffusion.

For the disc-planet models here, with fixed planet mass $M_p = 3 \times 10^{-4} M_*$ and fixed disc aspect-ratio $h = 0.05$, edge modes develop for gap profiles with an average Toomre $Q \lesssim 2$ exterior to the planet's orbital radius. In the unperturbed smooth disc, this corresponds to $Q \lesssim 2.6$ at the planet's location and $Q \lesssim 1.5$ at the outer boundary, or equivalently a disc mass $M_d \gtrsim 0.06 M_*$. For fixed M_d , nonlinear simulations show edge modes develop readily for uniform kinematic viscosity $\nu \lesssim 2 \times 10^{-5}$ and self-gravitational softening $\epsilon_g \lesssim 0.8H$.

A theoretical description of edge modes was developed. The edge mode is interpreted as a disturbance associated with an interior vortensity maximum that requires self-gravity to be sustained, and which further perturbs the remainder of the disc through its self-gravity causing the excitation of density waves. These density waves acts to destabilise the edge disturbance. Linear calculations confirm this picture and also show the expected strengthening of the instability when self-gravity is increased via the disc mass and the expected weakening of the instability through increasing gravitational softening.

Hydrodynamic simulations of disc-planet interactions confirm earlier suggestions by Meschiari & Laughlin (2008), who found gravitational instabilities associated with prescribed gap profiles without a planet. Indeed, edge modes can develop for gaps self-consistently opened by introducing a planet into the disc. However, models here showed their development *during* gap formation of a Saturn-mass planet, when the gap only consists of a 20–30% deficit in surface density relative to the unperturbed disc. This is much shallower than the Jovian-mass planetary gaps considered in Meschiari & Laughlin (2008). Edge modes exist for the disc models considered here with masses $M_d \gtrsim 0.06 M_*$, again this is less massive than required in Meschiari & Laughlin (2008), but more massive than those typically used in modelling protoplanetary discs.

Simulations show edge modes with $m = 2$ and $m = 3$. They have surface density maxima localised near the outer gap edge, but the disturbance they produce extends

throughout the entire disc, consistent with analytical expectation and linear calculations. One important difference between edge modes and the vortex forming modes in non-self-gravitating discs is that the latter require low viscosity. The typical viscosity values adopted for protoplanetary discs, $\nu = 10^{-5}$, will suppress vortex formation but not edge modes around a gap opened by a Saturn-mass planet.

Edge modes produce large oscillations in the disc torques acting on the planet, which can be positive or negative. The presence of edge modes reverses the trend of the time averaged disc-on-planet torque as a function of disc mass: the torque being more positive as disc mass increases. Direct interaction between the planet and a spiral arm associated with an edge mode is possible. These should be more prominent in the disc section with the lower Q . In practise this corresponds to $r > r_p$, so the planet tends to interact with the outer spirals, which results in a scattering of the planet outwards.

4. EDGE MODES IN SELF-GRAVITATING DISCS

Chapter 5

Planetary migration in the presence of edge modes

Edge modes significantly modify disc-planet torques from the classic picture of monotonic inwards migration in low-mass discs. Hence, migration of gap-opening planets in massive protoplanetary discs is expected to be qualitatively different to that in gravitationally stable discs. For example, §4.7.2 indicate the possibility of direct gravitational interaction with an edge mode spiral arm. This is analogous, although distinct, from vortex-planet scattering.

This Chapter explores the effect of edge modes associated with gap edges on planetary migration averaged over timescales much longer than that for scattering. Torque measurements from fixed-orbit simulations in Chapter 4 show the time-averaged disc-on-planet torques become more positive with disc mass. For Saturn-mass planets this effect is due to modification of Lindblad torques. However, for the Jovian mass planets adopted here, edge modes were found to supply torques from a region very close to the planet’s orbital radius.

This study is fully numerical and is organised as follows. The disc-planet model is described in §5.1 and their stability discussed in §5.2. Simulation results are presented in §5.3 and a fiducial case is analysed in §5.4. §5.5 concludes this Chapter with a discussion of the implications of the results.

5.1 Simulation setup

The class of disc profiles are identical to that used in Chapter 3—4. The discs are two-dimensional and are parameterised by their Keplerian Toomre value at the outer

5. PLANETARY MIGRATION IN THE PRESENCE OF EDGE MODES

boundary, Q_o , or equivalently the disc-to-star mass ratio M_d/M_* . These are chosen via the surface density scale. The discs have uniform kinematic viscosity $\nu = 10^{-5}$.

The disc spans $r \in [1, 25]$ and a larger planetary mass $M_p/M_* = 2 \times 10^{-3}$ is considered. This M_p opens a deep gap so that strong scattering with spiral arms leading to large jumps in orbital radius on orbital timescales, do not occur. It allows one to examine the long term effect of edge modes averaged over many orbits. Numerically, a larger M_p is advantageous because there are more cells per Hill radius than smaller M_p .

The planet is introduced on a circular orbit at $r_p = 10$, this is further from disc boundaries than in disc-planet models used previously. This attempts to reduce boundary effects, should the planet move significantly.

5.1.1 Equation of state

The equation of state is locally isothermal, so the vertically integrated pressure is $p = c_s^2 \Sigma$ where c_s is a specified function. However, as noted in Chapter 3, adopting $c_s^2 = h^2 GM_*/r$ (as used in most of the previous simulations) in a self-gravitating disc means the planet mass is no longer a free parameter. The effective planet mass is M_p plus disc material gravitationally bound to the planet. This effect will be significant for Jovian mass planets. To reduce it, one can artificially increase c_s near the planet by using the EOS described in §1.18. The parameter h_p in this EOS controls the temperature increase relative to the $1/r$ profile, and is fixed at $h_p = 0.5$. Far away from the planet, the disc aspect-ratio is fixed at $h = 0.05$ and the temperature profile returns to $\propto 1/r$. The motivation in using this EOS is entirely numerical.

5.1.2 Numerics

The disc is divided into $N_r \times N_\phi = 1024 \times 2048$ zones in radius and azimuth, logarithmically and uniformly spaced respectively. The resolution is approximately 16 cells per H (or 28 cells per r_h). The cells are nearly square ($\Delta r/r \Delta \phi = 1.02$). Open boundaries are applied at the inner boundary and non-reflecting at the outer boundary, as before. The self-gravity softening is $\epsilon_g = 0.3H$.

The planet is introduced with zero mass at $t = 20P_0$. Its mass is then switched on smoothly over $10P_0$ to its final value M_p . The planet is then allowed to respond to disc gravitational forces for $t > 30P_0$. A standard fifth order Runge-Kutta scheme is used to integrate its equation of motion. The softening length for the planet potential is $\epsilon_p = 0.6H$.

5.2 Disc models and their stability

Four discs are presented here to assess the stability of the adopted models: $Q_0 = 1.3, 1.5, 1.7, 2.0$. These correspond to a Toomre parameter at the planet's initial radius of $Q_p = 2.40, 2.77, 3.14, 3.70$ and total disc masses of $M_d/M_* = 0.092, 0.080, 0.071, 0.060$, respectively. The discs are gravitationally stable to axisymmetric perturbations. For smooth radial profiles, they are also expected to be gravitationally stable to non-axisymmetric perturbations near r_p because $Q_p > 1.5$.

Fig. 5.1 shows the surface density perturbation for the disc models when the planet mass reaches its final value. The gap structure is similar and relatively stable. The over-density at the outer gap edge upstream of the planet ($r \simeq 14, \phi - \phi_p \simeq 0.5\pi$) resembles an edge mode. However, this is likely associated with the introduction of a massive planet over a short time-scale, rather than instability of an axisymmetric gap. This is because it is present in all disc masses with the same azimuthal wavenumber ($m = 1$), whereas the most unstable edge mode is expected to be of higher m with increasing disc mass. This transient feature increases in magnitude with M_d , and for $Q_0 = 1.3$ it visibly makes the gap less clean.

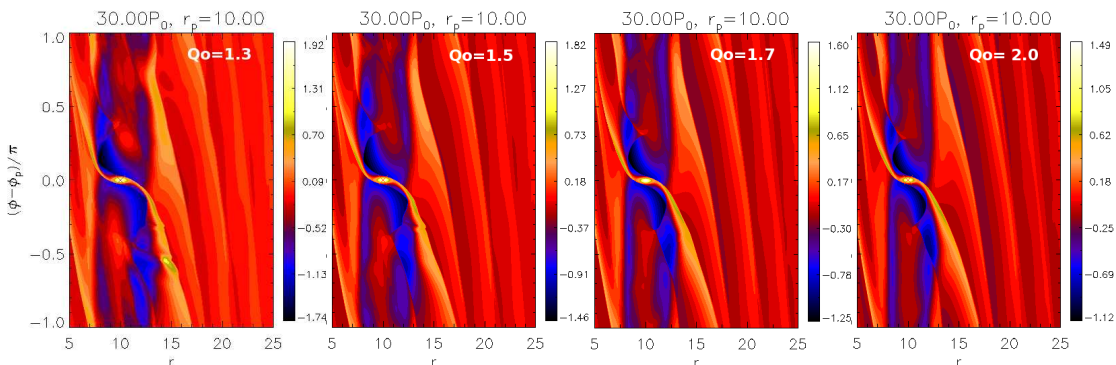


Figure 5.1: *Logarithmic relative surface density perturbation $\log[\Sigma/\Sigma(t=0)]$ when the planet reaches its final mass. The colour bar corresponds to the maximum and minimum perturbations within each plotted region, which occur near the planet.*

For stability discussion, one dimensional gap profiles should be used, because linear analysis is usually applied to axisymmetric backgrounds. Fig. 5.2 compares the azimuthally averaged surface density perturbation and Toomre Q profiles at $t = 30P_0$. The gap is almost cleared ($\Delta\Sigma/\Sigma \simeq -0.7$). The surface density perturbation is similar to that induced by a Saturn-mass planet, apart from the spike near r_p due to disc material bound to the planet. As expected, this causes the gap edges to become more

5. PLANETARY MIGRATION IN THE PRESENCE OF EDGE MODES

steep with lowering Q_o .

The fundamental quantity for stability is the vortensity η . The edge mode is associated with $\max(\eta)$, which for gap profiles coincide with $\max(Q)$. Fig. 5.2 shows that edge mode corotation r_c should occur inside the gap. Assuming the coorbital region of a giant planet is $|r - r_p| \lesssim 2.5r_h$ (Artymowicz, 2004b; Paardekooper & Papaloizou, 2009), r_c is just within the outer gap edge. This suggests the development of edge modes could bring over-densities into the coorbital region of the planet.

Edge modes also require coupling to the outer smooth disc, which is easier with lower Q_o . Interestingly, in Fig. 5.2 the Q profile beyond the outer gap edge becomes smoother as Q_o is reduced. The local $\min(Q)$ at the outer gap edge is most pronounced for $Q_o = 2.0$, which favour vortex modes. The profiles here indicate edge modes will develop more readily with increasing disc mass and therefore have more significant effect on migration as Q_o is lowered.

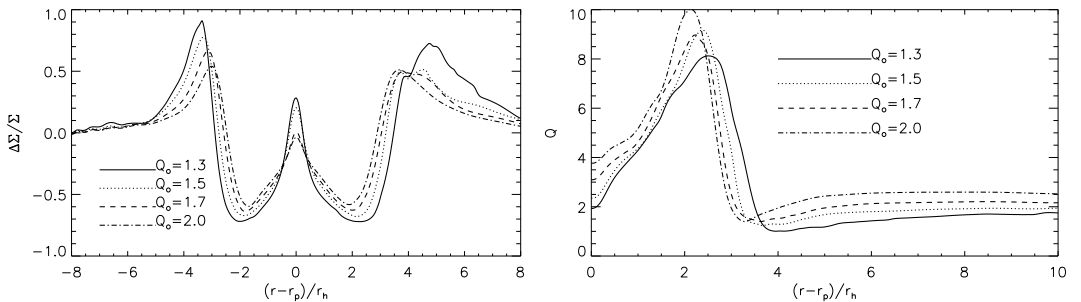


Figure 5.2: One dimensional gap profiles in terms of relative surface density perturbation (left) and the Toomre Q profile of the outer disc. The horizontal axis is the displacement from the planet in units of its Hill radius. The snapshots are taken at $t = 30P_0$.

5.2.1 Fragmentation of high mass discs

Giant planets can induce fragmentation in marginally stable discs. This phenomenon was previously identified in smooth particle hydrodynamic simulations by Armitage & Hansen (1999) and Lufkin et al. (2004). However, grid-based codes are generally better at resolving shocks and therefore planetary wakes. Fig. 5.1 confirms that planetary wakes may become gravitationally unstable as a clump forms along the outer wake in $Q_o = 1.3$.

Fig. 5.3 shows that fragmentation indeed occur for $Q_o = 1.3$ within $10P_0$ of releasing the planet, and there is no annular gap. This means that for very massive discs ($M_d \gtrsim$

$0.1M_*$), migration will be significantly affected by these clumps, rather than large-scale spirals associated with edge modes.

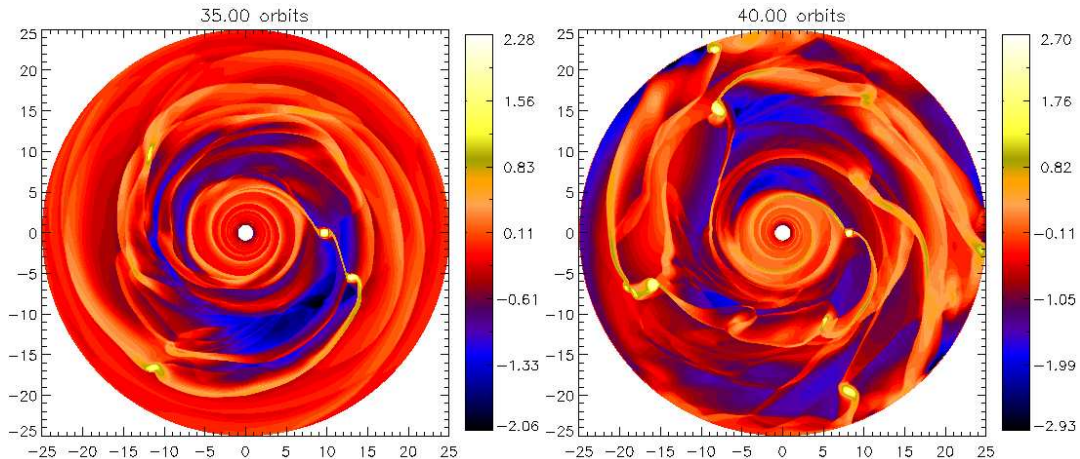


Figure 5.3: *Fragmentation induced by the planet in the most massive disc model $Q_o = 1.3$, $M_d/M_* = 0.092$. $\log[\Sigma/\Sigma(t=0)]$ is shown.*

5.3 Migration and disc structure

Edge modes lead to non-axisymmetric over-densities inside the gap. This is a potential source for torques to originate from within the planet’s coorbital region. Because r_c is expected near the outer edge of the coorbital region, edge mode over-densities may undergo horseshoe turns upon approaching the planet. Furthermore, fluid elements just inside the separatrix will traverse close to the planet when it executes the U-turn, and this may provide significant torque.

Fig. 5.4 shows the instantaneous orbital radius of the planet in non-fragmenting disc models. Migration is non-monotonic and can be inwards or outwards on short time-scales ($\sim P_0$). However, with increasing disc mass, outwards migration is favoured on timescales of a few 10’s of orbits. For $Q_o = 1.5$, r_p increases by 20% in only $70P_0$.

The behaviour of $r_p(t)$ correlates with the coherence of the gap structure. Fig. 5.5 shows that $Q_o = 2.0$ first develops vortex mode (with $m > 6$). Although a viscosity of $\nu = 10^{-5}$ suppressed vortex modes in previous studies (for $M_p \leq 0.001M_*$), increasing M_p has the same effect as lowering ν , hence vortex modes are not completely suppressed. $Q_o = 2.0$ does develop spiral arms later on, but these probably result from the vortices perturbing the disc, rather than the linear edge mode instability. The important feature with $Q_o = 2.0$ is that instability leaves the gap edge intact and identifiable, unlike more

5. PLANETARY MIGRATION IN THE PRESENCE OF EDGE MODES

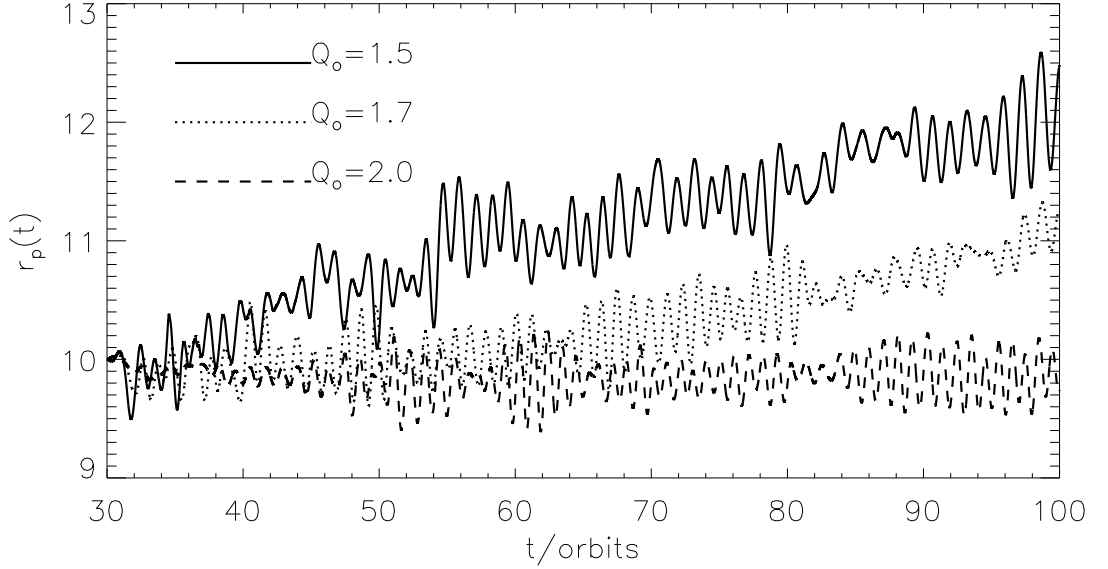


Figure 5.4: Migration of a 2-Jupiter mass planet ($M_p/M_* = 0.002$) in massive discs. The instantaneous orbital radius is shown.

massive discs below. Related to this is that Fig. 5.6 shows up to $t \leq 40P_0$, the torque magnitudes for $Q_o = 2.0$ are typically smaller than in that $Q_o = 1.5, 1.7$.

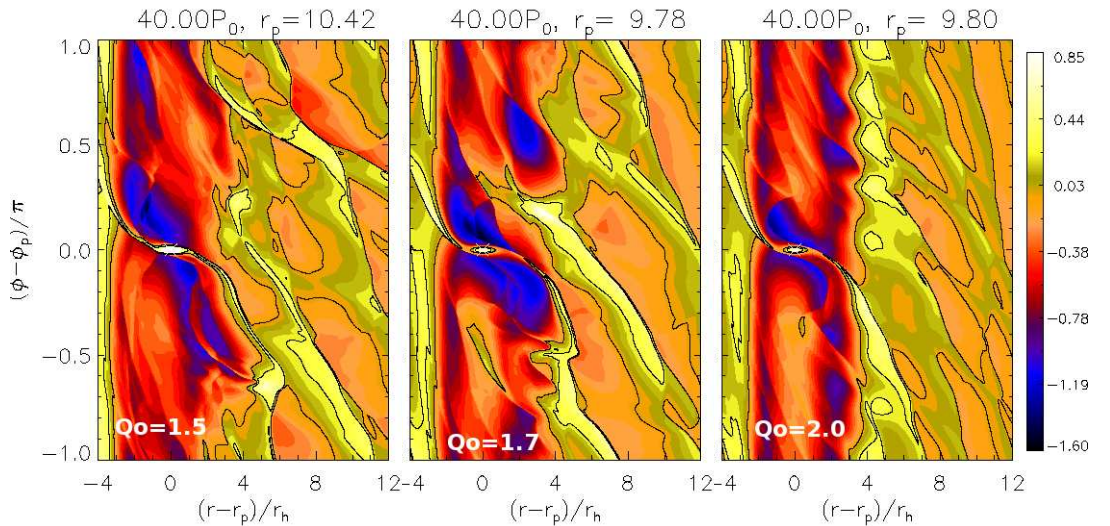


Figure 5.5: Relative surface density perturbation on a logarithmic scale, $\log[\Sigma/\Sigma(t=0)]$, showing gap instability.

$Q_o = 1.7$ develops a $m = 2-3$ edge mode, making the gap less well-defined. The

chosen snapshot is just prior to a small outwards jump in r_p for $Q_o = 1.7$ (Fig. 5.4) and Fig. 5.6 shows a clear positive disc-on-planet torque at this instant. Fig. 5.5 shows that the over-density inside the gap and associated with the edge mode, just upstream of the planet, could provide such positive torque. The outer gap edge in $Q_o = 2.0$ is not as disrupted and this results in no secular increase in r_p over the simulated timescale.

In Fig. 5.5, large-scale spirals can still be seen for $Q_o = 1.5$. Weak fragmentation occurs but collapse into clumps are not seen. Like $Q_o = 1.7$, the gap is unclear, but now with significant disruption to the outer gap edge. It is no longer a clear feature like in $Q_o = 2.0$.

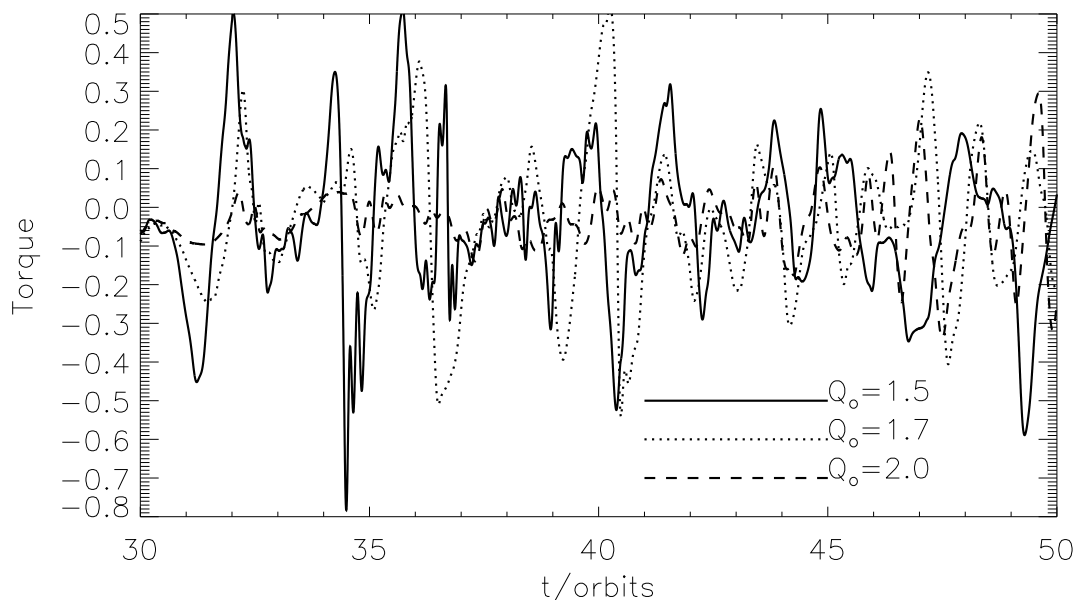


Figure 5.6: *Instantaneous disc-on-planet torques. Tapering has been applied to contributions from the Hill sphere to make this plot clearer. This procedure does not affect the torque’s rapid oscillatory behaviour, or its behaviour as a function of Q_o . Migration was evolved without tapering since the simulations are fully self-gravitating (Crida et al., 2009). The snapshots in Fig. 5.5 are taken at $t = 40P_0$.*

5.3.1 Gap evolution

The migration curves in Fig. 5.4 results from modification to the gap structure by instabilities. The following prescription was used to monitor gap structure. A one-dimensional profile of the relative surface density perturbation $\Delta\Sigma/\Sigma$ is first calculated (as in Fig. 5.2). The *outer gap edge* is defined as $r_{e,\text{out}} > r_p$ such that $\Delta\Sigma/\Sigma = 0$

5. PLANETARY MIGRATION IN THE PRESENCE OF EDGE MODES

there, and similarly for the *inner gap edge* $r_{e,\text{in}} < r_p$. The *outer gap depth* is $\Delta\Sigma/\Sigma$ averaged over $r \in [r_p, r_{e,\text{out}}]$. The *outer gap width*, in units of the Hill radius, is $w_{\text{out}} = |r_{e,\text{out}} - r_p|/r_h$ and the *inner gap width* is $w_{\text{in}} = |r_{e,\text{in}} - r_p|/r_h$. The *gap asymmetry* is $w_{\text{out}} - w_{\text{in}}$. Running time-averaged ($t \geq 40P_0$) plots are shown in Fig. 5.7

Edge modes have a gap-filling effect. The case $Q_o = 1.5$, in which the planet immediately migrates outwards, has the smallest gap depth throughout. $Q_o = 2.0$ has the deepest gap and is non-migrating over the simulation timescale. In $Q_o = 1.7$, the magnitude of gap depth decreases relative to that for $Q_o = 2.0$ after $t = 80P_0$. Note that this corresponds to outwards migration in $Q_o = 1.7$. These observations suggests that material brought into the gap by edge modes, is responsible for outwards migration.

Gap asymmetry was defined as the distance from the planet to the outer gap edge minus the distance between the planet and the inner gap edge. It follows that the more negative gap asymmetry is, the closer the planet is to the outer gap edge. Fig. 5.7 shows the planet is located closer to the outer gap edge in discs with lower Q_o than in discs with higher Q_o . Furthermore, the non-migrating $Q_o = 2.0$ reaches a constant asymmetry, whereas in the outwards-migrating cases asymmetry decreases with time.

The trend above suggests that the planet is moving material associated with the outer gap edge, which is unstable to edge modes, inwards via horseshoe turns (this would also be consistent with shallower gaps with decreasing Q_o). This provides a positive torque on the planet. If, on average, this effect dominates over sources of negative torques, e.g. negative Lindblad torques and the enhancement of which by the coincidence of an edge mode spiral arm with the outer planetary wake, then the planet should on average migrate outwards, i.e. closer to the outer gap edge.

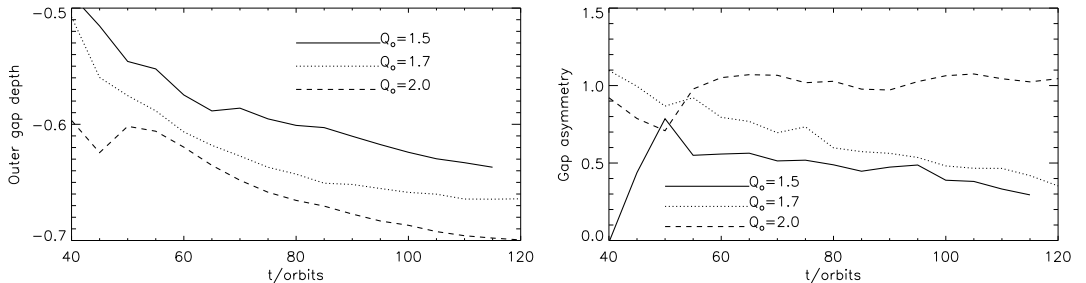


Figure 5.7: *Running-time averages of gap properties: the dimensionless outer gap depth (left) and the gap asymmetry in units of Hill radius (right). These quantities are defined in §5.3.1.*

5.4 A fiducial case in detail

Fig. 5.8 and Fig. 5.9 presents a summary of the simulation with $Q_o = 1.7$, in terms of the evolution of the planet's orbit and the disc's surface density perturbation. The orbit remains fairly circular ($e < 0.06$) with no secular change in eccentricity. However, the semi-major axis a increases more noticeably than orbital radius r_p .

Fig. 5.9 shows the outer disc ($r > r_p$) is unstable to edge modes whereas the inner disc ($r < r_p$) remains relatively stable. Large-scale spirals are maintained throughout the simulation. The gap is unclean with material brought into it by the edge modes. Notice in the $t = 105P_0$ and $t = 115P_0$ snapshots the over-density just upstream and exterior to the planet. It resides within the planet's coorbital region ($2.5r_h \gtrsim r - r_p > 0$) so this material is expected to execute inwards horseshoe turns and exert a positive torque on the planet. If this torque is sufficiently large, it may trigger type III migration because it promotes the planet to move closer to the higher density gap edge and interact with it. This can explain the faster-than-linear increase in a . In this case the edge mode over-densities provides the coorbital mass deficit.

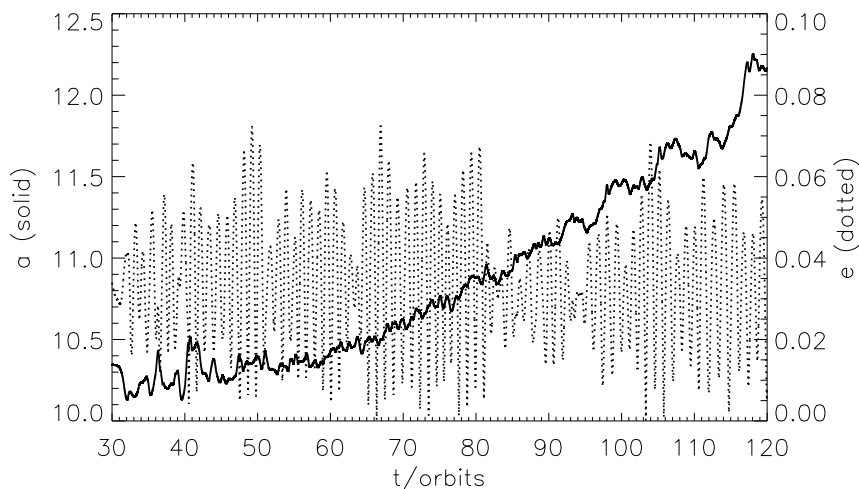


Figure 5.8: *Orbital evolution of the planet in the $Q_o = 1.7$ disc, in terms of Keplerian semi-major axis a (solid) and eccentricity e (dotted). These have been calculated without accounting for the disc potential.*

5.4.1 Torques and mode amplitudes

Fig. 5.10 shows the time-averaged evolution of non-axisymmetric modes in the outer disc, together with evolution of disc-on-planet torques. The effect of edge modes can

5. PLANETARY MIGRATION IN THE PRESENCE OF EDGE MODES

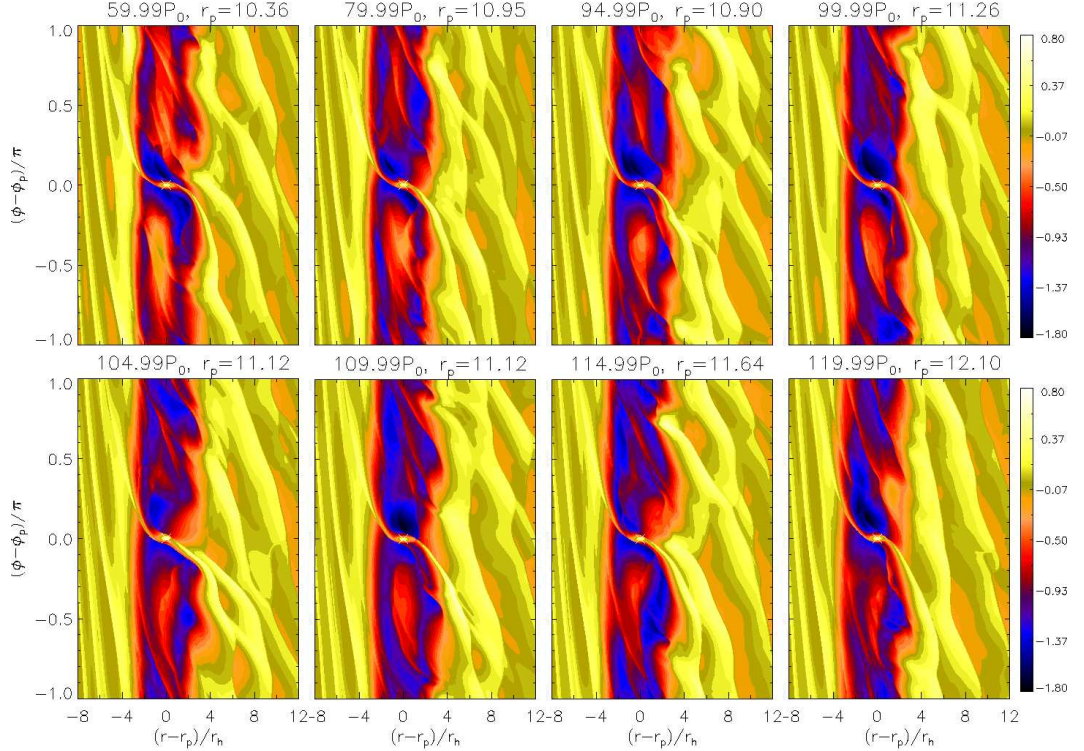


Figure 5.9: Evolution of the $Q_o = 1.7$ case. The logarithmic relative surface density perturbation $\log[\Sigma/\Sigma(t=0)]$ is shown.

be seen by comparison to the non-migrating case $Q_o = 2.0$. The modes have larger amplitudes in $Q_o = 1.7$, and the total torque becomes more positive with time¹. The figure shows this is attributed to torques from within the Hill radius, which is positive. By contrast, although $Q_o = 2.0$ also develop large-scale spirals later on, they are of smaller amplitude and do not make the Hill torque positive. The total time-averaged torque remains fairly constant.

Fig. 5.10, along with Fig. 5.9, confirms that large-scale spirals which extend into the gap can provide significant torques. Edge modes naturally fit this description since they are associated with vortensity maxima just inside the gap. Because the outer disc is more unstable, the corotation radius r_c of edge modes lie beyond r_p so its pattern speed is smaller than the planet's rotation. Thus, over-densities associated with corotation approach the planet from upstream, but because r_c is actually within the planet's coorbital region, the associated fluid elements are expected to execute

¹The total torque values are negative, despite overall outwards migration, because the running time-average is plotted. The average includes the time at planet release, when the disc-planet torque is large negative.

(inwards) horseshoe turns upon approaching the planet. This provides a positive torque on the planet, and if the edge mode amplitude is large enough, it may reverse the usual tendency for inwards migration.

The picture outlined above is consistent with the fact that edge modes have corotation at vortensity maxima. A giant planet can induce shocks very close to its orbital radius (Chapter 2) and vortensity is generated as fluid particles execute horseshoe turns across the shock. That is, the vortensity rings are associated with particles on horseshoe orbits. So when non-axisymmetric disturbances associated with vortensity rings develop (i.e. edge modes), the associated over-densities can be expected to execute horseshoe turns.

5.4.2 Torque distribution

The analysis above suggests that torques from within the gap are responsible for the gradual increase in r_p or a . Fig. 5.11 compares torque densities in the fiducial case to the non-migrating case. The torques have been averaged over $30P_0$ at two time intervals. For $t \in [50, 80]P_0$ radial plots for both cases show a positive torque at r_p . By $t \in [80, 110]P_0$ this torque has diminished for $Q_o = 2.0$. This is expected for Jovian planets as they open a clean gap, leaving little material near r_p to torque the planet. However, in $Q_o = 1.7$, this torque is maintained (and even slightly increased) by the edge modes because they bring material into the co-orbital region.

Azimuthal plots in Fig. 5.11 only account for material in between the gap edges (quantitatively defined in §5.3.1). Both cases have similar distributions for $t \in [50, 80]P_0$. For $t \in [80, 110]P_0$ though, the torque becomes symmetric about the planet's azimuth for $Q_o = 2.0$, whereas for $Q_o = 1.7$ asymmetry is maintained, leading to a positive torque. This torque is caused by over-density ahead in comparison to that behind the planet, i.e. material flowing radially inwards across the planet's orbital radius. However, unlike the single scattering events by vortices or spirals, which dominate most of the migration, here migration occurs on much longer timescales in comparison.

5.5 Summary and discussion

The numerical experiments above have shown that edge modes can induce outwards migration over timescales of a few tens of initial orbital periods. In the planet's frame, an edge mode spiral, with corotation at vortensity maximum exterior to r_p , approaches the planet from $\phi > \phi_p$. However, the mode extends into the planet's coorbital re-

5. PLANETARY MIGRATION IN THE PRESENCE OF EDGE MODES

gion, so there are associated over-densities that execute inwards horseshoe turns, which provides a positive torque.

The torque responsible for outwards migration can be identified as positive bump

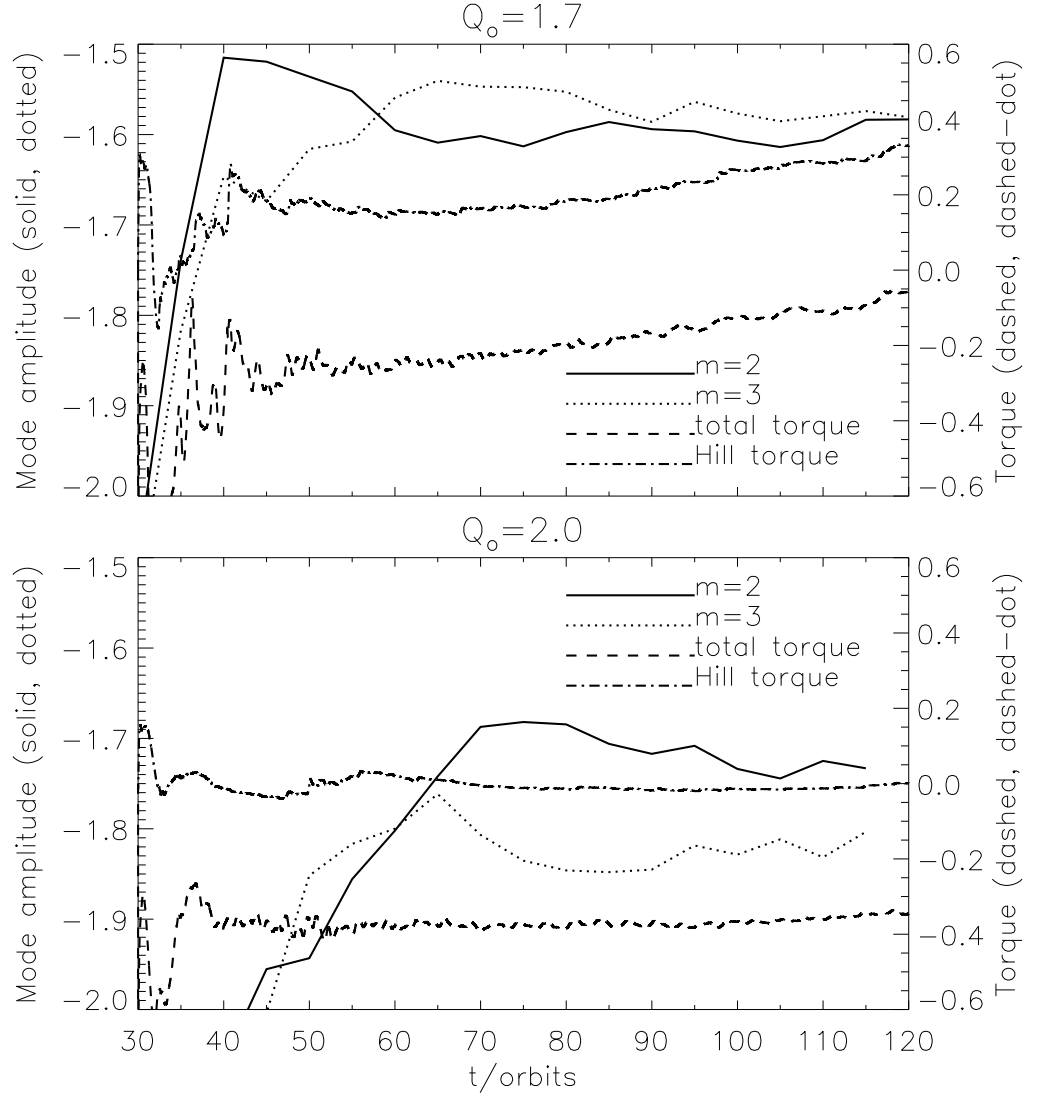


Figure 5.10: Evolution of the amplitude of the $m = 2$ (solid) and $m = 3$ (dotted) modes in surface density in the outer disc $r \in [r_p, r_o]$ in comparison to the total disc-on-planet torque (dashed) and the torque contribution from the Hill sphere (dashed-dot). All quantities are running-time averaged. Disc modes have been normalised by the axisymmetric amplitude and is plotted on a logarithmic scale. The upper plot is $Q_o = 1.7$, in which the planet migrates outwards. The lower plot is the non-migrating $Q_o = 2.0$ case.

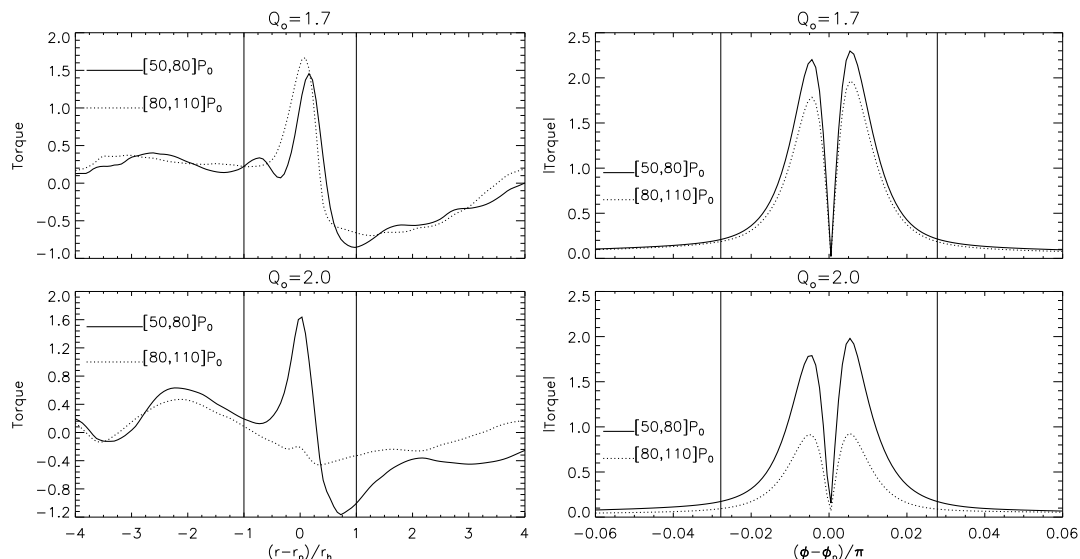


Figure 5.11: Torque densities. Radial plot (left) and azimuthal plot (right), averaged over $t \in [50, 80]P_0$ (solid) and over $t \in [80, 110]P_0$ (dotted). The upper plots are from the $Q_o = 1.7$ case, which display outwards migration and the lower plots are from the non-migrating $Q_o = 2.0$ case. Vertical lines in each plot indicate the Hill radius.

at $r = r_p$ in the time-averaged radial torque plots in Fig. 5.11, resulting from azimuthal gap asymmetry about the planet. This torque diminishes with time in the non-migrating $Q_o = 2.0$ case, thus migration itself helps to maintain this torque. Note that such torque was also absent in the Saturn-mass case when held on fixed orbit (Fig. 4.23). Another difference to Saturn is Jovian mass planets induce larger perturbations and the gap-filling effect of edge modes tend to be stronger.

Edge modes must be sustained in order to provide this torque, this in turn requires the existence of vortensity maximum, despite the development of edge modes destroying them. However, this is easily regenerated by giant planets because they induce shocks which act as a source of vortensity. Furthermore, since edge modes also require gravitational coupling to an exterior smooth disc, outwards migration of this type is probably more likely to occur than inwards, because typically discs become more self-gravitating with increasing radius.

Planetary migration observed here closely resembles classic type III migration (Masset & Papaloizou, 2003) because it relies on torques from the coorbital region and the accelerated increase in semi-major axis indicate positive feedback. The main difference to type III migration originally described by Masset & Papaloizou is that the flow across the planet's orbit is non-smooth, because edge modes present distinct fluid blobs,

5. PLANETARY MIGRATION IN THE PRESENCE OF EDGE MODES

rather than a continuous flow across r_p . Also, where type III migration usually apply to partial gaps and therefore Saturn-mass planets, the Jovian mass planet used here reside in a much deeper gap. Radial mass flux across the planet is still possible, despite the gap-opening effect of Jovian planets, because the unstable edge modes brings over-densities inside the gap to interact with the planet.

A consequence of the above is that formation of a clean gap is expected to become increasingly difficult in massive discs. Planetary gaps correspond to existence of vortensity maximum, but these become unstable to edge modes in massive discs and instability tends to fill the gap. In very massive discs giant planets induce fragmentation, their planetary wakes become gravitationally unstable and no annular gap forms. Therefore classic type II migration, where the planet moves smoothly with the viscous gap evolution, probably does not apply in massive discs.

5.5.1 Outstanding issues

It should be noted that the adopted disc model is biased towards outwards migration because the Toomre Q decreases outwards (approximately like $r^{-1/2}$) so edge modes associated with the outer disc develops preferentially. If a disc model such that the inner disc is more unstable was used, then inwards migration should occur. And if independent edge modes of comparable amplitude develop on either side of the gap, there should be little migration overall.

Significant torques originate from material close to the planet, so numerical treatment of the Hill sphere is a potential issue. This concern also applies to standard type III migration, but note that the adopted equation of state was originally designed for numerical studies of type III migration (Pepliński et al., 2008a). Furthermore, fully self-gravitating discs are suited for type III migration (Crida et al., 2009). Lower resolution simulations have been performed which also resulted in outwards migration.

This EOS mimics heating near the planet but is not a quantitative model for the true thermodynamics. If this EOS under-estimates the heating, then the positive torque could be over-estimated and vice versa. On the other hand, edge mode corotation reside outside the Hill radius but still inside the gap. So the numerical treatment of the Hill radius does not affect existence of edge modes and their associated over-densities should affect torques originating from the coorbital region in the way described in this Chapter.

Chapter 6

Three-dimensional simulations

Razor-thin discs have been used in all the previous numerical simulations. This approximation can be justified for gap-opening planets, whose definition can be taken as $M_p/M_* \gtrsim h^3$, so that the planet's Hill radius exceeds the local disc thickness. Nevertheless, three-dimensional discs should be considered because instabilities studied above are associated with radial structures with characteristic length scales comparable to the local thickness.

In this Chapter, numerical simulations are developed to confirm properties of gap stability studied previously in 2D, persist in 3D geometry. The 3D disc-planet models are presented in §6.1. §6.2 describes adaptation of the **ZEUS-MP** software to perform disc-planet simulations with and without disc self-gravity. Three-dimensional counterparts to previous 2D results are presented in §6.3, including vortex formation and vortex-induced migration in non-self-gravitating discs, vortex modes in self-gravitating discs and edge modes in massive discs. §6.4 concludes this Chapter.

6.1 Three-dimensional models

The governing equations for three-dimensional discs are listed in Chapter 1. The disc model is extended from that used in Chapter 3—5 by introducing vertical structure. The disc is described in spherical polar co-ordinates (r, θ, ϕ) with origin on the central star, r is now the spherical polar radius, and the cylindrical radius denoted by $R = r \sin \theta$ and the cylindrical $z = r \cos \theta$.

The disc occupies $r \in [r_i, r_o]$, $\theta \in [\theta_{\min}, \theta_{\max}]$ and $\phi \in [0, 2\pi]$. The vertical extent is such that $|\pi/2 - \theta_{\min}|$ corresponds to n_H number of scale-heights. The initial density

6. THREE-DIMENSIONAL SIMULATIONS

profile is

$$\begin{aligned}\rho_0(R, z) &= \frac{\Sigma_0 R^{-p}}{\sqrt{2\pi H^2}} \exp\left(-\frac{\Phi_*}{c_s^2} - \frac{1}{h^2}\right) \times \left[1 - \sqrt{\frac{r_i}{R + hr_i}}\right] \\ &\equiv f(R) \exp(-\Phi_*/c_s^2),\end{aligned}\tag{6.1}$$

where $H = hR$ is the scale-height with $h = 0.05$ the fixed aspect-ratio, Σ_0 is a constant parameterising the density scale. The power-law index in surface density is fixed at $p = 3/2$ for all the simulations performed. Recall Φ_* is the stellar potential. Eq. 6.1 results from exact vertical hydrostatic equilibrium assuming an isothermal equation of state with sound-speed $c_s = c_s(R)$. As before, the density scale Σ_0 is chosen via the Keplerian Toomre parameter Q_o at the outer boundary:

$$Q_o \equiv \frac{c_s \Omega_k}{\pi G \int_{z_{\min}}^{z_{\max}} \rho_0 dz} \Big|_{R=r_o}.\tag{6.2}$$

Note that vertical integration in Eq. 6.2 corresponds to the computational domain so exterior mass is either zero or assumed negligible. Thus models with fixed Q_o but increasing n_H will have decreasing midplane density scale represented by Σ_0 .

The disc is initialised without meridional velocity ($u_r = u_\theta = 0$) and azimuthal velocity given by centrifugal balance between stellar potential and pressure:

$$u_\phi^2 = R \frac{\partial c_s^2}{\partial R} + \Phi_* \frac{\partial \ln c_s^2}{\partial \ln R} + R c_s^2 \frac{\partial \ln f}{\partial R}.\tag{6.3}$$

Exact balance implies $\partial u_\phi / \partial z \neq 0$, but since the disc is thin ($h \ll 1$), the azimuthal velocity is well approximated to be Keplerian ($u_\phi \simeq R\Omega_k$).

6.1.1 Self-gravity

The three-dimensional disc potential Φ is obtained by solving the differential Poisson equation $\nabla^2 \Phi = 4\pi G \rho$. This removes the need for a softening length used for 2D self-gravity. However, boundary conditions must be explicitly imposed and will be described in §6.2.

Instead of solving for the initial vertical structure consistent with self-gravity, a simpler numerical approach is taken. All calculations are initialised with the non-self-gravitating equilibrium described above. To account for radial self-gravity, the initial azimuthal velocity is reset to $v_\phi^2 \rightarrow v_\phi^2 + rg_r$. The vertical self-gravity force, g_θ , is initially set to zero and then smoothly increased to its full value over time.

6.1.2 Planet treatment and equation of state

The planet potential is softened by a fraction of its Hill radius

$$\Phi_p = -\frac{GM_p}{\sqrt{|\mathbf{r} - \mathbf{r}_p|^2 + (\epsilon_{p0}r_h)^2}}. \quad (6.4)$$

The softening is set to $\epsilon_{p0} = 0.1$. In 2D, softening by a fraction of the local scale-height approximately accounts for vertical structure, whereas softening in 3D is purely to avoid numerical divergence. The Hill radius is calculated with $|\mathbf{r}_p|$, but this is very close to the planet's cylindrical radius from the star because the orbit inclination is negligible for all the migrating cases presented later. The planet is introduced on fixed circular orbit of radius r_p in the midplane with Keplerian rotation speed and its mass is smoothly switched on.

The disc-on-planet forces are tapered by multiplying all components of the force by the factor $1 - \exp\left[-\frac{|\mathbf{r}-\mathbf{r}_p|^2}{2(\epsilon_c r_h)^2}\right]$. The cut-off parameter ϵ_c is set to unity unless otherwise stated. Tapering is applied whether or not disc self-gravity is enabled. Without self-gravity, the physical motivation for tapering is that material close to the planet should be gravitationally bound to it and should be excluded in the disc-on-planet force. There is no physical need for tapering in a self-gravitating disc, but limited numerical resolution in a 3D calculation prevents accurate determination of torques from close to the planet. Thus, the tapering is largely motivated by numerics.

The equation of state is isothermal, so the pressure is $p = c_s^2 \rho$. Before introducing the planet, $c_s^2 = h^2 GM_*/R$ is the locally isothermal sound-speed giving the standard Gaussian atmosphere. After the planet is introduced, the sound-speed is modified to c_s given by Eq. 1.18 which heats up the disc near the planet relative to the $1/R$ profile. This prescription is used to limit mass pile-up near the planet and reduce numerical errors resulting from this effect. It also helps to reduce the increase in effective planet mass and therefore the effect of self-gravity on instabilities through the background planetary gap.

6.2 Adapting ZEUS-MP

ZEUS is a general-purpose astrophysical fluid dynamics code based on finite-difference schemes. Several versions are publicly available¹, including ZEUS-2D (Stone & Norman, 1992a,b; Stone et al., 1992) which FARGO is based on, and its three-dimensional version

¹<http://lca.ucsd.edu/portal/software>

6. THREE-DIMENSIONAL SIMULATIONS

ZEUS-3D. The version adapted here is ZEUS-MP , developed and described by Hayes et al. (2006), which places emphasis on parallel computing. ZEUS-MP is capable of solving hydrodynamics, self-gravity, magnetohydrodynamics and radiative transfer, although the latter two modules are disabled in this study.

6.2.1 Source terms

The most important new source term is the gravitational forces on disc fluid due to the planet potential. The acceleration $-\nabla\Phi_p$ is added directly to the momentum equations to update each velocity component (u_r, u_θ, u_ϕ) as part of the source term step in ZEUS-MP . Indirect disc and planet potentials are also added, to account for the non-inertial reference frame.

All simulations presented have zero physical viscosity (artificial viscosities is employed by ZEUS-MP to treat shocks). However, a module for viscous forces has also been added to ZEUS-MP . Velocity components can be updated by the full viscous stress tensor immediately after the main source step. The magnitude of the viscous force is characterised by kinematic viscosity ν , which can be a constant or $\nu = \alpha c_s H$, where α is a constant. The effect of viscosity on vortex and edge modes in 3D is not expected to differ to that in 2D discs, which has already been investigated.

6.2.2 Disc potential

ZEUS-MP was chosen for its ability to solve 3D self-gravity in spherical polar co-ordinates with parallelisation. This is an uncommon feature among public codes. The discretised Poisson equation is a linear system, $A\mathbf{x} = \mathbf{b}$. ZEUS-MP solves this using the conjugate gradient (CG) method (see Hayes et al., 2006, for details). It is an iterative scheme that minimises the quadratic form $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T A\mathbf{x} - \mathbf{b}^T \mathbf{x}$. At the k^{th} iteration, the CG method finds the scalar c_{k+1} and conjugate direction \mathbf{p}_{k+1} (such that $\mathbf{p}_i^T A\mathbf{p}_j = 0$ if $i \neq j$) to give the improved solution as $\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + c_{k+1}\mathbf{p}_{k+1}$ (e.g. Press et al., 1992; Barrett et al., 1994).

6.2.2.1 Boundary potential

A potential solver based on Boss (1980) has been added to ZEUS-MP to obtain potentials at r - and θ -boundaries, which are set as Dirichlet boundary conditions for the CG solver. The potential is periodic in ϕ . The potential and density are expanded in spherical

harmonics, so that

$$\Phi = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} \Phi_{lm}(r) Y_{lm}(\theta, \phi); \quad \rho = \sum_{l=0}^{\infty} \sum_{m=-l}^{m=l} \rho_{lm}(r) Y_{lm}(\theta, \phi). \quad (6.5)$$

In practice the sum is truncated at $l = l_{\max}$ and $m = m_{\max}$. To further reduce computation cost, only $l + m = \text{even}$ modes are used, corresponding to symmetry with respect to the disc midplane. ρ_{lm} is found by direct integration:

$$\rho_{lm} = \int_0^{2\pi} \int_{\theta_{\min}}^{\theta_{\max}} \rho Y_{lm}^* \sin \theta d\phi d\theta. \quad (6.6)$$

This integration is parallelised by domain decomposition along with the hydrodynamics. Each processor holds a copy of the global $\rho_{lm}(r)$ but performs the integration over its local sub-domain of the global disc. The result is written to the appropriate sub-portion of the global $\rho_{lm}(r)$. A global reduction is then performed, summing contributions to $\rho_{lm}(r)$ from all angles that is held on different processors.

After expansion, the Poisson equation becomes

$$\frac{d^2 \Phi_{lm}}{ds^2} + \frac{d\Phi_{lm}}{ds} - l(l+1)\Phi_{lm} = 4\pi G \rho_{lm} r^2, \quad (6.7)$$

where $s \equiv \ln r$, and radial boundary conditions

$$\left[\frac{d}{ds} + (l+1) \right] \Phi_{lm} \Big|_{r_o} = 0; \quad \left[\frac{d}{ds} - l \right] \Phi_{lm} \Big|_{r_i} = 0 \quad (6.8)$$

are applied. The discretised version of Eq. 6.7 is a tridiagonal matrix equation. It can be solved relatively quickly with the LAPACK software package to yield $\Phi_{lm}(r)$. The potential at r and θ boundaries can be found from Eq. 6.5. Note that the solution method outline here assumes no mass outside the computational domain.

6.2.3 Planetary migration

Planetary migration is mostly neglected in order to focus on the issue of gap stability. However, a fifth order Runge-Kutta integrator (Press et al., 1992) for the equation of motion of the planet has been implemented in ZEUS-MP. During one hydrodynamic time step Δt , the planet's position and velocity is integrated from $t \rightarrow t + \Delta t$ with adaptive sub-steps but using the density field at t . If only the upper disc ($z > 0$) is simulated, then the vertical force is set to zero and the planet remains in the midplane.

6.3 Gap stability in 3D

The substantial cost of a 3D numerical calculation prevent extensive parameter studies. Instead, described below are specific simulations demonstrating gap instability. The general numerical setup is as follows.

The disc is divided into $N_r \times N_\theta \times N_\phi$ zones in the (r, θ, ϕ) directions, uniformly spaced in (θ, ϕ) and logarithmically spaced in polar r . During $t \leq t_0 \equiv 10P_0$ the disc is evolved without the planet. Between $t_0 \leq t \leq 2t_0$, the planet mass is switched on. If the disc is self-gravitating then for $0 \leq t \leq t_0$, $(l_{\max}, m_{\max}) = (48, 0)$ is set for the boundary potential and g_θ is smoothly switched on. For $t > t_0$, $(l_{\max}, m_{\max}) = (16, 10)$ is set for the boundary potential. Hydrodynamic boundary conditions are outflow in radius, reflecting in θ and periodic in ϕ .

6.3.1 Vortex instability in a non-self-gravitating disc

In this simulation the disc model is $Q_o = 4$. The domain is $r \in [1, 10]$ and $n_H = 2$ scale-heights above and below the midplane. The resolution is $(N_r, N_\theta, N_\phi) = (256, 64, 696)$, corresponding to 16 cells per H in θ , but only $\simeq 5.6$ cells per H in r and ϕ . The planet mass is $M_p = 3 \times 10^{-4} M_*$ and is fixed in the midplane at radius $r_p = 5$ after its introduction. The EOS parameter is $h_p = 0.7$.

Fig. 6.1 show density perturbations relative to the azimuthally-averaged density field at $t = 10P_0$. Two cuts in θ are shown, corresponding to the midplane and at $z = 2H$. The vortex instability clearly persists for planetary gaps in a three-dimensional disc. The relative density perturbation has no vertical dependence (except near the planet since it is embedded). This is also seen in the three-dimensional plot in Fig. 6.2. Since the background density is essentially Gaussian in z , the vortex density field will also be Gaussian vertically.

The evolution is very similar to 2D discs simulated in Chapter 2–3. The only noticeable difference is that in 2D without self-gravity, 3 vortices develop initially then merge (Fig. 3.7). In the 3D run, one vortex develops from the beginning (although a weak, second vortex can be identified in Fig. 6.1 at $t = 60P_0$). Reduction in the number of vortices from 2D to 3D was also noted by Meheut et al. (2010). However, in quasi-steady state the vortex in 3D and 2D are essentially identical. These results show that the vortex instability is a 2D instability. Thus, the condition for instability can be largely determined by vertically integrated properties of the disc.

The vortex vertical structure is shown in Fig. 6.3. The vortex lies just outside the outer gap edge at $R \simeq 5.6$. Its associated relative density and vortensity pertur-

6.3 Gap stability in 3D

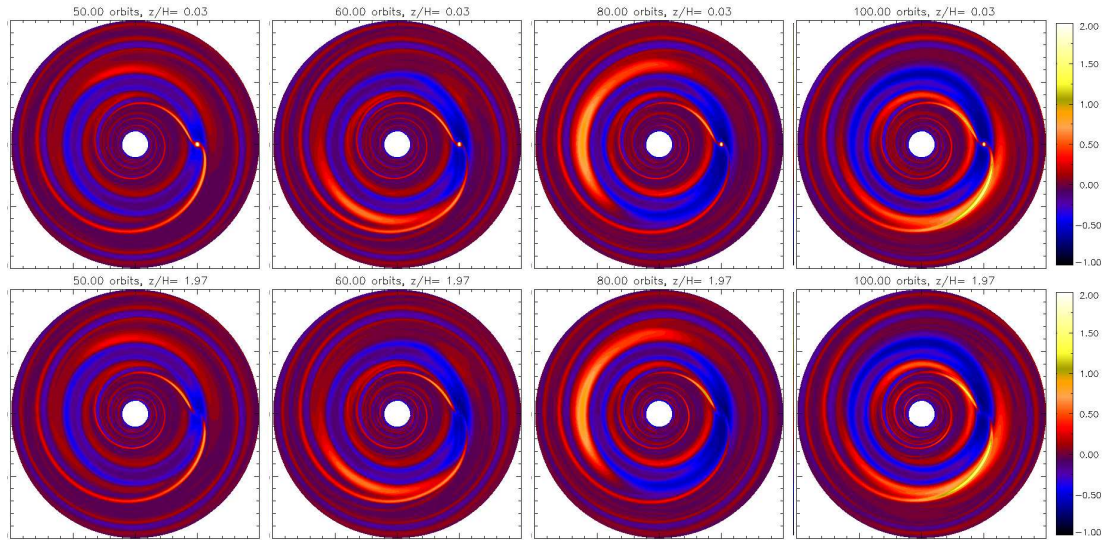


Figure 6.1: *Relative density perturbation near the midplane (top) and near $z = 2H$ (bottom).*

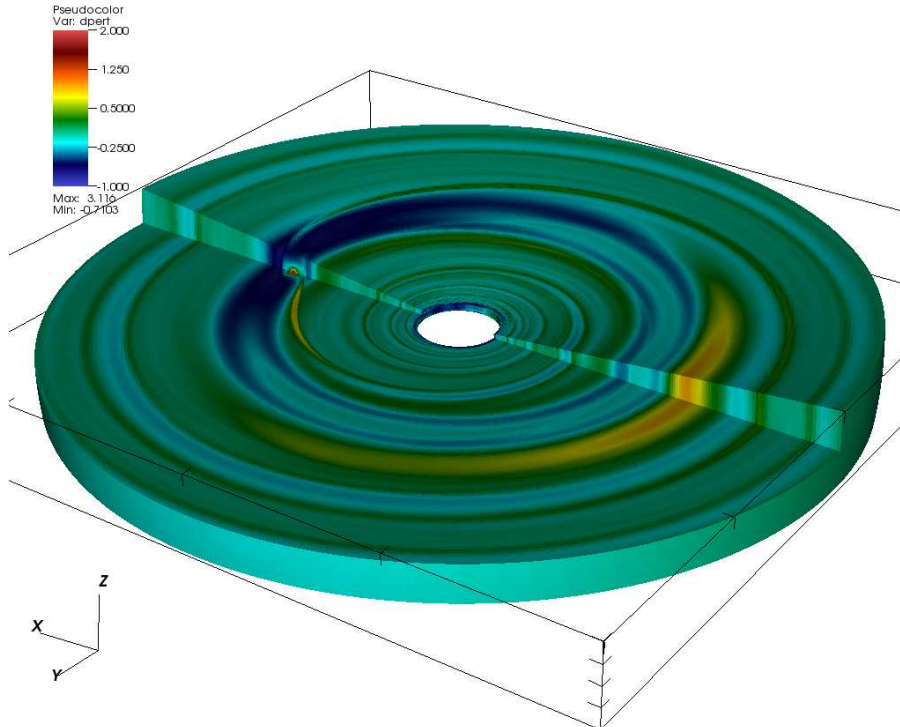


Figure 6.2: *Three-dimensional visualisation of the vortex instability at $t = 80P_0$. The colour bar is relative density perturbation.*

6. THREE-DIMENSIONAL SIMULATIONS

bations have very weak vertical dependence, despite the background being Gaussian. Stratification has no effect on the instability.

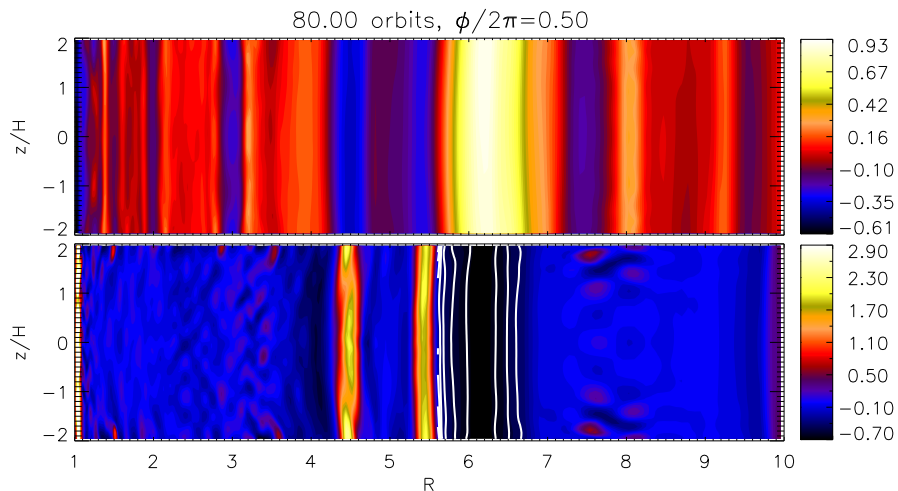


Figure 6.3: *Relative perturbation of density (top) and vertical component of vortensity η_z (bottom) in a meridional plane through the vortex at $R \in [5.6, 6.7]$ (white contour lines).*

The mass flux $\rho \mathbf{u}$ of the main vortex at $t = 80P_0$ is drawn in Fig. 6.4 for several meridional planes. As expected for an anticyclonic vortex, the radial flow transits from inwards to outwards as one traverses anti-clockwise (increasing ϕ) through the vortex. The transition point is the vortex centroid. The $\phi/2\pi = 0.45$ slice shows that the transition involves vertical flux, but its magnitude is small even near the midplane, where the mass density is largest. This means the vortex centroid has very little motion. There is no evidence for strong meridional circulation found by Meheut et al. (2010), but the flow towards the midplane highlighted in Fig. 6.4 suggest vertical motion near the vortex centroid, as seen in Meheut et al.. Note that symmetry about $z = 0$ is broken, so mass flux can overshoot across the midplane.

6.3.2 Vortex-planet interaction

Fig. 6.5 shows the planet orbital radius after it is released at $t = 100P_0$ for the simulation in §6.3. Also shown is a second simulation where only $z > 0$ was simulated and the planet released at $t = 50P_0$ (before vortices fully develop). Results are very similar to 2D, the migration composing of slow and rapid phases, the latter induced by the vortex associated with the inner gap edge. Since the interaction is gravitational,

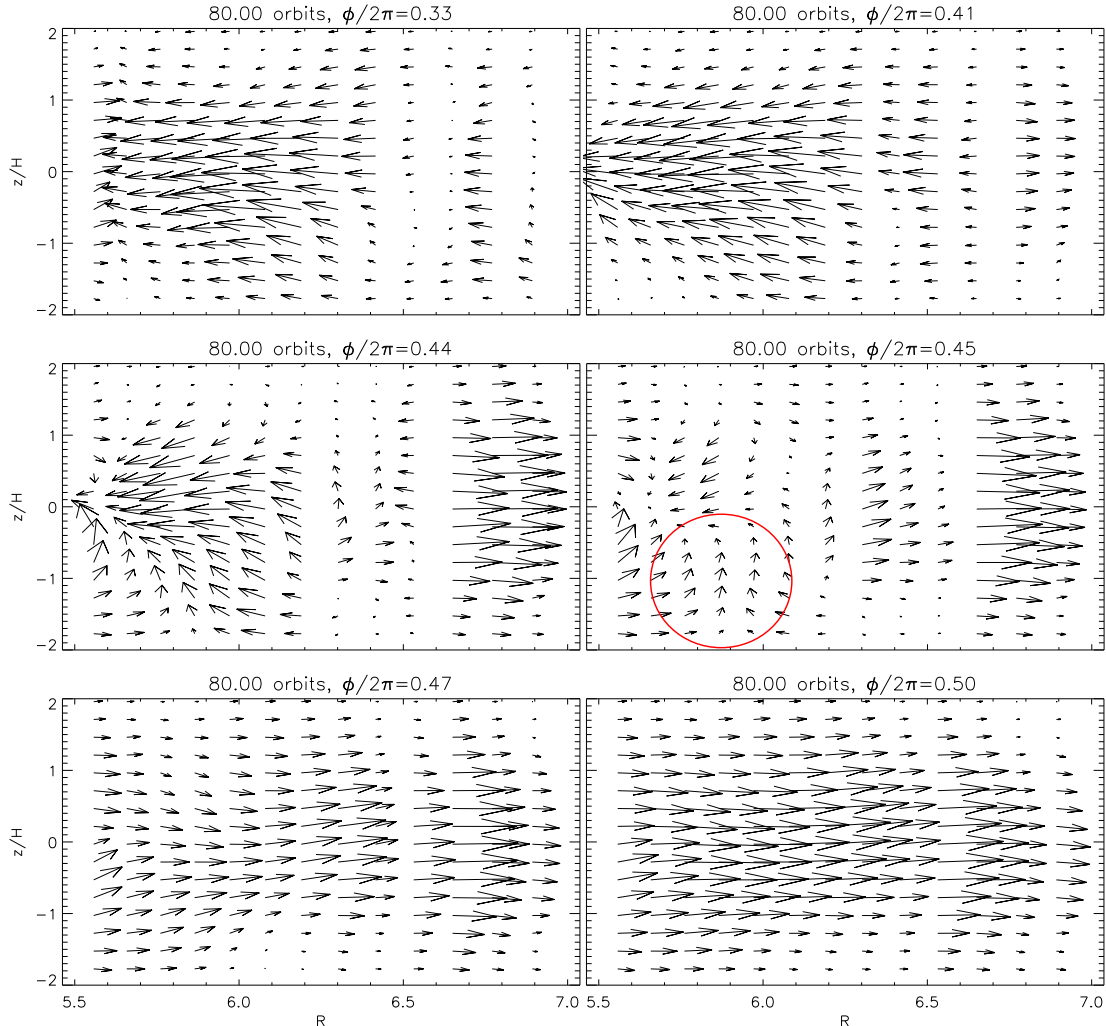


Figure 6.4: Meridional flow in vertical planes through the vortex. Velocity vectors have been scaled by the local density. The vertical velocity has also been divided by H . The circled region highlights flux towards the midplane near the vortex centroid.

and the perturbed density field associated with the vortex has no vertical structure, it is not surprising that the same interaction occurs in 3D as in 2D.

6.3.3 Vortex modes with self-gravity

For 3D self-gravitating simulations, the computational domain is $r \in [1, 25]$ and $0 \leq z \leq 2H$. For simulations in this subsection the resolution is $(N_r, N_\theta, N_\phi) = (256, 32, 512)$. Jovian planets are used to open the gap because the steeper edges imply higher growth rates than gaps opened by a Saturn-mass planet. This shortens the simulation time.

6. THREE-DIMENSIONAL SIMULATIONS

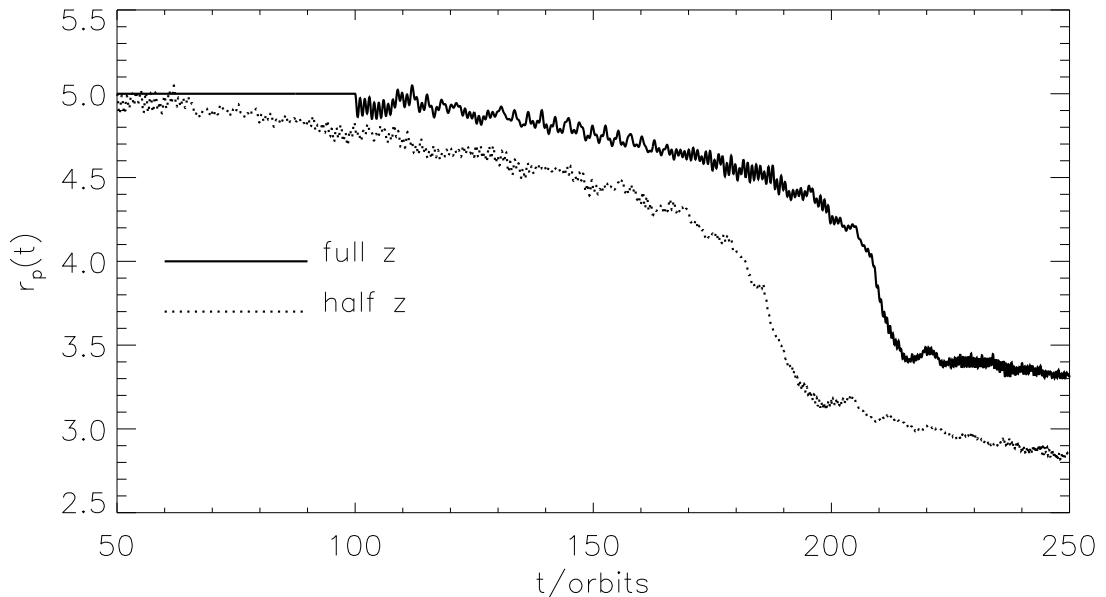


Figure 6.5: *Vortex-induced migration of a Saturn-mass planet in 3D discs. The run ‘full z’ simulates the entire disc and ‘half z’ only simulates the upper half-space ($z > 0$).*

The planet is fixed at $r_p = 10$ and simulated up to $t = 50P_0$. The EOS parameter is $h_p = 0.5$.

Fig. 6.6 compares vortex formation in the $Q_o = 3, 4, 8$ disc models, with $M_p = 10^{-3}M_*$. It confirms that the effect of self-gravity on the vortex instability discussed in Chapter 3, persists in 3D. Comparing the outer gap edge at $t = 30P_0$ shows self-gravity is stabilising, because in $Q_o = 8$ vortices can already be identified but not so readily seen in $Q_o = 4$. At $t = 40P_0$, more vortices develop with lowering Q_o because stabilisation is more effective for low m , thus higher m vortex modes are favoured, just as in 2D. The 6-vortex and 5-vortex configuration remains in $Q_o = 3, 4$, respectively, at $t = 50P_0$ whereas only 3 remain in $Q_o = 8$ (reduced from 4). This is due to resisted merging as vortices behave like co-orbital planets with increasing self-gravity. Again, this was seen in 2D.

The vertical structure of a vortex in the $Q_o = 4$ disc is shown in Fig. 6.7. Because the initial disc is not in self-gravitating equilibrium, the density perturbation relative to $t = 0$ and to $t = 10P_0$ (just after g_θ is fully switched on) are both shown. The increasing perturbation towards the midplane is partly due to vertical self-gravity increasing the background midplane density. Between $t = 10P_0$ and $t = 50P_0$ there may also be

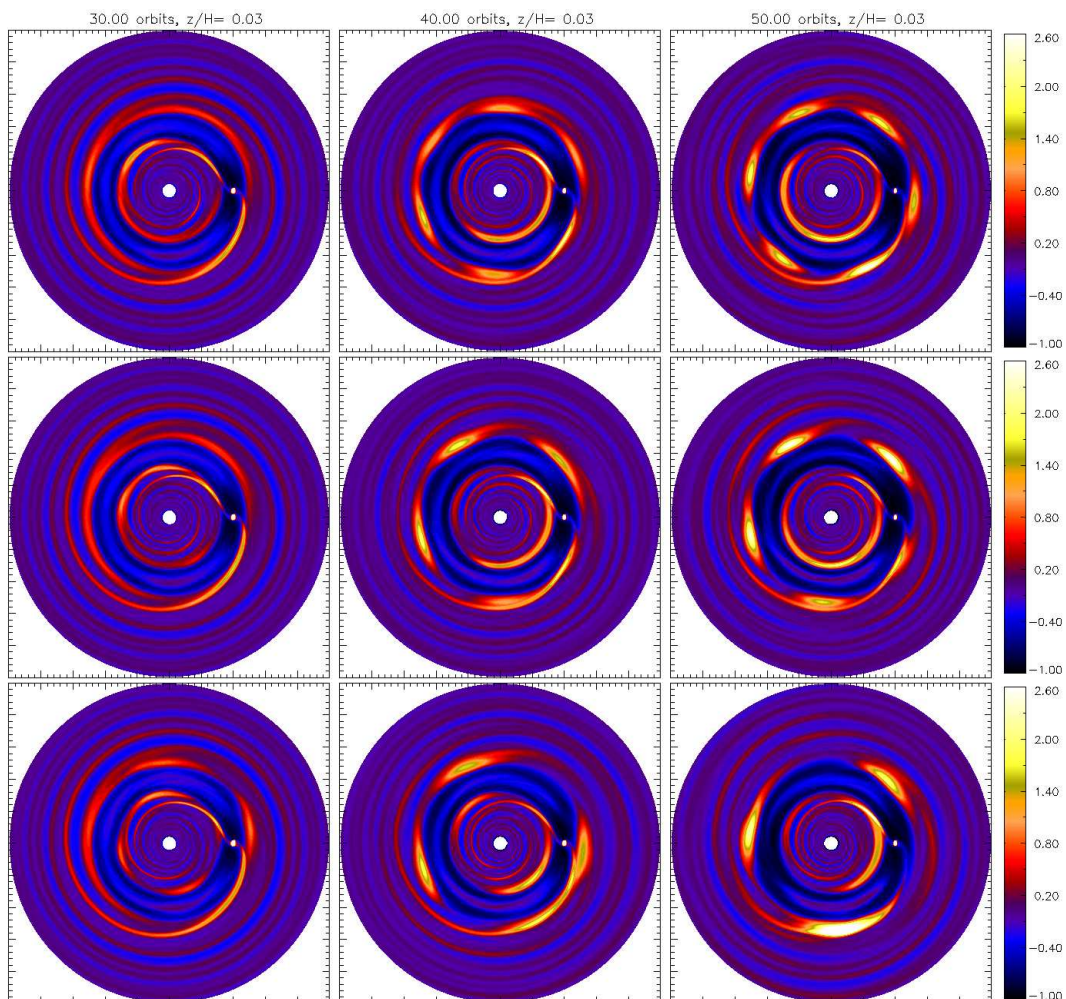


Figure 6.6: *Vortex instability in self-gravitating 3D disc models $Q_o = 3$ (top), $Q_o = 4$ (middle) and $Q_o = 8$ (bottom). The relative density perturbation near the midplane is shown.*

additional stratification due to vertical self-gravity of the vortex, since the midplane density is largest. Stratification is reduced if perturbation is taken relative to $t = 10P_0$, so the instability indeed tends to involve the entire column of fluid.

The vortensity field in Fig. 6.7 shows some disturbance along the contour lines at the vortex boundary, as well as closed contours within the vortex, which is associated with anti-clockwise motion indicated by mass flux vectors in Fig. 6.8. Meheut et al. (2010) found similar flow to occur in pairs — clockwise ($r < r_c$) and anti-clockwise ($r > r_c$), where r_c denotes the vortex centroid. Such symmetry may not be expected here since there is a planet located on one side of the gap edge. It is also likely

6. THREE-DIMENSIONAL SIMULATIONS

that the flow here has been modified by self-gravity of the vortex and/or background. However, despite additional vertical structure and flow, the vortensity perturbation remains largely columnar.

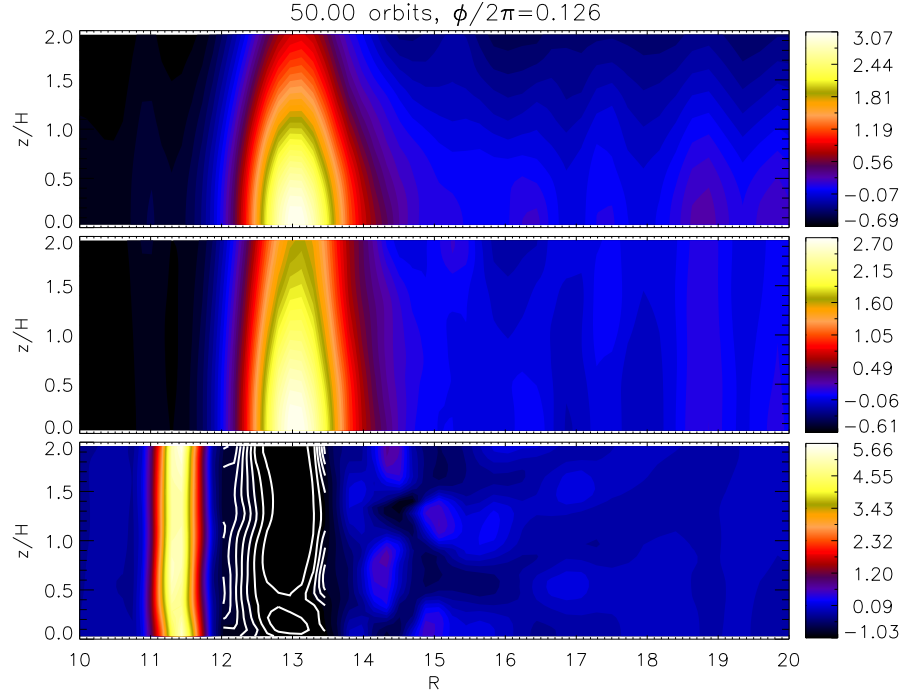


Figure 6.7: Density perturbation relative to $t = 0$, $10P_0$ (top, middle) and perturbation to the vertical component of vortensity relative to $t = 10P_0$ (bottom). The disc model is $Q_o = 4$ with self-gravity.

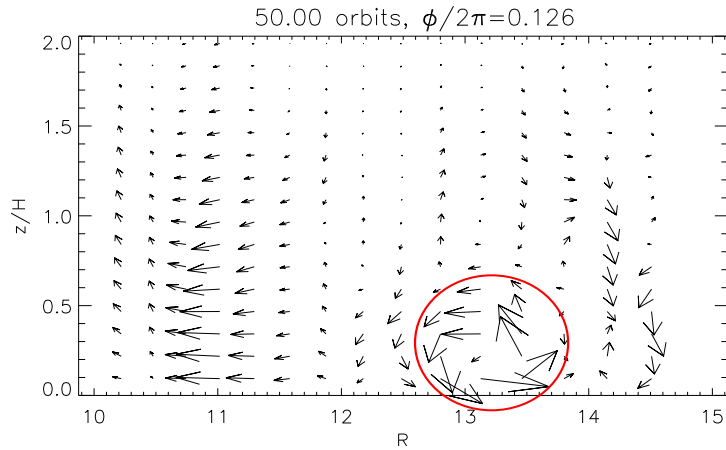


Figure 6.8: Mass flux corresponding to Fig. 6.7.

6.3.4 Edge modes

The disc model $Q_o = 1.5$ is used to demonstrate edge modes. The resolution is increased to $(N_r, N_\theta, N_\phi) = (256, 64, 516)$, on account of reduction in effective scale-height due to vertical self-gravity.

Fig. 6.9 shows the development of an $m = 2$ edge mode and Fig. 6.10 is a three-dimensional visualisation. Perturbations are calculated relative to $t = 10P_0$. The behaviour near the midplane is similar to 2D simulations, but the edge mode amplitude decreases noticeably away from the midplane. The background density ($t = 10P_0$) near $z = 2H$ has been reduced by vertical self-gravity from $t = 0$. This means that perturbations high in the atmosphere appear enhanced, but their actual densities are small compared to the midplane.

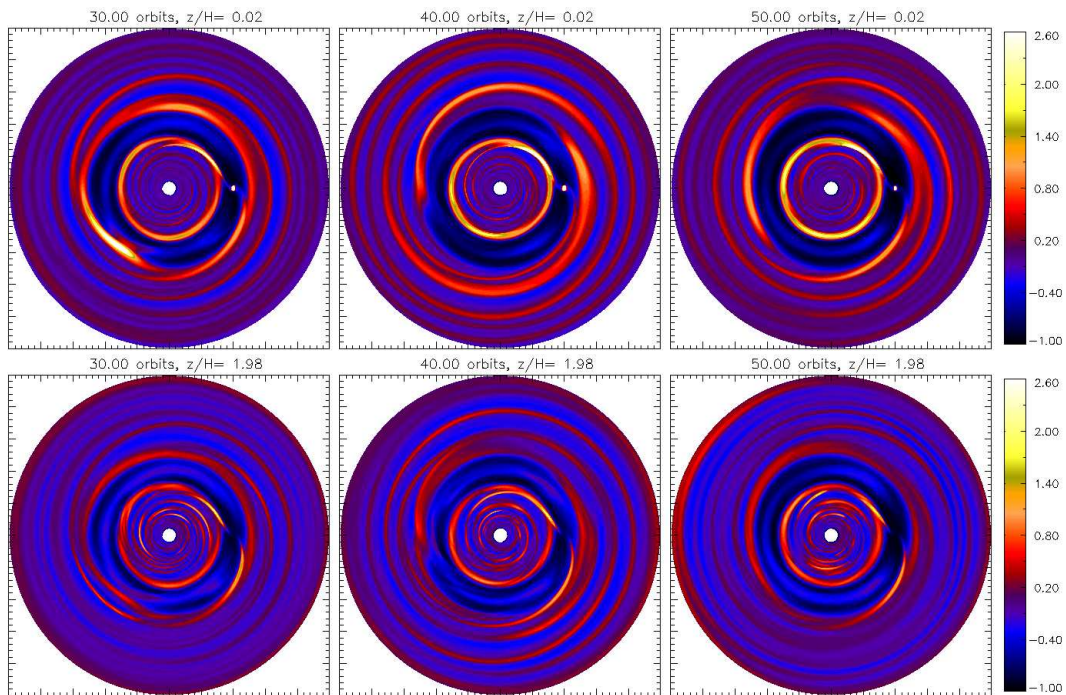


Figure 6.9: Edge mode in the 3D disc model $Q_o = 1.5$. The density perturbation relative to $t = 10P_0$ is shown near the midplane (top) and $z = 2H$ (bottom).

Fig. 6.11—6.12 shows the vertical structure of the edge mode near the gap edge. The density perturbation is even more stratified than a self-gravitating vortex. Most of the disturbance is contained within $z < H$ and the corresponding vortensity perturbation is no longer columnar. Mass flux goes into the over-density radially from either side, as well as vertically from above. Since the density perturbation in the midplane is

6. THREE-DIMENSIONAL SIMULATIONS

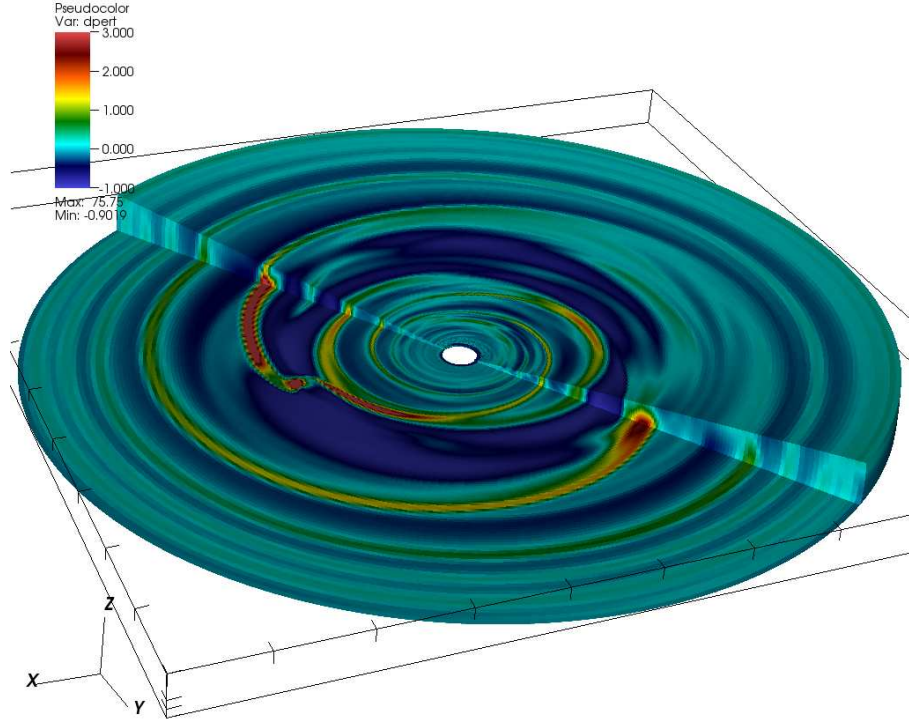


Figure 6.10: *Three-dimensional visualisation the edge mode instability for the simulation in Fig. 6.9. The snapshot is taken at $t = 35P_0$. One of the edge mode spiral arms coincide with the planetary wake. The colour bar is relative density perturbation.*

large (up to 2–3 times over-dense than the background), this blob of fluid perturbs the surrounding flow gravitationally. The vertical flux is likely attributable to this effect. Recall that the vertical boundaries are solid, so flow must reflect off it.

6.3.5 Outwards migration induced by edge modes

This simulation is the 3D version of the fiducial case in Chapter 5. The disc model is $Q_o = 1.7$ and the resolution is $(N_r, N_\theta, N_\phi) = (256, 32, 512)$. N_θ is reduced to enable longer simulation times. The planet mass is now $M_p = 0.002M_*$ and it is allowed to migrate for $t > 20P_0$. Since torques originating from close to the planet were found to be important in the 2D simulations, the torque cut-off parameter is reduced to $\epsilon_c = 0.5$ (about 3.5 cells).

Fig. 6.13 and Fig. 6.13 shows the migration history and snapshots of density perturbation. These figures can be compared to the 2D case, Fig. 5.8–5.9. Whereas the 2D disc developed $m = 2$ –3 edge modes, only $m = 2$ was seen in 3D. This could

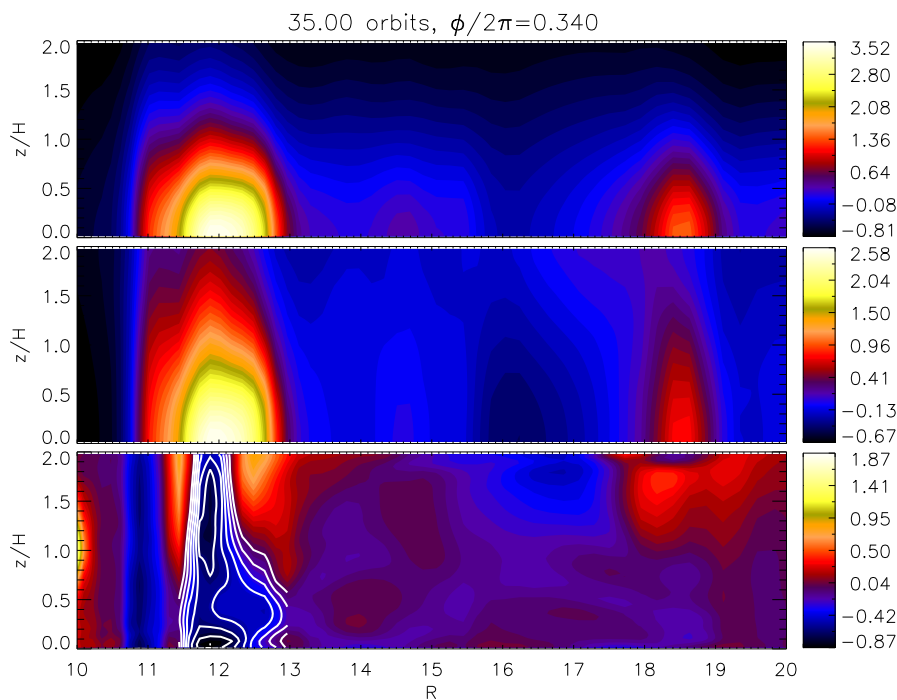


Figure 6.11: *Density perturbation relative to $t = 0$, $10P_0$ (top, middle) and perturbation to the vertical component of vortensity relative to $t = 10P_0$ (bottom) in a vertical plane through the edge mode in Fig. 6.10.*

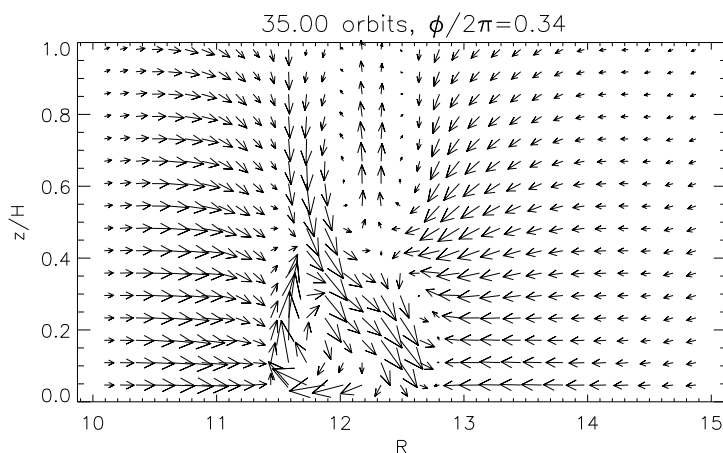


Figure 6.12: *Mass flux corresponding to Fig. 6.11. Note the vertical axis only covers $z < H$.*

be due to the much reduced resolution in 3D (by a factor of 4 in r and ϕ compared to 2D), or reduction in the strength of self-gravity due to disc thickness.

6. THREE-DIMENSIONAL SIMULATIONS

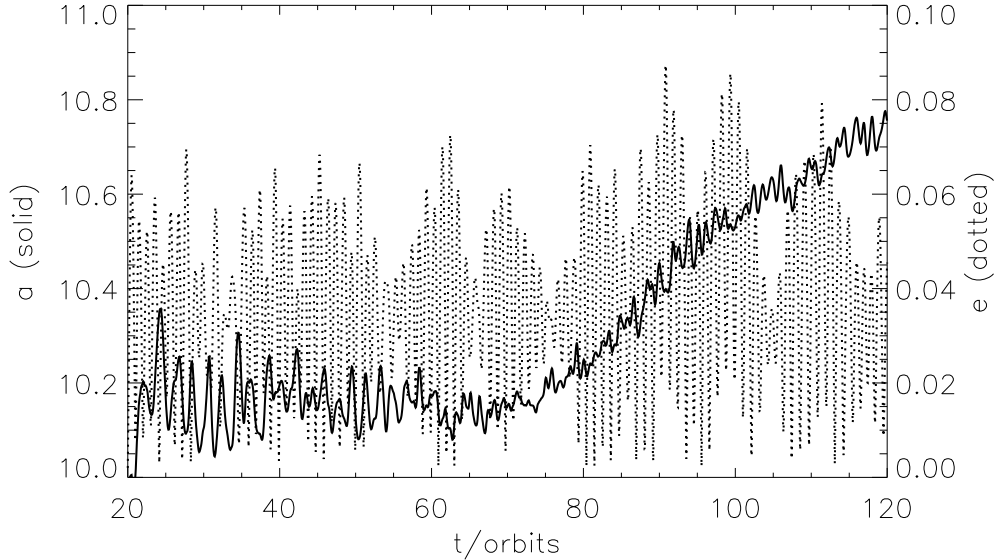


Figure 6.13: *Keplerian semi-major axis a and eccentricity e for planet ($M_p = 0.002M_*$) in the $Q_o = 1.7$ three-dimensional disc model.*

Despite the low resolution, this simulation demonstrates that edge modes can induce outwards migration in 3D discs. Outwards migration only begins around $50P_0$ after release, whereas it starts immediately in the 2D run. The migration in 3D is also on average slower than in 2D. This is probably because the applied tapering (for numerical considerations) artificially reduces the torque magnitudes. However, the existence of edge modes does not depend on treatment close to the planet. Their development will likely bring material into the gap for interaction with the planet, because they are associated with the gap edge. This is shown in Fig. 6.14, which is qualitatively very similar to the 2D run. The planet is asymmetrically positioned with respect to gap edges. Edge mode spiral arms protrude into the gap and supply an over-density within the coorbital region of the planet upstream to it. As the spiral pattern pass by from $\phi > \phi_p$ to $\phi < \phi_p$, edge mode material in the planet's coorbital region is expected to execute inward horseshoe turns ahead of the planet. This contributes positively to the disc-on-planet torque.

6.4 Summary and discussion

The first 3D numerical simulations demonstrating planetary gap instability were presented. The instabilities behave similarly in 2D and 3D. In retrospect, 2D is a better approximation for self-gravitating discs than for non-self-gravitating discs because ver-

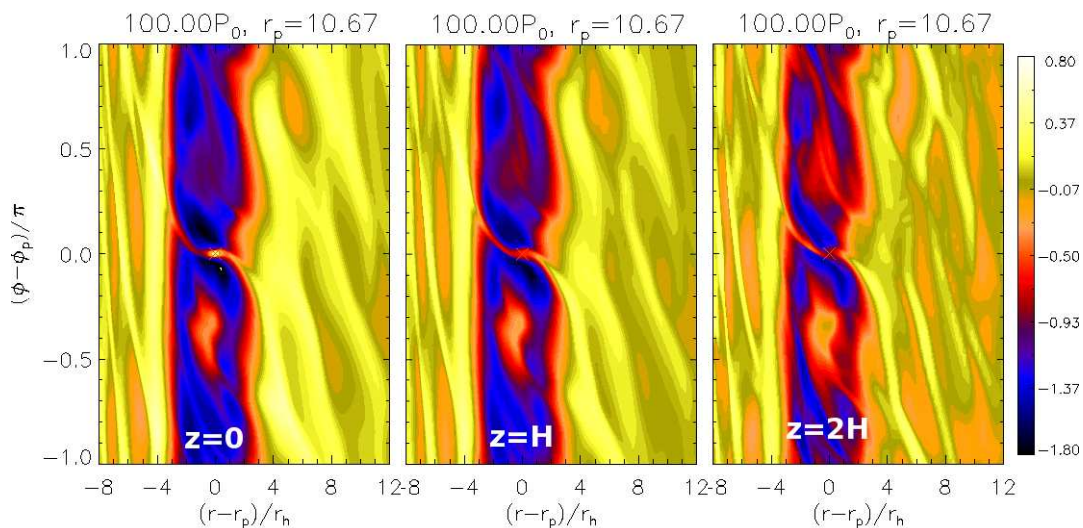


Figure 6.14: *Logarithmic density perturbation $\log[\rho/\rho(t = 10P_0)]$ at three heights, during outwards migration of a planet induced by edge modes in a 3D disc.*

tical self-gravity compresses the midplane.

The vertical structure of the vortex instability was first studied in non-self-gravitating discs by Meheut et al. (2010). They found prominent circulatory flow in the (R, z) plane, but these features are not seen in the current simulations. However, Meheut et al. used prescribed disc profiles which lead to strong instability (growth rate $\simeq 0.4\Omega$) whereas profiles here are induced by a planet, for which 2D linear calculations gave smaller growth rates (a few per cent of Ω). There could also be influences from the planet. Simulations here show vortices and edge mode spirals have increasing vertical dependence as the strength of self-gravity is increased. This is probably due to additional vertical acceleration due to self-gravity by the higher midplane density.

Computational cost present significant obstacle to 3D simulations. Vortices and edge modes are global instabilities associated with localised gradients. Thus entire discs must be simulated whilst maintaining good resolution locally. In the 3D runs here, resolution in (r, ϕ) is only 4–6 cells per H , compared to 16 in 2D simulations. Improved resolution and a convergence study should be performed in the future. Despite numerical limitations, the similarity between 3D and 2D results validates the razor-thin disc approximation employed in previous Chapters. In fact, it may be more advantageous to use 2D in some situations, e.g. massive discs and planet migration, because high resolution is achievable. Migration should only depend on the density field, which has essentially the same structure in 2D and 3D.

6. THREE-DIMENSIONAL SIMULATIONS

Chapter 7

Conclusions and future work

7.1 Main results

7.1.1 Unstable gap structure

Planetary gaps are characterised by thin vortensity rings outlining its edges. They are generated across shocks induced by the planet. Their formation can be anticipated from models of the planet-perturbed flow and expected shock location. For planetary masses of Saturn or above, these shocks extend well into the gap and the rings reside in the planet's co-orbital region. The existence of vortensity extrema satisfies the necessary condition for instability.

7.1.2 Vortex modes and induced migration

In a low viscosity, non-self-gravitating disc, corotation radii of unstable modes are associated with local vortensity minima at gap edges. Their nonlinear outcome is the formation of a few vortices that merge on dynamical timescales. The result is one azimuthally-extended vortex circulating just outside either gap edge.

For the standard disc models adopted, differential Lindblad torques favour inwards migration and lead to gravitational scattering of the inner vortex outwards. The planet therefore receives an inwards kick of a few Hill radii. Migration alternates between a slow phase, when gap and vortex formation takes places, and brief rapid inward jumps.

The role of vortices in such non-smooth migration can be understood as a modified form of type III migration. Vortices provide a coorbital mass deficit when it is circulating on the gap edge, and remove it when they flow across the gap and fill it to some extent.

7. CONCLUSIONS AND FUTURE WORK

7.1.3 Effect of self-gravity on vortex modes

Increasing self-gravity is stabilising for the vortex mode, if the background disc profile remains fixed. However, in practice the effective planetary mass increases with the strength of self-gravity due to disc material bound to the planet. This results in steeper gap edges with increasing surface density scales, and this destabilises the vortex mode.

The combined effect of self-gravity through the background profile and through the linear response is to shift the most unstable vortex mode to higher azimuthal wavenumbers. This leads to more vortices with increasing disc mass. These self-gravitating vortices are localised in azimuth as well as radius, and they can behave like co-orbital planets. Their mutual gravitational interaction allows them to survive against vortex merging for a longer period of time than in non-self-gravitating discs.

7.1.4 Edge modes in massive discs

The vortex mode is suppressed in sufficiently massive discs. Instead, unstable modes have co-rotation radii at local vortensity maximum, which lie closer to the planet than vortensity minimum. These edge modes can be interpreted as a disturbance associated with the gap edge, which induces activity in the smooth part of the disc via self-gravity. Thus they are global modes. For the disc models adopted here, they are seen predominantly as one-sided spiral arms in the disc beyond the planet's orbital radius.

Edge modes associated with the outer gap edge make the disc-on-planet torque on average more positive than that in stable discs. Outwards migration is thus promoted. A Saturn-mass planet can be scattered by an edge mode spiral and jump outwards significantly on orbital timescales. More massive Jovian planets reside in deeper gaps but can migrate outwards on timescales of 10's of orbits. These effects originate from material brought into the gap by the edge mode.

7.1.5 Validity of razor-thin discs

The use of 2D discs to study planetary gap stability has been validated by counterpart 3D numerical simulations. Vortex modes and edge modes behave similarly in both geometries. Not surprisingly, the vertical dependence of their structure increases with the strength of self-gravity.

7.2 Directions for future work

Attention has mostly been placed on understanding the phenomenon found in each of the preceding Chapters. The trade-off is the realism of disc models and the size of parameter space that can be practically surveyed. Extensions to this work are listed below.

7.2.1 Three-dimensional linear calculations

Although the non-self-gravitating 3D simulations did not reproduce vertical structure found by Meheut et al. (2010), it is clear that vortices involve vertical motion. In order to gain a better understanding of three-dimensional effects, if any, it would be ideal to have 3D linear stability analysis to complement nonlinear simulations.

A simple starting point would be three-dimensional, locally isothermal discs with radial structure, such as a surface density bump (e.g. profiles used by Li et al., 2000). The vertical structure is Gaussian, so if the disc occupies $z \in [-\infty, \infty]$, one can expand the vertical dependence of perturbations as $W(R, z) = \sum_{l=0}^{\infty} W_l(R) \mathcal{H}_l(z)$, where \mathcal{H}_l is the l^{th} Hermite polynomial. W_0 would represent the 2D solution and W_2 would be the first 3D contribution to an even vortex mode. This leads to the system of coupled first order ODE's analysed by Tanaka et al. (2002), but the problem here would be solving for the homogeneous solution, with eigenfrequencies to be determined. Preliminary results show that 3D has negligible effect on growth rates and, e.g. $|W_2| \ll |W_0|$, so perturbations are indeed columnar. Nonetheless, the small vertical velocities still grow exponentially and this should be important in the nonlinear regime.

However, it is more realistic to have finite vertical boundaries, particularly for the vortex modes. This is because they are easily suppressed by viscosity, and would only exist in ‘dead zones’ of protoplanetary discs which have sufficiently low viscosity. However, dead zones exist close to the disc midplane and do not occupy the entire vertical extent (Gammie, 1996). In this case one requires the solution to a 2D partial differential equation eigenvalue problem.

7.2.2 Planet-edge mode interactions

Systematic studies of planetary migration in massive discs have only begun recently (Michael et al., 2011; Baruteau et al., 2011c). These first simulations consider disc masses of $M_d/M_* \geq 0.14$. Such discs are gravitationally active even without a planet.

Lower mass discs that are gravitationally stable without a planet can become gravitationally active via edge modes associated with planetary gaps. Edge mode properties

7. CONCLUSIONS AND FUTURE WORK

depend on planetary properties, but the unstable disc has significant back-reaction on planet migration, which could subsequently affect further edge mode development. An example is edge mode-induced migration moving the planet to a region even more prone to edge modes.

These inter-dependencies will be explored with high resolution 2D numerical simulations. This is an expansion of Chapter 5, which focused on the migration mechanism, to a parameter survey. Questions include the relation between planet and disc masses that allow edge modes, and how the extent of migration depend on planetary and disc properties.

7.2.3 Numerical simulations with ZEUS-MP

Three-dimensional simulations in Chapter 6 were primarily performed to confirm 2D results. Future numerical studies with ZEUS-MP will focus on properties of vortex and edge modes in 3D self-gravitating discs.

Further development of ZEUS-MP is needed. Currently, the non-self-gravitating initial conditions are evolved towards equilibrium. This inevitably leaves disturbances in the disc, relative to which perturbations are defined. Although subsequent perturbations due to planet, vortices and edge modes are much larger than the noise in the background, it would be ideal to have a module to generate clean initial discs in equilibrium with self-gravity. Development of polytropic disc models is also planned, as a preliminary step towards including thermodynamics, which is important for gravitational instabilities.

Many interesting issues arise in 3D. The vertical structure of vortex and edge modes as a function of gap properties should be investigated. Vertical boundary conditions (including the treatment of the gravitational potential in terms of l_{\max}, m_{\max}) may be important for vortex modes because they tend to involve entire fluid columns, i.e. they are global in z . The evolution of 3D self-gravitating disc vortices, including the possibility of self-gravitational collapse, vortex merging (a phenomenon common in 2D) and their migration (so far only simulated in 2D by Paardekooper et al., 2010a) should be studied. These numerical works require improved numerical resolutions.

Appendices

Appendix A

Upper limit on the horseshoe width

It is possible to place an upper limit on the horseshoe width x_s that is valid in the limit of either zero or constant pressure. Consider a local Cartesian frame (x, y) that has origin at the planet and hence co-rotates with it with the Keplerian angular velocity $\Omega_p = \sqrt{GM_*/r_p^3}$. The x direction corresponds to increasing radius and y is azimuthal direction (same sense as Ω_p). In this frame, consider fluid elements approaching the planet from $(x = x_0 > 0, y = \infty)$ with velocity $(0, -3\Omega_p x_0/2)$. Suppose such a fluid element executes a horseshoe turn crossing the co-orbital radius at position $(0, y)$ with velocity $(u_x, 0)$. The maximum value of x_0 , below which such motions occur, is determined as follows.

The equation of motion governing a fluid element or particle, in the absence of pressure forces, is

$$\frac{D\mathbf{u}}{Dt} + 2\Omega_p \hat{\mathbf{z}} \times \mathbf{u} = -\nabla\Phi_{\text{tot}}, \quad (\text{A.1})$$

where

$$\Phi_{\text{tot}} = -\frac{GM_p}{\sqrt{x^2 + y^2}} - \frac{3}{2}\Omega_p^2 x^2. \quad (\text{A.2})$$

Φ_{tot} contains contributions from the gravitational potential of the planet of mass M_p (with the softening length neglected) and the tidal potential associated with the central object. From Eq. A.1 it follows that the Jacobi invariant

$$J = \frac{1}{2}|\mathbf{u}|^2 + \Phi_{\text{tot}} \quad (\text{A.3})$$

is constant along a particle path. Equating J evaluated at $(x, y) = (x_0, \infty)$ to J

A. UPPER LIMIT ON THE HORSESHOE WIDTH

evaluated at $(0, y)$ gives

$$-\frac{3}{8}\Omega_p^2 x_0^2 = \frac{1}{2}u_x^2 - \frac{GM_p}{y}. \quad (\text{A.4})$$

The steady-state Euler equation of motion for u_y evaluated at $(0, y)$ is

$$u_x \frac{\partial u_y}{\partial x} + 2\Omega_p u_x = -GM_p/y^2. \quad (\text{A.5})$$

In the neighbourhood of $(0, y)$ for the type of streamline being considered, $u_x < 0$ and $\partial_x u_y < 0$ because the particle is going from $x > 0$ to $x < 0$ by executing a U-shape turn. This implies

$$|u_x| > \frac{GM_p}{2\Omega_p y^2}. \quad (\text{A.6})$$

Combining Eq. A.4 and Eq. A.6 gives

$$-\frac{3}{8}\Omega_p^2 x_0^2 > \frac{1}{2} \left(\frac{GM_p}{2\Omega_p y^2} \right)^2 - \frac{GM_p}{y}. \quad (\text{A.7})$$

Writing x_0 and y in units of the Hill radius, i.e. setting $x_0 = \hat{x}_0 r_h$ and $y = \hat{y} r_h$, gives

$$\hat{x}_0 < \sqrt{\frac{8}{\hat{y}} - \frac{3}{\hat{y}^4}} \lesssim 2.3. \quad (\text{A.8})$$

Thus because the maximum possible value of the right hand side of the above is 2.3, particles executing a U-turn could not have originated further than $2.3r_h$ from the planet's orbital radius. This is comparable to the value of $2.5r_h$ that has been estimated from hydrodynamic simulations (Artymowicz, 2004b; Paardekooper & Papaloizou, 2009).

A.1 Application to co-orbital vortices

It is straight forward to apply the equations above to describe gravitational interaction between two vortices in the shearing sheet. The equations have exactly the same form, with a slight change of notation $x \rightarrow X$, $y \rightarrow Y$ where X, Y denote the relative displacement of the two vortices in the shearing sheet frame. $M_p \equiv qM_*$ becomes the sum of vortex masses. The frame is now centred on one of the vortices, its average orbital radius from the central star being $r = r_p$.

A.1 Application to co-orbital vortices

Non-dimensionalising by r_p , Eq. A.7 becomes

$$\frac{3}{8}\hat{X}_0^2 < \frac{q}{\hat{Y}} - \frac{q^2}{8\hat{Y}^4} \quad (\text{A.9})$$

where $\hat{X}_0 = X_0/r_p$, $\hat{Y} = Y/r_p$, and the angular speed of the rotating frame $\Omega_p = \sqrt{GM_*/r_p^3}$ has been used. Eq. A.9 relates the initial mutual radial separation of the vortex pair, \hat{X}_0 , to their minimum azimuthal separation, \hat{Y} .

Eq. A.9 is useful for the case there vortices are just able to undergo U turns. For fixed q , the function

$$f(\hat{Y}; q) = \frac{q}{\hat{Y}} - \frac{q^2}{8\hat{Y}^4} \quad (\text{A.10})$$

has a maximum value at $\hat{Y} = (q/2)^{1/3}$, corresponding to the maximum conceivable initial separation $X_0 = X_s$, where

$$X_s = 2^{2/3}q^{1/3}r_p. \quad (\text{A.11})$$

If $X_0 > X_s$ then Eq. A.9 cannot be satisfied and there can be no mutual horseshoe turns (as in the test particle and planet case). For initial separations $X_0 < X_s$, Eq. A.9 implies that the minimal inter-vortex distance (in Y) must exceed $q^{1/3}/2$ (so that $f, \hat{X}_0^2 > 0$). Now for sufficiently large q , $q^{1/3}/2$ will be larger than the critical separation for merging (set by hydrodynamics), so merging is avoided during the encounter.

It is interesting to compare Eq. A.11 to the estimate of the horseshoe half-width x_s of Paardekooper & Papaloizou (2009). They found $x_s = 1.68(q/h)^{1/2}r_p$ based on hydrodynamic simulations for low mass planets. Equating x_s and X_s with $h = 0.05$ gives $q = 8.9 \times 10^{-5}$. This should give the minimum q for which pressure effects could be ignored. Inserting this value in Murray & Dermott (2000)'s model of coorbital satellites gives a minimal separation of 0.58 (with other parameter values estimated in §3.5.4), which is close to simulation results (Fig. 3.21). Hence, if the vortex-pair interaction in Chapter 3 is purely gravitational, a single vortex behaves in a similar way to ~ 15 Earth masses, i.e. a low-mass protoplanet.

Considering a vortex size of order H , the vortex-to-star mass ratio is $q \sim \pi H^2 \Sigma / M_* \simeq h^3 / Q$. For a self-gravitating vortex where the Toomre $Q \sim 1$, $q \simeq h^3 = 1.25 \times 10^{-4}$ for $h = 0.05$, slightly exceeding the threshold value above. Hence the pressureless treatment of self-gravitating vortex-vortex interactions should be acceptable for the purpose of explaining the resisted-merging of self-gravitating vortices.

A. UPPER LIMIT ON THE HORSESHOE WIDTH

Appendix B

Linear model for flow perturbed by a planet

When modelling the pre-shock flow in the presence of a planet (§2.3.1), the pressureless momentum equations were integrated directly. Here, a simple linear model for the pre-shock velocity field is presented, which was found to give qualitatively correct results compared to hydrodynamic simulations.

Low mass planets cannot have shocks close to $r = r_p$ because flow is purely azimuthal and becomes subsonic close to the planet (in the planet's frame). The goal of this calculation is to introduce non-zero radial velocity by treating the planet potential as a perturbation. The planet mass M_p is assumed to be sufficiently small to allow linearisation, but large enough to ignore pressure. The zero-pressure, steady state momentum equations in the shearing sheet are

$$u_x \frac{\partial u_x}{\partial x} + u_y \frac{\partial u_x}{\partial y} - 2\Omega_p u_y = -\frac{\partial \Phi_p}{\partial x} + 3\Omega^2 x \quad (\text{B.1})$$

$$u_x \frac{\partial u_y}{\partial x} + u_y \frac{\partial u_y}{\partial y} + 2\Omega_p u_x = -\frac{\partial \Phi_p}{\partial y}. \quad (\text{B.2})$$

The centre of the frame co-rotates with the planet. Linearising about the Keplerian flow

$$u_x \rightarrow 0 + \delta u_x \quad (\text{B.3})$$

$$u_y \rightarrow -\frac{3}{2}\Omega_p x + \delta u_y, \quad (\text{B.4})$$

where δ denote small Eulerian perturbations, gives

B. LINEAR MODEL FOR FLOW PERTURBED BY A PLANET

$$\begin{aligned} \frac{3}{2}x\Omega_p \frac{\partial \delta u_x}{\partial y} + 2\Omega_p \delta u_y &= \frac{\partial \Phi_p}{\partial x} \\ \Rightarrow \frac{\partial \delta u_x}{\partial y} + \frac{4}{3x} \delta u_y &= \frac{2}{3\Omega_p x} \frac{\partial \Phi_p}{\partial x} \end{aligned} \quad (\text{B.5})$$

from the x -momentum equation. The linearised y -momentum equation is:

$$\begin{aligned} \delta u_x \left(-\frac{3}{2}\Omega_p \right) - \frac{3}{2}x\Omega_p \frac{\partial \delta u_y}{\partial y} + 2\Omega_p \delta u_x &= -\frac{\partial \Phi_p}{\partial y} \\ \Rightarrow \delta u_x &= 3x \frac{\partial \delta u_y}{\partial y} - \frac{2}{\Omega_p} \frac{\partial \Phi_p}{\partial y}. \end{aligned} \quad (\text{B.6})$$

Eliminating δu_x gives

$$\frac{\partial^2 \delta u_y}{\partial y^2} + k^2 \delta u_y = \frac{k}{\Omega_p} \frac{\partial^2 \Phi_p}{\partial y^2} + \frac{k^2}{2\Omega_p} \frac{\partial \Phi_p}{\partial x} \equiv F, \quad (\text{B.7})$$

where $k \equiv 2/3x$. Eq. B.7 is an ODE with x as a parameter. Multiplying by k^{-2} and considering the limit $x \rightarrow 0$ ($k^{-1} \rightarrow \infty$) demands the balance:

$$\delta u_y = \frac{1}{2\Omega_p} \frac{\partial \Phi_p}{\partial x}, \quad (\text{B.8})$$

which implies, from Eq. B.6, that close to $x = 0$,

$$\delta u_x = \frac{3x}{2\Omega_p} \frac{\partial^2 \Phi_p}{\partial x \partial y} - \frac{2}{\Omega_p} \frac{\partial \Phi_p}{\partial y}. \quad (\text{B.9})$$

The expression for δu_y is equivalent to Eq. 28 of Paardekooper & Papaloizou (2009) when pressure is neglected, and δu_x is equivalent to their Eq. 30 if $\partial_x \Phi_p$ is neglected.

Eq. B.7 describes a forced oscillator and solutions are readily obtained by writing

$$\delta u_y = A(y) \cos ky + B(y) \sin ky, \quad (\text{B.10})$$

and imposing

$$A' \cos ky + B' \sin(ky) = 0,$$

where the prime denotes ∂_y . For $x > 0$, the coefficients are written as

$$A = k^{-1} \int_y^\infty F(y') \sin ky' dy', \quad (\text{B.11})$$

$$B = -k^{-1} \int_y^\infty F(y') \cos ky' dy'. \quad (\text{B.12})$$

So that $\delta u_y, \delta u_x \rightarrow 0$ as $y \rightarrow \infty$ for fixed x . The $\partial_y \Phi_p$ part of the forcing F can be integrated by parts (twice) to give explicit velocity expressions

$$\delta u_y = -\frac{k\Phi_p}{\Omega_p} - \frac{k^2 GM_p}{\Omega_p} \int_y^\infty \frac{\sin k(y-y') dy'}{(x^2 + y'^2 + \epsilon_p^2)^{1/2}} + \frac{GM_p}{3\Omega_p} \int_y^\infty \frac{\sin k(y'-y) dy'}{(x^2 + y'^2 + \epsilon_p^2)^{3/2}}, \quad (\text{B.13})$$

$$\begin{aligned} \delta u_x &= -\frac{2k^2 GM_p}{\Omega_p} \int_y^\infty \frac{\cos k(y-y') dy'}{(x^2 + y'^2 + \epsilon_p^2)^{1/2}} - \frac{2GM_p}{3\Omega_p} \int_y^\infty \frac{\cos k(y'-y) dy'}{(x^2 + y'^2 + \epsilon_p^2)^{3/2}} \\ &= 2B \cos ky - 2A \sin ky - \frac{2}{\Omega_p} \frac{\partial \Phi_p}{\partial y}. \end{aligned} \quad (\text{B.14})$$

The second line for δu_x follows from Eq. B.6 once A, B are known, and this is the expression used in practice. Since $k \propto x^{-1}$, linearisation fails for small x .

The planet-perturbed velocity field is now available for use. To first approximation, let the velocity field be $(\delta u_x, u_y)$, i.e. the y -velocity perturbation is neglected in comparison to Keplerian flow. The unperturbed uniform density field, $\Sigma = \Sigma_0 \times 10^{-4}$, is assumed when the surface density is needed. Taking the shock front from simulation data, Fig. B.1 compares shock strength predicted by the above calculation to that found in disc-planet simulations. The linear solution over-estimates M^2 but its position of maximum, at $r - r_p = 1.1r_h, 1.5r_h$ (for $M_p = 2.8 \times 10^{-4} M_*, 2 \times 10^{-4} M_*$ respectively) is close to simulation results. In particular, the $M_p = 2 \times 10^{-4} M_*$ curve follows closely to simulation. This simple linear calculation adequately predicts the qualitative behaviour of M^2 along a given shock. More importantly it confirms the planet-induced u_x enable shocks near $r = r_p$, and that vortensity is generated within about a Hill radius from the planet (which is the region of increasing M^2 with respect to x).

Fig. B.2 compare vortensity jumps. Linear theory under-estimates vortensity generation, probably because $|\omega/\Sigma| \propto 1/\Sigma$ and the unperturbed uniform density was used, which is larger than the actual pre-shock surface density (this is seen in Fig. 2.4 as diverging particle paths approaching the U-turn). The agreement between the linear model and simulation is better in $M_p = 2 \times 10^{-4} M_*$, in terms of position of maximum vortensity jump and shock strength, since linearisation become better approximation as M_p decreases (despite the pressureless approximation becoming worse).

B. LINEAR MODEL FOR FLOW PERTURBED BY A PLANET

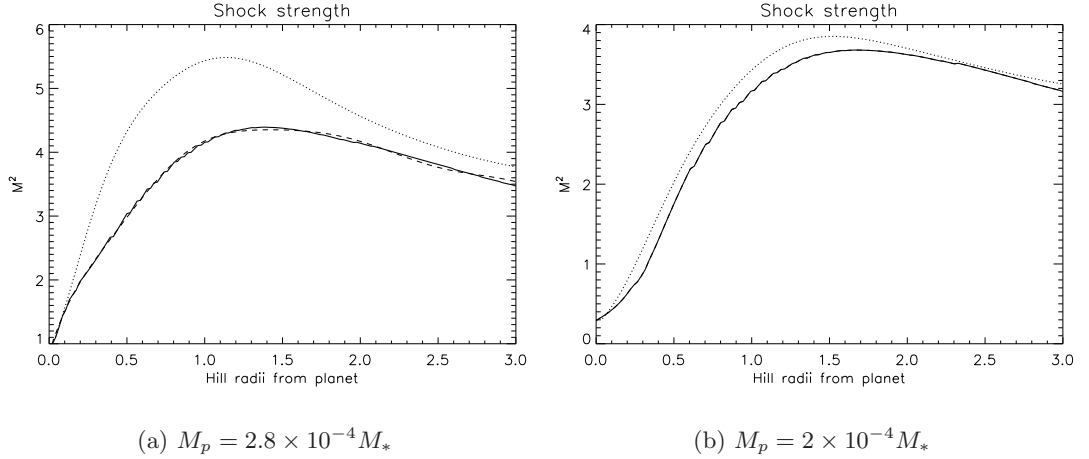


Figure B.1: Perpendicular Mach number squared, M^2 , along the outer spiral shock induced by a giant planet. The location of the shock front is taken from simulation. The horizontal axis is the x co-ordinate of the shock in the shearing sheet. The solid line is M^2 measured from simulation, with the dashed curve as a polynomial fit, and the dotted line is M^2 calculated from the velocity field given by the linear model (but with perturbation to the y -velocity neglected).

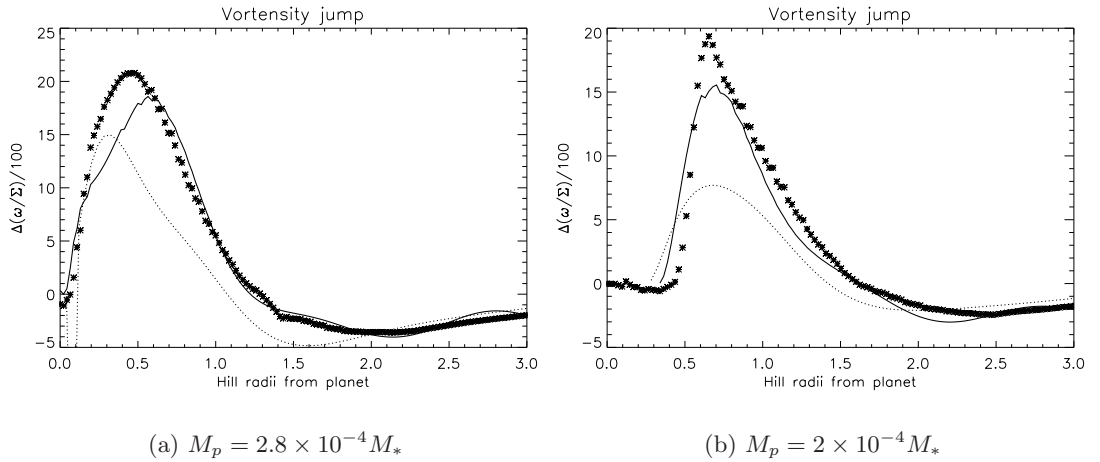


Figure B.2: Vortensity jump across the outer spiral shock induced by a giant planet. The location of the shock front is taken from simulation. Asterisks are jumps measured from simulation. The solid line is the predicted jump given pre-shock data from simulation using the analytic jump expression (Eq. 2.12). The dotted curve is the jump predicted using the linear model for pre-shock velocity and using the unperturbed uniform surface density in the pre-shock flow.

Appendix C

Shocks induced by a planet and the vorticity jump across it

C.1 Estimating shock fronts

Papaloizou et al. (2004) presented a method based on characteristic rays to estimate the shape of wakes induced by a planet. It was applied to Keplerian flow. Their formalism is outlined below and applied to general two-dimensional flow. Note that the method applies to local or global frames.

Consider a frame which rotates at the planet's fixed circular orbital frequency. The hydrodynamic variables are assumed to have wave-like spatial and time dependencies, $\exp i(\mathbf{k} \cdot \mathbf{x} - \sigma t)$, where \mathbf{k} is a wavenumber and $\sigma(\mathbf{k}, \mathbf{x}, t)$ a frequency. Seeking a local dispersion relation from the hydrodynamic equations yields

$$\sigma = -\mathbf{k} \cdot \mathbf{v} + c_s |\mathbf{k}|. \quad (\text{C.1})$$

The wavefronts or characteristic curves $y(x)$ are related to the dispersion relation by

$$\begin{aligned} \frac{dy}{dx} &= \frac{\partial \sigma / \partial k_y}{\partial \sigma / \partial k_x} \\ &= \frac{-\hat{v}_y + k_y / |\mathbf{k}|}{-\hat{v}_x + k_x / |\mathbf{k}|} \\ &= \frac{-\hat{v}_y (1 + \beta^2)^{1/2} + \beta}{-\hat{v}_x (1 + \beta^2)^{1/2} + 1}, \end{aligned} \quad (\text{C.2})$$

where $\hat{v}_y = v_y / c_s$ is the non-dimensional velocity and similarly for v_x . The second line in Eq. C.2 applies the dispersion relation Eq. C.1, and the third line follows by defining $\beta \equiv k_y / k_x$.

C. SHOCKS INDUCED BY A PLANET AND THE VORTICITY JUMP ACROSS IT

The solution of interest is the characteristic ray with $\sigma = 0$, corresponding to a stationary wave in the planet frame. Setting the frequency to zero, Eq. C.1 reads:

$$0 = -\hat{v}_x - \beta\hat{v}_y + (1 + \beta^2)^{1/2}, \quad (\text{C.3})$$

which is a quadratic equation for β with solution

$$\beta = \frac{\hat{v}_x\hat{v}_y \pm (\hat{v}^2 - 1)^{1/2}}{1 - \hat{v}_y^2}, \quad (\text{C.4})$$

where $\hat{v}^2 = \hat{v}_x^2 + \hat{v}_y^2$. Using Eq. C.3 to replace $(1 + \beta^2)^{1/2}$ in Eq. C.2 gives

$$\frac{dy}{dx} = \frac{\beta(1 - \hat{v}_y^2) - \hat{v}_x\hat{v}_y}{1 - \hat{v}_x^2 - \beta\hat{v}_x\hat{v}_y}. \quad (\text{C.5})$$

Finally, inserting the expression for β from Eq. C.4 gives

$$\frac{dy}{dx} = \frac{\pm(\hat{v}^2 - 1)^{1/2}(\hat{v}_y^2 - 1)}{(\hat{v}^2 - 1) \pm \hat{v}_x\hat{v}_y(\hat{v}^2 - 1)^{1/2}} \quad (\text{C.6})$$

$$= \frac{\hat{v}_y^2 - 1}{\hat{v}_x\hat{v}_y \pm (\hat{v}^2 - 1)^{1/2}}, \quad (\text{C.7})$$

which can be solved once a flow field is given. The solution $y(x)$ correspond to wakes induced by the planet, with the minus (plus) sign corresponding to wakes exterior (interior) to the planet. These wakes develop into shocks for sufficiently large planetary masses.

C.2 Vorticity jump across a steady isothermal shock

Consider an isothermal shock that is stationary in a frame rotating with angular velocity $\Omega_p \hat{\mathbf{z}}$. In order to evaluate the vorticity jump across the shock it is convenient to use a two dimensional orthogonal coordinate system (x_1, x_2) defined in the disc mid-plane such that one of the curves $x_2 = \text{constant} = x_s$ coincides with the shock. The curves $x_1 = \text{constant}$ will then be normal to the shock where they intersect it. In addition the coordinates are set up so that (x_1, x_2, z) is a right handed system. This co-ordinate system is depicted in Fig.C.1 where distances along the normal (x_2) and tangential (x_1) directions to the shock are labelled N, S respectively.

The $\hat{\mathbf{z}}$ -component of relative vorticity ω_r can then be written as

$$\omega_r = \frac{1}{h_1 h_2} \left(\frac{\partial(u_2 h_2)}{\partial x_1} - \frac{\partial(u_1 h_1)}{\partial x_2} \right), \quad (\text{C.8})$$

C.2 Vorticity jump across a steady isothermal shock

where (h_1, h_2) are the components of the coordinate scale factor.

Note that on $x_2 = x_s$, u_1 is the velocity tangential to the shock and is unchanged across the shock. The normal component u_2 and other state variables undergo a jump from pre-shock values to post-shock values on normally traversing the curve $x_2 = x_s$. For an isothermal shock

$$\frac{u_{2,\text{post}}}{u_2} = M^{-2} = \frac{\rho}{\rho_{\text{post}}}, \quad (\text{C.9})$$

where $M = u_2/c_s$ is the pre-shock perpendicular Mach number. Here and in similar expressions below connecting pre-shock and post-shock quantities, post-shock values are denoted with a subscript $_{\text{post}}$ while pre-shock quantities are left without a corresponding subscript. Thus the jump in relative vorticity is

$$[\omega_r] = \omega_{r,\text{post}} - \omega_r. \quad (\text{C.10})$$

Quite generally the x_1 component of the equation of motion for a steady state flow is

$$\frac{u_1}{h_1} \frac{\partial u_1}{\partial x_1} + \frac{u_2}{h_2} \frac{\partial u_1}{\partial x_2} - u_2 \left(\frac{u_2}{h_1 h_2} \frac{\partial h_2}{\partial x_1} - \frac{u_1}{h_1 h_2} \frac{\partial h_1}{\partial x_2} \right) = -\frac{1}{\rho h_1} \frac{\partial p}{\partial x_1} - \frac{1}{h_1} \frac{\partial \Phi_{\text{eff}}}{\partial x_1} + 2\Omega_p u_2, \quad (\text{C.11})$$

or equivalently

$$\frac{u_1}{h_1} \frac{\partial u_1}{\partial x_1} + \frac{u_2}{h_1} \frac{\partial u_2}{\partial x_1} - (2\Omega_p + \omega_r) u_2 = -\frac{1}{\rho h_1} \frac{\partial p}{\partial x_1} - \frac{1}{h_1} \frac{\partial \Phi_{\text{eff}}}{\partial x_1}, \quad (\text{C.12})$$

where p is the pressure and Φ_{eff} the total potential (the latter quantity being continuous across the shock). From Eq. C.12, the relative vorticity can be written in the form

$$\omega_r = \frac{1}{h_1} \frac{\partial u_2}{\partial x_1} + \frac{u_1}{u_2 h_1} \frac{\partial u_1}{\partial x_1} + \frac{1}{\rho u_2 h_1} \frac{\partial p}{\partial x_1} + \frac{1}{u_2 h_1} \frac{\partial \Phi_{\text{eff}}}{\partial x_1} - 2\Omega_p. \quad (\text{C.13})$$

Applying Eq. C.13 to give an expression for the post shock relative vorticity, and using Eq. C.9 to express the post-shock normal velocity and density in terms of the corresponding pre-shock quantities, the vorticity jump can be written in the form

$$[\omega_r] = \frac{1}{h_1} \frac{\partial(M^{-2} u_2)}{\partial x_1} + \frac{M^2 u_1}{u_2 h_1} \frac{\partial u_1}{\partial x_1} + \frac{M^2}{u_2 h_1} \frac{\partial \Phi_{\text{eff}}}{\partial x_1} + \frac{1}{\rho u_2 h_1} \frac{\partial p_{\text{post}}}{\partial x_1} - (\omega_r + 2\Omega_p), \quad (\text{C.14})$$

where continuity of ρu_2 across the shock has been used in the pressure term. Adopting a locally isothermal equation of state implies $p_{\text{post}} = M^2 p$. Substituting this into Eq.

C. SHOCKS INDUCED BY A PLANET AND THE VORTICITY JUMP ACROSS IT

C.14, while making use of Eq. C.13 together with the relation $u_2 = c_s M$ gives

$$[\omega_r] \equiv [\omega] = -\frac{c_s(M^2 - 1)^2}{M^2 h_1} \frac{\partial M}{\partial x_1} + (M^2 - 1)\omega - \frac{(M^4 - 1)}{M h_1} \frac{\partial c_s}{\partial x_1}, \quad (\text{C.15})$$

where $\omega = 2\Omega_p + \omega_r$ is the absolute vorticity.

Setting $h_1 dx_1 = dS$ with dS being the corresponding element of distance measured parallel to the shock, this takes the form

$$[\omega_r] \equiv [\omega] = -\frac{c_s(M^2 - 1)^2}{M^2} \frac{\partial M}{\partial S} + (M^2 - 1)\omega - \frac{(M^4 - 1)}{M} \frac{\partial c_s}{\partial S}. \quad (\text{C.16})$$

This gives the vorticity jump across a shock in terms of pre-shock quantities measured in the rotating frame in which it appears steady. It is important to note that Eq. C.16 applies specifically in the right handed coordinate system adopted with shock location given by $x_2 = x_s$. If instead $x_1 = x_s$ was adopted for this location, the signs of the derivative terms would be reversed as in the expression (2.23) given in Kevlahan (1997). Note too that the last term on the right hand side that is proportional to the gradient of c_s along the shock arises from the assumption of a *locally* isothermal equation of state and is not present in the treatment given by Kevlahan (1997).

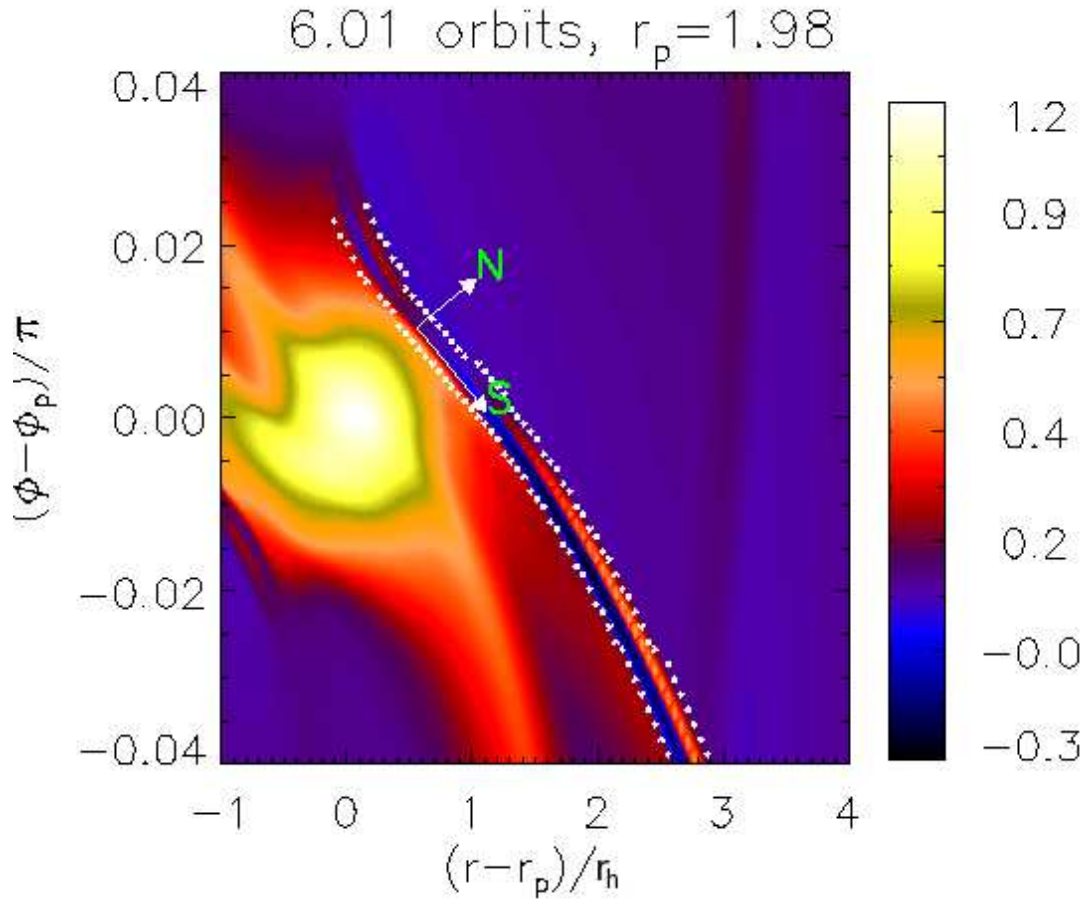


Figure C.1: Vorticity generation across the outer spiral shock induced by planet. The absolute vorticity is shown on a linear scale. Arrows define directions normal (N , equivalent to x_2) and tangential (S , equivalent to x_1) to the shock. The upper (lower) dotted line indicate pre-shock (post-shock) flow, since in the planet frame, for $r > r_p$ material flows towards the planet from $\phi > \phi_p$.

**C. SHOCKS INDUCED BY A PLANET AND THE VORTICITY JUMP
ACROSS IT**

Appendix D

Approximate equations in linear calculations

The full governing equation for linear perturbations in a self-gravitating disc with fixed temperature profile was summarised in Chapter 3 as

$$\mathcal{L}(W) = 0. \tag{D.1}$$

An expression for the linear operator \mathcal{L} is given by Eq. 3.33. The eigenvalue σ is contained in \mathcal{L} as $\bar{\sigma}$ in a complicated manner. However, once appearances of $\bar{\sigma}$ as denominators are removed by multiplying Eq. D.1 by the appropriate factor, one can neglect various powers of $\bar{\sigma}$, to approximate the above equation in quadratic and linear order in σ .

D.1 Governing equation to σ^2

In the quadratic approximation, the governing equation is

$$(\sigma^2 \mathcal{L}_2 + \sigma \mathcal{L}_1 + \mathcal{L}_0)W = 0, \tag{D.2}$$

D. APPROXIMATE EQUATIONS IN LINEAR CALCULATIONS

where

$$\begin{aligned} \mathcal{L}_0 = & m\Omega\kappa^2\mathcal{D}_1 + m\Omega[\kappa^2/r - \kappa^{2'} + \kappa^2\Sigma'/\Sigma + 2m^2\Omega\Omega']\mathcal{D}_2 \\ & + m\{\kappa^2[2r(\Omega' + \Omega\Sigma'/\Sigma) - m^2\Omega] - 2r\Omega[\kappa^{2'} + m^2\Omega(\Omega\Sigma'/\Sigma - \Omega')]\}\mathcal{D}_3/r^2 \\ & + m(2c_s^{2'}\kappa^2\Omega/r - 2m^2\Omega^3c_s^{2'}/r - \kappa^4\Omega)\mathcal{D}_4 \end{aligned} \quad (\text{D.3})$$

$$\begin{aligned} \mathcal{L}_1 = & \kappa^2\mathcal{D}_1 + (\kappa^2/r - \kappa^{2'} + \kappa^2\Sigma'/\Sigma + 4m^2\Omega\Omega')\mathcal{D}_2 - m^2(\kappa^2 + 4r\Omega^2\Sigma'/\Sigma)\mathcal{D}_3/r^2 \\ & - (4m^2\Omega^2c_s^{2'}/r + \kappa^4)\mathcal{D}_4 \end{aligned} \quad (\text{D.4})$$

$$\mathcal{L}_2 = 2m\Omega'\mathcal{D}_2 - 2m(\Omega' + \Omega\Sigma'/\Sigma)\mathcal{D}_3/r - 2m\Omega c_s^{2'}\mathcal{D}_4/r. \quad (\text{D.5})$$

The operators \mathcal{D}_i are:

$$\mathcal{D}_1 = c_s^2 \frac{d^2}{dr^2} + c_s^{2'} \frac{d}{dr} + \mathcal{J}_2, \quad (\text{D.6})$$

$$\mathcal{D}_2 = c_s^2 \frac{d}{dr} + \mathcal{J}_1, \quad (\text{D.7})$$

$$\mathcal{D}_3 = c_s^2 + \mathcal{J}_0, \quad (\text{D.8})$$

$$\mathcal{D}_4 = 1. \quad (\text{D.9})$$

The operators \mathcal{J}_i are defined by Eq. 3.34.

D.2 Governing equation to σ

The approximation for corotational or vortex modes made in Chapter 2 can also be recast in a standard eigenvalue form as above. Recall that in this approximation, $\bar{\sigma}$ is only retained when it appears as $1/\bar{\sigma}$ associated with the vortensity gradient term. In other words, $D \simeq \kappa^2$ and $D' \simeq \kappa^{2'}$. The simplified equation is then

$$\mathcal{A}W = \sigma\mathcal{B}W, \quad (\text{D.10})$$

where

$$\begin{aligned} \mathcal{A} = & m\Omega[\mathcal{D}_1 + (1/r + \Sigma'/\Sigma - \kappa^{2'}/\kappa^2)\mathcal{D}_2] \\ & + [2m\Omega(\Sigma'/\Sigma + \Omega'/\Omega - \kappa^{2'}/\kappa^2)/r - m^3\Omega/r^2]\mathcal{D}_3 \\ & + m\Omega(2c_s^{2'}/r - \kappa^2)\mathcal{D}_4, \end{aligned} \quad (\text{D.11})$$

$$\mathcal{B} = -\mathcal{D}_1 - (\Sigma'/\Sigma + 1/r - \kappa^{2'}/\kappa^2)\mathcal{D}_2 + m^2\mathcal{D}_3/r^2 + \kappa^2\mathcal{D}_4. \quad (\text{D.12})$$

Self-gravity was neglected in Chapter 2, in which case the operators \mathcal{J}_i are set to zero.

Appendix E

Artificial vortices in an accretion disc

The artificial Kida-like vortices used in §3.5 were set up by introducing velocity perturbations $(\delta u_r, \delta u_\phi)$ that correspond to the elliptical vortices of Kida (1981) in an incompressible shear flow.

Consider a small patch of the disc, whose centre (r_p, ϕ_p) rotates at angular speed Ω_p about the primary and set up local Cartesian co-ordinates $x = r_p(\phi - \phi_p)$, $y = r_p - r$. In the (x, y) frame, there exists a Kida vortex solution for incompressible flow, whose velocity field (u_x, u_y) is

$$u_x = \frac{3\Omega_p \Theta y}{2(\Theta - 1)}, \quad u_y = -\frac{3\Omega_p x}{2\Theta(\Theta - 1)} \quad (\text{E.1})$$

inside the vortex core. The ratio of the vortex semi-major to semi-minor axis being $\Theta = a/b$ is a free parameter. This velocity field is such that the vorticity ω is constant in the rotating frame. The elliptical boundary of the vortex is defined such that $\omega = -3\Omega_p(1 + \Theta^2)/[2\Theta(\Theta - 1)] = \omega_v - 3\Omega_p/2$ inside the boundary of the vortex and $\omega = -3\Omega_p/2$ outside. The quantity $\omega_v = -3\Omega_p(1 + \Theta)/[2\Theta(\Theta - 1)]$ is then the vorticity of the vortex core relative to the background. Being negative, this corresponds to an anticyclonic vortex. In order to introduce perturbations corresponding to Kida vortices, perturbations $\delta u_r \equiv -u_y$, $\delta u_\phi \equiv u_x$ are introduced inside a specified elliptical boundary with an exponential decay outside. The boundary is fixed by specifying $\Theta = 8$ and semi-major axis $a = H(r_p)$ where the reference radius is $r_p = 1$.

E. ARTIFICIAL VORTICES IN AN ACCRETION DISC

Appendix F

Energy densities in linear theory of edge modes

In order to understand edge modes, in §4.5 the thermal-gravitational energy (TGE) was defined and the different contributions to it were discussed. There, a factor D^2 was applied to the TGE per unit length and to each contributing term in order to overcome numerical difficulties associated with Lindblad resonances. For completeness, the behaviour of the TGE and the various contributions to it is considered here without the additional factor D^2 .

Fig. F.1 shows very similar features to the corresponding curves discussed in §4.5.2. $\text{Re}(\rho)$ is negative around corotation, again signifying a self-gravity driven edge mode. Beyond $r \simeq 6.4$, $\text{Re}(\rho)$ becomes positive due to pressure. The most extreme peak at $r \simeq 5.6$ coincides with the background vortensity edge (Fig. 4.5). The negative contribution from the corotation term is larger in magnitude than that from the positive wave term, resulting in $\text{Re}(\rho) < 0$. The co-rotation radius is such that $r_c < 5.6$, making the shifted frequency $\bar{\sigma}_R(5.6) = m[\Omega(5.6) - \Omega(r_c)] < 0$. Also at $r = 5.6$, $d\eta/dr < 0$, resulting in $\text{Re}(\rho_{\text{corot}}) < 0$ at this point.

$\text{Re}(\rho_{\text{wave}})$ and $\text{Re}(\rho_{\text{corot}})$ have relatively small spurious bumps around $r = 7.2$ that are associated with the outer Lindblad resonance and are numerical. The eigenfunctions W , $\delta\Phi$ are well-defined without singularities there, but evaluating ρ_{corot} and ρ_{wave} involves division by D , which can amplify numerical errors at Lindblad resonances where $D \rightarrow 0$. Integrating $\text{Re}(\rho)$ over $r \in [5, 10]$ gives $U \equiv \text{Re} \int_5^{10} \rho dr < 0$. The TGE is negative, which means gravitational energy dominates over the pressure contribution

F. ENERGY DENSITIES IN LINEAR THEORY OF EDGE MODES

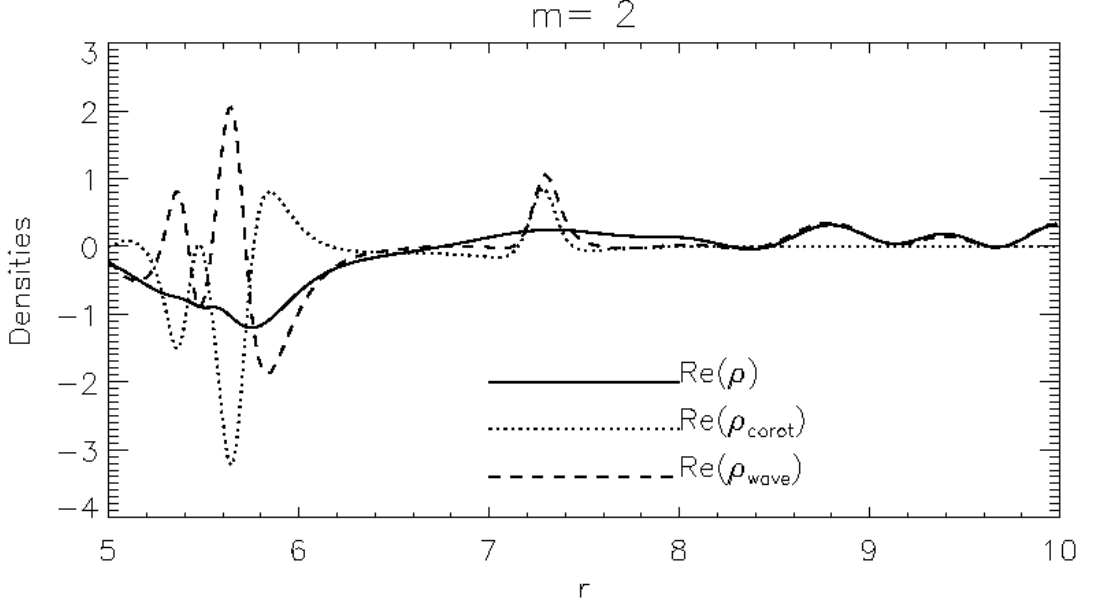


Figure F.1: *The thermal and gravitational energy (TGE) per unit length (solid) computed from eigensolutions for the fiducial case $Q_o = 1.5$ from linear theory. The contributions from the vortensity term (dotted line) and other terms (dashed line), defined by Eq. 4.54, are also shown. The relatively small bumps near $r = 7$ are numerical. However as they are the same magnitude for both contributions, incorporating them does not affect conclusions concerning the overall energy balance in the system.*

for $r \geq 5$. Integrating the contributions to the TGE separately gives

$$U_{\text{corot}} \equiv \text{Re} \int_5^{10} \rho_{\text{corot}} dr \simeq -0.94|U|,$$

$$U_{\text{wave}} \equiv \text{Re} \int_5^{10} \rho_{\text{wave}} dr \simeq -0.077|U|.$$

The integration range includes the OLR and thus the spurious bumps in ρ_{wave} and ρ_{corot} . However, U still approximately equals $U_{\text{corot}} + U_{\text{wave}}$. The correct energy balance is maintained despite being subject to possible numerical error due to the diverging factor $1/D$. The above means that, aside from the spurious bumps, the energy density values for $r \in [5, 10]$ may still be used to interpret energy balance. As as before, the TGE is predominantly accounted for by the vortensity term, since $|U_{\text{corot}}/U| \sim 0.9$.

Bibliography

- Adams, F. C., Ruden, S. P., & Shu, F. H. 1989, *ApJ*, 347, 959
- Armitage, P. J. 2010, *ArXiv e-prints*
- Armitage, P. J., & Hansen, B. M. S. 1999, *Nature*, 402, 633
- Artymowicz, P. 1993, *ApJ*, 419, 155
- Artymowicz, P. 2004a, in *Astronomical Society of the Pacific Conference Series*, Vol. 324, *Debris Disks and the Formation of Planets*, ed. L. Caroff, L. J. Moon, D. Backman, & E. Praton, 39–+
- Artymowicz, P. 2004b, in *KITP Conference: Planet Formation: Terrestrial and Extra Solar*
- Balbus, S. A., & Hawley, J. F. 1991, *ApJ*, 376, 214
- Balmforth, N. J., Cunha, M. S., Dolez, N., Gough, D. O., & Vauclair, S. 2001, *MNRAS*, 323, 362
- Barrett, R., et al. 1994, *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, 2nd Edition (Philadelphia, PA: SIAM)
- Baruteau, C., Cuadra, J., & Lin, D. N. C. 2011a, *ApJ*, 726, 28
- Baruteau, C., Fromang, S., Nelson, R. P., & Masset, F. 2011b, *ArXiv e-prints*
- Baruteau, C., & Masset, F. 2008, *ApJ*, 678, 483
- Baruteau, C., Meru, F., & Paardekooper, S.-J. 2011c, *MNRAS*, 1086
- Boss, A. P. 1980, *ApJ*, 236, 619
- Bracco, A., Chavanis, P. H., Provenzale, A., & Spiegel, E. A. 1999, *Physics of Fluids*, 11, 2280

BIBLIOGRAPHY

- Bryden, G., Chen, X., Lin, D. N. C., Nelson, R. P., & Papaloizou, J. C. B. 1999, *ApJ*, 514, 344
- Chavanis, P. H. 2000, *A&A*, 356, 1089
- Crida, A., Baruteau, C., Kley, W., & Masset, F. 2009, *A&A*, 502, 679
- Crida, A., Morbidelli, A., & Masset, F. 2006, *Icarus*, 181, 587
- D'Angelo, G., Bate, M. R., & Lubow, S. H. 2005, *MNRAS*, 358, 316
- D'Angelo, G., Durisen, R. H., & Lissauer, J. J. 2010, *Giant Planet Formation*, ed. Seager, S., 319–346
- de Val-Borro, M., Artymowicz, P., D'Angelo, G., & Peplinski, A. 2007, *A&A*, 471, 1043
- de Val-Borro, M. e. 2006, *MNRAS*, 370, 529
- Durisen, R. H., Boss, A. P., Mayer, L., Nelson, A. F., Quinn, T., & Rice, W. K. M. 2007, *Protostars and Planets V*, 607
- Fu, W., & Lai, D. 2011, *MNRAS*, 410, 399
- Gammie, C. F. 1996, *ApJ*, 457, 355
- . 2001, *ApJ*, 553, 174
- Godon, P. 1996, *MNRAS*, 282, 1107
- Goldreich, P., & Tremaine, S. 1979, *ApJ*, 233, 857
- . 1980, *ApJ*, 241, 425
- Hawley, J. F. 1987, *MNRAS*, 225, 677
- Hayashi, C. 1981, *Progress of Theoretical Physics Supplement*, 70, 35
- Hayes, J. C., Norman, M. L., Fiedler, R. A., Bordner, J. O., Li, P. S., Clark, S. E., ud-Doula, A., & Mac Low, M. 2006, *ApJS*, 165, 188
- Kevlahan, N.-R. 1997, *J. Fluid Mech.*, 341, 371
- Kida, S. 1981, *Journal of the Physical Society of Japan*, 50, 3517
- Klahr, H. H., & Bodenheimer, P. 2003, *ApJ*, 582, 869

BIBLIOGRAPHY

- Kley, W., & Dirksen, G. 2006, *A&A*, 447, 369
- Kocsis, B., Yunes, N., & Loeb, A. 2011, *Phys. Rev. D*, 84, 024032
- Koller, J., Li, H., & Lin, D. N. C. 2003, *ApJ*, 596, L91
- Korycansky, D. G., & Papaloizou, J. C. B. 1995, *MNRAS*, 274, 85
- Lesur, G., & Papaloizou, J. C. B. 2009, *A&A*, 498, 1
- Li, H., Colgate, S. A., Wendroff, B., & Liska, R. 2001, *ApJ*, 551, 874
- Li, H., Finn, J. M., Lovelace, R. V. E., & Colgate, S. A. 2000, *ApJ*, 533, 1023
- Li, H., Li, S., Koller, J., Wendroff, B. B., Liska, R., Orban, C. M., Liang, E. P. T., & Lin, D. N. C. 2005, *ApJ*, 624, 1003
- Li, H., Lubow, S. H., Li, S., & Lin, D. N. C. 2009, *ApJ*, 690, L52
- Lin, D. N. C., & Papaloizou, J. 1986, *ApJ*, 309, 846
- Lin, D. N. C., & Papaloizou, J. C. B. 1993, in *Protostars and Planets III*, ed. E. H. Levy & J. I. Lunine, 749–835
- Lin, D. N. C., Papaloizou, J. C. B., & Kley, W. 1993, *ApJ*, 416, 689
- Lin, M.-K., & Papaloizou, J. C. B. 2010, *MNRAS*, 405, 1473
- . 2011a, *MNRAS*, 415, 1426
- . 2011b, *MNRAS*, 415, 1445
- Lissauer, J. J., & Stevenson, D. J. 2007, *Protostars and Planets V*, 591
- Lovelace, R. V. E., & Hohlfield, R. G. 1978, *ApJ*, 221, 51
- Lovelace, R. V. E., Li, H., Colgate, S. A., & Nelson, A. F. 1999, *ApJ*, 513, 805
- Lufkin, G., Quinn, T., Wadsley, J., Stadel, J., & Governato, F. 2004, *MNRAS*, 347, 421
- Lynden-Bell, D., & Ostriker, J. P. 1967, *MNRAS*, 136, 293
- Lynden-Bell, D., & Pringle, J. E. 1974, *MNRAS*, 168, 603
- Lyra, W., Johansen, A., Zsom, A., Klahr, H., & Piskunov, N. 2009, *A&A*, 497, 869

BIBLIOGRAPHY

- Marcy, G., Butler, R. P., Fischer, D., Vogt, S., Wright, J. T., Tinney, C. G., & Jones, H. R. A. 2005, *Progress of Theoretical Physics Supplement*, 158, 24
- Masset, F. 2000a, *A&AS*, 141, 165
- Masset, F. S. 2000b, in *Astronomical Society of the Pacific Conference Series*, Vol. 219, *Disks, Planetesimals, and Planets*, ed. G. Garzón, C. Eiroa, D. de Winter, & T. J. Mahoney, 75–+
- Masset, F. S. 2002, *A&A*, 387, 605
- Masset, F. S. 2008, in *EAS Publications Series*, Vol. 29, *EAS Publications Series*, ed. M.-J. Goupil & J.-P. Zahn, 165–244
- Masset, F. S., & Papaloizou, J. C. B. 2003, *ApJ*, 588, 494
- Mayor, M., & Queloz, D. 1995, *Nature*, 378, 355
- McKernan, B., Ford, K. E. S., Lyra, W., Perets, H. B., Winter, L. M., & Yaqoob, T. 2011, *ArXiv e-prints*
- Meheut, H., Casse, F., Varniere, P., & Tagger, M. 2010, *A&A*, 516, A31+
- Melander, M. V., Zabusky, N. J., & McWilliams, J. C. 1988, *Journal of Fluid Mechanics*, 195, 303
- Meschiari, S., & Laughlin, G. 2008, *ApJ*, 679, L135
- Michael, S., Durisen, R. H., & Boley, A. C. 2011, *ApJ*, 737, L42+
- Murray, C. D., & Dermott, S. F. 2000, *Solar System Dynamics*
- Nelson, A. F., & Benz, W. 2003, *ApJ*, 589, 556
- Nelson, R. P., & Papaloizou, J. C. B. 2004, *MNRAS*, 350, 849
- Ou, S., Ji, J., Liu, L., & Peng, X. 2007, *ApJ*, 667, 1220
- Paardekooper, S., Lesur, G., & Papaloizou, J. C. B. 2010a, *ApJ*, 725, 146
- Paardekooper, S.-J., Baruteau, C., Crida, A., & Kley, W. 2010b, *MNRAS*, 401, 1950
- Paardekooper, S.-J., Baruteau, C., & Kley, W. 2011, *MNRAS*, 410, 293
- Paardekooper, S.-J., & Mellema, G. 2006, *A&A*, 459, L17

- . 2008, *A&A*, 478, 245
- Paardekooper, S. J., & Nelson, R. P. 2009, ArXiv e-prints
- Paardekooper, S.-J., & Papaloizou, J. C. B. 2009, *MNRAS*, 394, 2297
- Papaloizou, J. C., & Savonije, G. J. 1991, *MNRAS*, 248, 353
- Papaloizou, J. C. B. 2005, *Celestial Mechanics and Dynamical Astronomy*, 91, 33
- Papaloizou, J. C. B., & Lin, D. N. C. 1989, *ApJ*, 344, 645
- Papaloizou, J. C. B., Nelson, R. P., Kley, W., Masset, F. S., & Artymowicz, P. 2006, *Protostars and Planets V*, 655
- Papaloizou, J. C. B., Nelson, R. P., Kley, W., Masset, F. S., & Artymowicz, P. 2007, in *Protostars and Planets V*, ed. B. Reipurth, D. Jewitt, & K. Keil, 655–668
- Papaloizou, J. C. B., Nelson, R. P., & Snellgrove, M. D. 2004, *MNRAS*, 350, 829
- Papaloizou, J. C. B., & Pringle, J. E. 1984, *MNRAS*, 208, 721
- . 1985, *MNRAS*, 213, 799
- . 1987, *MNRAS*, 225, 267
- Papaloizou, J. C. B., & Terquem, C. 2006, *Reports on Progress in Physics*, 69, 119
- Pepliński, A., Artymowicz, P., & Mellema, G. 2008a, *MNRAS*, 386, 164
- . 2008b, *MNRAS*, 386, 179
- . 2008c, *MNRAS*, 387, 1063
- Petersen, M. R., Julien, K., & Stewart, G. R. 2007a, *ApJ*, 658, 1236
- Petersen, M. R., Stewart, G. R., & Julien, K. 2007b, *ApJ*, 658, 1252
- Pierens, A., & Huré, J.-M. 2005, *A&A*, 433, L37
- Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 1992, *Numerical recipes in FORTRAN. The art of scientific computing*
- Sellwood, J. A., & Kahn, F. D. 1991, *MNRAS*, 250, 278
- Shakura, N. I., & Sunyaev, R. A. 1973, *A&A*, 24, 337

BIBLIOGRAPHY

- Stone, J. M., Mihalas, D., & Norman, M. L. 1992, *ApJS*, 80, 819
- Stone, J. M., & Norman, M. L. 1992a, *ApJS*, 80, 753
- . 1992b, *ApJS*, 80, 791
- Tanaka, H., Takeuchi, T., & Ward, W. R. 2002, *ApJ*, 565, 1257
- Tassoul, J. 1978, *Theory of rotating stars*
- Toomre, A. 1964, *ApJ*, 139, 1217
- Udry, S., & Santos, N. C. 2007, *ARA&A*, 45, 397
- Uribe, A. L., Klahr, H., Flock, M., & Henning, T. 2011, *ApJ*, 736, 85
- Ward, W. R. 1997, *Icarus*, 126, 261
- Weidenschilling, S. J. 1977, *MNRAS*, 180, 57
- Yu, C., & Li, H. 2009, *ApJ*, 702, 75
- Yu, C., Li, H., Li, S., Lubow, S. H., & Lin, D. N. C. 2010, *ApJ*, 712, 198
- Zabusky, N. J., Hughes, M. H., & Roberts, K. V. 1979, *Journal of Computational Physics*, 30, 96
- Zhang, H., Yuan, C., Lin, D. N. C., & Yen, D. C. C. 2008, *ApJ*, 676, 639