

Special features of galactic dynamics: Disc dynamics

C. Efthymiopoulos^a

Research Center for Astronomy and Applied Mathematics, Academy of Athens, Greece

Received 12 July 2010 / Received in final form 3 August 2010

Published online 14 September 2010

Abstract. This is a tutorial presentation of special features of galactic disc dynamics, which completes our introduction to galactic dynamics initially presented in [30]. The emphasis is on topics where galactic dynamics and celestial mechanics share common starting points and/or methods of approach. We start by giving some definitions and general notions on the link between observations and dynamical modeling of discs. Then we focus on the application of resonant Hamiltonian perturbation theory in disc resonances. By examining in detail the case of the Inner Lindblad resonance, we demonstrate how resonant perturbation theory leads to an orbital theory of spiral structure in normal galaxies. Passing to barred galaxies, the phase space structure and the role of chaos in the corotation region are analyzed. This is accomplished by a summary of the modern theory of invariant manifolds of unstable periodic orbits in the vicinity of L_1 or L_2 , which can interpret the generation of spiral patterns by chaotic orbits beyond corotation. Some additional topics, potentially important for disc dynamics, are briefly commented.

1 Introduction

This is a continuation of the lecture notes published in [30] (called hereafter Part I). Part I dealt with general concepts of galactic dynamics, placing particular emphasis on those topics where *stellar dynamics* and *celestial mechanics* share common starting points and/or methods of approach. Furthermore, detailed reference was made to the dynamics of the so-called ellipsoidal components of galaxies, i.e. i) elliptical galaxies and ii) the dark haloes and the bulges of disc galaxies. On the other hand, the dynamics of bars and spiral arms were only occasionally referred to in part I. This will be remedied in the present article.

It is necessary to repeat here, as in Part I, that the present notes are not a review on the subject, but only an introduction emphasizing how to employ concepts of dynamics in the study of galactic discs. Excellent textbooks are now available in the literature (for example [6–8, 24]) which contain more advanced material, as well as a treatment of other related topics, like gas dynamics, star formation and chemical evolution, whose significance in understanding the structure and evolution of galactic discs has proved to be essential in recent years. Here, instead, we discuss only a limited (and classical) list of topics regarding the stellar dynamics of discs. We insist, however, on giving sufficient details for self-contained understanding. For example, we shall demonstrate how the orbital version of the so-called density wave theory, which establishes the possibility of creating a density wave by stars moving in elliptic orbits with coherent phases (the so-called ‘precessing ellipses’ flow [43]) follows from implementing the general principles of resonant Hamiltonian perturbation theory in the case of galactic resonances [15]. This implementation has, in fact, many analogies with the theory of mean motion

^a e-mail: cefthim@academyofathens.gr

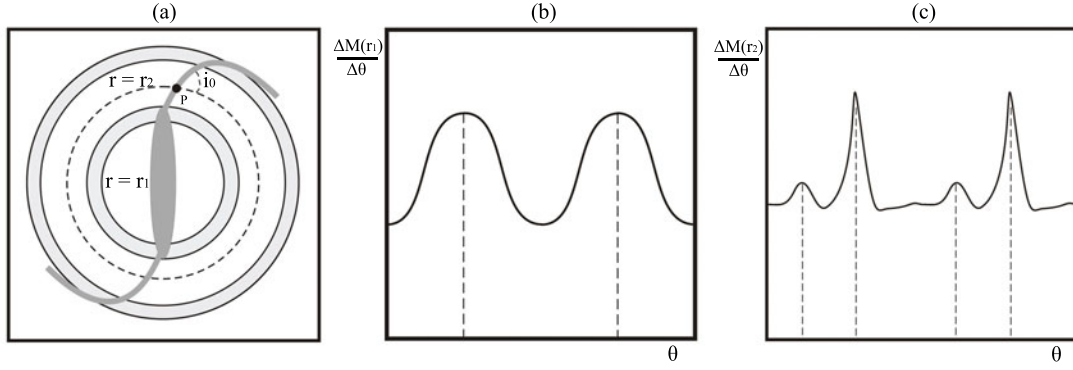


Fig. 1. (a) Face-on view (schematic) of a disc galaxy with bar and spiral arms. (b) The function $\frac{\Delta M}{\Delta \theta}$ in a ring at $r = r_1$. (c) The same function at $r = r_2$.

resonances in Celestial Mechanics [63]. In the same spirit, we discuss how the concept of invariant manifolds of unstable periodic orbits leads to an interpretation of the spiral arms in barred galaxies, while the same formulation in Celestial Mechanics has led to a theory on low-cost transfers via unstable Lagrangian points (e.g. of the Earth-Moon system).

In section 2 we introduce some terminology and discuss the structure of the potential – density field, as well as kinematics of disc galaxies in general. Then, the concept of *self-consistency*, as well as the framework of density wave theory are discussed. Section 3 focuses on the orbits and phase-space structure in disc galaxies, as well as on the orbital version of density wave theory, which is based on resonant Hamiltonian perturbation theory. Then, we discuss vertical resonances as well as the role of chaos in strongly nonlinear models of barred-spiral galaxies. Section 4 discusses a modern topic, namely spiral arms sustained by chaotic orbits, due to the invariant manifolds of families of unstable periodic orbits in the corotation region. Finally, in section 5 we briefly present topics that have attracted current interest, namely i) proposed methods for observational determination of pattern speeds, ii) live haloes and secular evolution of disc galaxies, and iii) multiple pattern speeds.

2 Potential-density field in rotating disc galaxies

2.1 Fourier decomposition of the surface density field

Let us consider a disc galaxy which contains a combination of non-axisymmetric features (e.g. a bar-spiral structure, Fig. 1, schematic). Figure 1 represents the ‘face-on’ view of such a disc. Photometric observations yield a surface brightness map (amount of light per unit area of the disc), which, assuming a constant mass-to-light ratio, represents a surface mass density $\Sigma(r, \theta)$ (in cylindrical coordinates). All disc morphological features are imprinted in the form of the function $\Sigma(r, \theta)$. It is customary to analyze these features in terms of the Fourier decomposition:

$$\Sigma(r, \theta) = \sum_{m=0}^{\infty} \Sigma_m(r) \cos [m\theta - \phi_m(r)]. \quad (1)$$

The functions $\Sigma_m(r)$ are to be calculated by the following procedure: If we fix the radial distance in a small annulus $r_1 - \Delta r/2 \leq r < r_1 + \Delta r/2$, we may think of $\Delta M(\theta; r_1)/\Delta \theta = \Sigma(r_1, \theta)r_1\Delta r$ as the angular mass (or light) distribution, as function of θ , in the annulus. The coefficients $\Sigma_m(r_1)$ and $\phi_m(r_1)$ are then given by:

$$\Sigma_m(r_1) = \frac{\sqrt{A_m^2(r_1) + B_m^2(r_1)}}{r_1 \Delta r}, \quad \phi_m(r_1) = \tan^{-1} \left[\frac{B_m(r_1)}{A_m(r_1)} \right] \quad (2)$$

where

$$A_m(r_1) = \frac{1}{\pi} \int_0^{2\pi} \frac{\Delta M(\theta; r_1)}{\Delta \theta} \cos(m\theta) d\theta, \quad B_m(r_1) = \frac{1}{\pi} \int_0^{2\pi} \frac{\Delta M(\theta; r_1)}{\Delta \theta} \sin(m\theta) d\theta$$

for $m \neq 0$, or

$$A_0(r_1) = \frac{1}{2\pi} \int_0^{2\pi} \frac{\Delta M(\theta; r_1)}{\Delta \theta} d\theta, \quad B_0(r_1) = 0.$$

Repeating this procedure for many consecutive narrow rings of radii $r_i = (i - 0.5)\Delta r$, $i = 1, 2, \dots, i_{max}$ we determine observationally the form of all functions $\Sigma_m(r)$, $\phi_m(r)$ using the information from the sampling points r_i . For orbital studies, it is necessary to fit the functions $\Sigma_m(r)$, $\phi_m(r)$ by some analytic formula, allowing for a representation of $\Sigma(r, \theta)$ by a smooth model. This, in turn, yields a smooth potential, whose derivatives are the forces acting upon stars. A similar technique can be applied in N-body simulations (see also subsection 2.2 below).

We now briefly discuss the physical meaning of the functions $\Sigma_m(r)$, $\phi_m(r)$ for different m .

For $m = 0$ we have $\phi_0(r) = 0$. $\Sigma_0(r)$ represents the *axisymmetric background* of the surface density of the disc. This is practically always the leading component in the Fourier decomposition (1), i.e. as a rule galactic discs are nearly axisymmetric objects within which some non-axisymmetric features appear (i.e. bar or spiral arms) with amplitude smaller than that of the axisymmetric background.

Consider now the $m = 2$ terms. If we fix a value of r , say $r = r_1$, and approximate $\Sigma(\theta; r_1)$ by the sum of the first two even Fourier terms

$$\Sigma(\theta; r_1) \approx \Sigma_{02}(\theta; r_1) = \Sigma_0(r_1) + \Sigma_2(r_1) \cos[2\theta - \phi_2(r_1)] \quad (3)$$

we can obtain a clear picture of the graph of the function $\Sigma_{02}(\theta; r_1)$, shown schematically in Fig. 1b, by simple visual inspection of Fig. 1a, where we have assumed that the circle $r = r_1$ intersects a bar. Setting $\phi_2(r_1) = 0$ (or π), the surface density rises to a peak at the angles $\theta = \pi/2$ and $\theta = 3\pi/2$ which, according to Fig. 1a, are the angles where the circle $r = r_1$ intersects the bar's major axis. At these points the surface density exhibits a local maximum. If this maximum has a relative amplitude ϵ_{bar} with respect to the axisymmetric background of the disc, then $\Sigma_2(r_1) = \epsilon_{bar} \Sigma_0(r_1)$. We thus see that $\Sigma_2(r_1)$ yields the amplitude of the bi-symmetric perturbation at the distance r_1 , while $\phi_2(r_1)$ yields the orientation of the structure which produces this perturbation (e.g. the bar in Fig. 1a). Since, by definition, a bar is an object with well-defined major axis, we expect $\phi_2(r)$ to be a constant function of r for all radii where the bar is observed. On the other hand, spiral arms correspond to a function $\phi_2(r)$ varying monotonically with r . Observations suggest some preferential forms of the function $\phi_2(r)$ in this case. A well studied form is that of logarithmic spirals:

$$\phi_2(r) = \frac{2 \ln(r/r_0)}{\tan i_0} + \phi_{20} \quad (4)$$

where i_0 , r_0 and ϕ_{20} are constants.

The locus of maxima of the $m = 2$ component of the surface density field for some choice of $\phi_2(r)$ is shown schematically in Fig. 1a as the spiral structure starting after $r = r_0$, where r_0 is taken at the end of the bar. The maxima of the spiral are found by taking all pairs of points $(x_s, y_s) = (r_s \cos \theta_s, r_s \sin \theta_s)$ where $\theta_s = \phi_2(r_s)/2$ or $\theta_s = \phi_2(r_s)/2 + \pi$. The angle i_0 in Eq. (4) is called *pitch angle*. It is the angle formed at, say, a point P on the spiral between the spiral and the tangent of the circle passing through P . The spiral is called *trailing* if $dr_s/d\theta_s < 0$ ($i_0 < 0$) in Eq. (4) and *leading* if $dr_s/d\theta_s > 0$ ($i_0 > 0$), when θ_s is taken to increase in the same sense as pattern rotation (counterclockwise in Fig. 1a). Observations indicate that most galaxies have trailing spiral arms. A theoretical argument for why this should be so, based on the study of torques and the outward transfer of angular momentum on the disc as a spiral mode grows, has been offered by Lynden-Bell and Kalnajs [52].

In most 'grand design' spiral galaxies, the $m = 2$ component dominates over all other even components of the Fourier decomposition (1), except $\Sigma_0(r)$. Nevertheless, the remaining components should not be considered of negligible importance. Their role is to describe deformations of the form of the spirals from a purely bi-symmetric shape. For example, let us assume (as in Fig. 1a) that the outermost parts of the spiral arms tend to a more rectangular shape than

what would be predicted by the logarithmic law (4). We now see that the circle $r = r_2$ intersects the locus of maximum spiral density at two points, but it also approaches close to it at two more points. This may cause the angular density function $\Delta M(\theta; r_2)/\Delta\theta$ to have two more local maxima (smaller than the main ones, Fig. 1c). Its Fourier decomposition will then result in a significant $m = 4$ component superposed to the (already important) $m = 2$ component.

As regards the *odd* components $m = 1, 3, \dots$, these describe deviations of the pattern from symmetry with respect to the center. Of particular interest is the $m = 1$ component. This is the main component of lopsided structures which could result, for example, in cases where the disc galaxy interacts with a satellite galaxy. Early studies tended to disregard odd components as of smaller importance, but modern observations indicate that in particular galaxies they might play a key role in dynamics.

2.2 Potential

Assume that the Fourier decomposition (1) for a particular disc galaxy is known. Can we infer from it the gravitational potential due to the *total* mass distribution in the galaxy? This appears at first a trivial question, since a mass distribution (density function $\rho(\mathbf{r})$) corresponds to a unique potential $V(\mathbf{r})$ derived through the solution of $\nabla^2 V = 4\pi G\rho$. However, here we have assumed that we only know the surface density $\Sigma(r, \theta)$ instead of the spatial one $\rho(\mathbf{r})$. Thus, no conclusions can be drawn on the mass distribution unless extra assumptions are made on the vertical structure of the disc, which will allow us to pass from Σ to ρ . Furthermore, there is overwhelming evidence that disc galaxies are embedded into *dark haloes* (see [7]), whose contribution to the total mass distribution can be inferred only indirectly by *kinematic* data, i.e. by measurements of the velocities of stars (or gas) in the disc.

In view of the above, a process has to be devised through which to transform available disc photometric and kinematic data into estimates of the *total* gravitational potential. A frequently used process is to attempt splitting the mass distribution into distinct components (e.g. bulge, disc, bar, spiral arms, dark halo) and to use a fitting model for each component using either empirical laws or an inverse processing of observations. For example, let us consider the decomposition of a system into i) bulge, ii) disc, iii) dark halo. We assume that photometric data are available for the bulge and disc (which contain information on the non-axisymmetric patterns as well), and kinematic data are available for the disc. We may then proceed as follows:

i) *Isolate and fit the bulge component.* This can be done, for example, by assuming that the radial profile of the surface brightness of the disc follows an exponential law, namely $\overline{\Sigma}_d(r) = \Sigma_0 \exp(-\epsilon_d r)$ (overline denotes averaging of the disc surface brightness $\Sigma_d(r, \theta)$ with respect to θ). This assumption is usually verified in the outer parts of a disc, while in the inner parts we may observe $\overline{\Sigma}_d(r)$ rising beyond the exponential law. If we use the outer part to find fitting values of the parameters Σ_0 and ϵ_d , we can identify the non-zero residual $\Sigma_b(r) = \overline{\Sigma}_d(r) - \Sigma_0 \exp(-\epsilon_d r)$ at inner radii as the surface brightness profile of the bulge. The latter can then be fitted by an empirical law. We frequently use *Sersic's law*

$$\Sigma_b(r) = \Sigma_e \exp \left[-b_n \left\{ \left(\frac{r}{r_e} \right)^{1/n} - 1 \right\} \right] \quad (5)$$

which is a generalization of de Vaucouleur's law (see Part I, subsection 11.3.1). The new fitting parameters are the index n , and the constants b_n, r_e, Σ_e . The index n determines the steepness of the profile ($n = 4$ for elliptical galaxies while $n \sim 1-2$ for fast rotating bulges); r_e sets the scale of the bulge and b_e is regulated so that half of the total light is contained within the disc of radius r_e . Finally, Σ_e determines the total luminosity, or total mass in the bulge. If some simple 3D geometry, e.g. spherical, is assumed for the bulge, from $\Sigma_b(r)$ we can deduce the radial density profile (by inverting Eq. (11.10) of Part I), and hence the gravitational potential of the bulge.

ii) *Disc Fourier decomposition.* Having a model for $\Sigma_b(r)$ in hand, we now proceed in the inverse way, i.e. subtract $\Sigma_b(r)$ from the total surface brightness $\Sigma(r, \theta)$. The remaining function

$\Sigma_d(r, \theta) = \Sigma(r, \theta) - \Sigma_b(r)$ represents the pure disc with all patterns embedded in it. By Fourier-decomposing $\Sigma_d(r, \theta)$ (Eq. (1)), we find the functions $\Sigma_m(r)$, $\phi_m(r)$, or, equivalently, $A_m(r)$, $B_m(r)$ (we use the same symbols as in Eq. (2), but A , B are now normalized, i.e. divided by $r\Delta r$, and refer only to the residual disc surface density Σ_d). In terms of these functions, the Fourier decomposition reads:

$$\Sigma_d(r, \theta) = \sum_{m=0}^{m_{max}} [A_m(r) \cos(m\theta) + B_m(r) \sin(m\theta)] \quad (6)$$

(m_{max} is taken high enough to reach sufficient convergence of the Fourier series, typically $m_{max} \sim 10$).

To derive the gravitational potential of the disc, the Fourier decomposition (6) must be supplemented by an ansatz on the thickness profile $Z_d(z)$, defined by $\rho_d(r, \theta, z) = Z_d(z)\Sigma_d(r, \theta)$, where ρ_d is the disc's spatial density. Observations are usually rather inconclusive on this matter, unless the disc is observed 'edge-on', in which case, however, we lose information on $\Sigma_d(r, \theta)$. Thus, assumptions must be made for the form of $Z_d(z)$ as well. Some frequent choices are: a) the 'razor-thin' disc $Z_d(z) = \delta(z)$, and b) the exponential law $Z_d(z) \propto e^{-k_z|z|}$, where $k_z^{-1} > 0$ sets the exponential vertical scale of the disc. A 'smoothed' version of the exponential law is the 'hyperbolic secant law' $Z_d(z) \propto \text{sech}^2(k_z z)$, which yields substantially non-zero values only at $z < k_z^{-1}$. Independently of what has been the choice of $Z_d(z)$, the disc potential can now be determined by the theory of *potential-density pairs*. This is equivalent to the 'self-consistent field' method discussed for 3D galaxies in Part I, subsection 11.5.1, and the formulae for discs are also similar, except that they are expressed in cylindrical rather than spherical coordinates. Care must be taken of the fact that making a choice on how to represent $Z_d(z)$ via basis functions affects also the way the radial functions $A_m(r)$, $B_m(r)$ are represented. Consider, for example, the razor-thin disc $Z_d(z) = \delta(z)$. Green function theory yields that

$$\left(\frac{d^2}{dz^2} - k^2 \right) e^{-k|z|} = -2k\delta(z). \quad (7)$$

Since Poisson's equation $\nabla^2 V = 4\pi G\rho$ is linear, we may find separately the potential term generated by each term of the Fourier decomposition of the density, and the sum of these terms yields the Fourier decomposition of the total potential. Let

$$\rho_{c,m}(r, \theta, z) = \delta(z)A_m(r) \cos(m\theta), \quad \rho_{s,m}(r, \theta, z) = \delta(z)B_m(r) \sin(m\theta). \quad (8)$$

The following result is due to Toomre [77]: if we express the correspondent Fourier terms of the potential

$$V_{c,m}(r, \theta, z) = C_m(r, z) \cos(m\theta), \quad V_{s,m}(r, \theta, z) = D_m(r, z) \cos(m\theta) \quad (9)$$

via the following *Hankel transforms*

$$C_m(r, z) = -2\pi G \int_0^\infty \tilde{A}_m(k) e^{-k|z|} J_m(kr) dk, \quad D_m(r, z) = -2\pi G \int_0^\infty \tilde{B}_m(k) e^{-k|z|} J_m(kr) dk \quad (10)$$

where $J_m(kr)$ are Bessel functions and $\tilde{A}_m(k)$, $\tilde{B}_m(k)$ are unknown functions to be specified, then Poisson's equation $\nabla^2 V_{c,m} = 4\pi G\rho_{c,m}$, or $\nabla^2 V_{s,m} = 4\pi G\rho_{s,m}$, together with Eq. (7), leads to

$$A_m(r) = \int_0^\infty \tilde{A}_m(k) J_m(kr) k dk, \quad B_m(r) = \int_0^\infty \tilde{B}_m(k) J_m(kr) k dk. \quad (11)$$

Thus, the functions $\tilde{A}(k)$, $\tilde{B}(k)$ are directly linked to the (presumably known) functions $A_m(r)$, $B_m(r)$, and they can be computed by inverting Eqs. (11), namely

$$\tilde{A}_m(k) = \int_0^\infty A_m(r) J_m(kr) r dr, \quad \tilde{B}_m(k) = \int_0^\infty B_m(r) J_m(kr) r dr. \quad (12)$$

This completes the determination of the gravitational potential not only on the disc plane, but also for any other value of z . On the disc ($z = 0$), in particular, the Fourier decomposition of the potential corresponding to the surface density (6) reads:

$$V_d(r, \theta) = -2\pi G \sum_{m=0}^{m_{max}} \left\{ \left[\int_0^\infty dk J_m(kr) \int_0^\infty A_m(r') J_m(kr') r' dr' \right] \cos(m\theta) + \left[\int_0^\infty dk J_m(kr) \int_0^\infty B_m(r') J_m(kr') r' dr' \right] \sin(m\theta) \right\}. \quad (13)$$

If, now, $Z_d(z)$ is no longer taken to be the $\delta(z)$ function, but it is the exponential, or the hyperbolic secant function, we can derive analogous expressions by replacing $e^{-k|z|}$ in Eq. (10) by a different expression depending on the choice of $Z_d(z)$ (see [27], table 1; note however a missing factor $-1/2k$ in the potential factor of the table's first row).

The use of a particular set of basis functions, like the Bessel functions, is by no means unique in representing the disc potential, and other basis sets, or other methods can be devised. A 'numerical' method consists of using the fundamental expression

$$V_d(r, \theta) = -G \int_{r'=0}^\infty \int_{\theta'=0}^{2\pi} \frac{g(r', \theta') \Sigma_d(r', \theta') r' dr' d\theta'}{\sqrt{r^2 + r'^2 - 2rr' \cos(\theta - \theta')}} \quad (14)$$

where $g(r', \theta')$ gives the factor weighting the mass distribution at (r', θ') due to the chosen vertical profile of the disc. The kernel function $g(r', \theta') r' / \sqrt{r^2 + r'^2 - 2rr' \cos(\theta - \theta')}$ is first Fourier-analyzed, and the Fourier convolution theorem is invoked to find the Fourier transform of V_d from the known Fourier transforms of the kernel function and of $\Sigma_d(r', \theta')$. This method has been used to obtain potential models for galaxies observed in the near infrared (see [62]; for a recent implementation see [41]), since the distribution of light in IR wavelengths, coming from old stars, is considered as a better tracer of the total mass distribution in the galaxy.

The final missing term in the potential regards the contribution of the dark halo. This is fitted by exploiting *kinematic* data. That is, if the *rotation curve* is known (circular velocity of stars v_c as a function of distance from the center; this can be inferred from measuring the redshifts of well known spectral lines at various distances from the center), then, assuming e.g. a spherical halo, we have

$$v_c^2(r) = r \left(\frac{dV_{halo}(r)}{dr} + \frac{dV_b(r)}{dr} + \frac{dV_d(r)}{dr} \right) \quad (15)$$

whereby the halo potential can be computed after subtracting the contributions from the bulge and from the disc, that were estimated in previous steps, from $v_c^2(r)$ that was measured by kinematic data. The halo potential can again be fitted by a simple model. Examples are the Plummer, isochrone, Dehnen, Hernquist or Navarro-Frenk-White (NFW) models (see subsections 2.2 and 2.3 of [7] and references therein for an account of particular features of these models). It should be stressed that simple models can be invented for an axisymmetric disc as well, which fit the component $A_0(r)$ of the decomposition (1). Examples are the Kuzmin (Toomre), Freeman and Miyamoto models [7]. The features of all the above models are reviewed in practically all books on galactic dynamics, we shall thus not repeat this exposition here. Instead, we give the explicit formulae for a set of potential functions that we use below in some numerical demonstrations; these are convenient as simple models for basic dynamical studies (see e.g. [44]). For spheroidal components, i.e. the bulge or the dark halo, we use the Plummer model:

$$V_0(r) = -\frac{GM}{\sqrt{b^2 + r^2}} \quad (16)$$

which corresponds to a mass distribution $\rho(r) = (3M/4\pi b^3)(1 + r^2/b^2)^{-5/2}$ softened in the center by a length parameter b . For the axisymmetric component of the disc, we use [33]

$$V_d(r) = -\Sigma_0 \pi r \left(I_0 \left(\frac{1}{2} \epsilon_d r \right) K_1 \left(\frac{1}{2} \epsilon_d r \right) - I_1 \left(\frac{1}{2} \epsilon_d r \right) K_0 \left(\frac{1}{2} \epsilon_d r \right) \right), \quad (17)$$

which corresponds to an ‘exponential’ disc, i.e. with surface density law given by $\Sigma_d(r) = \Sigma_0 \exp(-\epsilon_d r)$. The functions I_0 , I_1 , and K_0 , K_1 are modified Bessel functions of the first and second kind respectively, and ϵ_d is an inverse length parameter which sets the exponential scale length of the disc. For an $m = 2$ representation of spiral arms, we use the ‘logarithmic spiral’ model

$$V_s(r, \theta) = A_0 r \exp(-\epsilon_s r) \cos\left(2\theta - \frac{2 \ln(r/a)}{\tan(i_0)}\right) \quad (18)$$

where A_0 sets the amplitude of spiral perturbation, ϵ_s^{-1} sets its radial exponential scale length, i_0 is a constant pitch angle, and a is a length parameter determining, together with i_0 , the ‘radial wavenumber’, i.e. the inverse of the distance between two successive maxima of the spiral density encountered along the same direction (the latter is given by $ae^{2\pi \tan i_0}$). Finally, in the case of bars we use the so-called $n = 2$ *Ferrer bar* model [32]. The density is stratified in concentric ellipsoids

$$\rho_b(x, y, z) = \frac{105GM_b}{32\pi abc} \left(1 - \frac{y^2}{a^2} - \frac{x^2}{b^2} - \frac{z^2}{c^2}\right)^2, \quad a > b > c, \quad (19)$$

with

$$1 - \frac{y^2}{a^2} - \frac{x^2}{b^2} - \frac{z^2}{c^2} \geq 0.$$

In the above formulae, a, b, c are the bar’s long, intermediate and short semi-axes respectively, and M_b is the total mass in the bar. Calculation of the potential can be more efficiently worked out in ellipsoidal coordinates (see [13] or [59] for explicit formulae). The final result is

$$V_{bar}(x, y) = -\frac{105GM_b}{96} \left(3(2W_{110}x^2y^2 - W_{120}x^4y^2 - W_{210}x^2y^4 - W_{100}y^2 + W_{020}x^4 + W_{200}y^4 - W_{010}x^2) + W_{000} - W_{030}x^6 - W_{300}y^6 \right) \quad (20)$$

where

$$W_{000} = \frac{2}{\sqrt{a^2 - c^2}} F(\phi, k), \quad W_{100} = \frac{2}{(a^2 - b^2)\sqrt{a^2 - c^2}} [F(\phi, k) - E(\phi, k)],$$

$$W_{001} = \frac{2}{b^2 - c^2} \sqrt{\frac{b^2 + l}{(a^2 + l)(c^2 + l)}} - \frac{2}{(b^2 - c^2)\sqrt{a^2 - c^2}} E(\phi, k),$$

$$W_{010} = \frac{2}{\Delta} - W_{100} - W_{001}, \quad W_{110} = \frac{W_{010} - W_{100}}{a^2 - b^2}, \quad W_{011} = \frac{W_{001} - W_{010}}{b^2 - c^2},$$

$$W_{101} = -\frac{W_{100} - W_{001}}{a^2 - c^2}, \quad W_{200} = \frac{1}{3} \left(\frac{2}{\Delta(a^2 + l)} - W_{110} - W_{101} \right),$$

$$W_{020} = \frac{1}{3} \left(\frac{2}{\Delta(b^2 + l)} - W_{011} - W_{110} \right), \quad W_{120} = \frac{W_{020} - W_{110}}{a^2 - b^2},$$

$$W_{201} = -\frac{W_{200} - W_{101}}{a^2 - c^2}, \quad W_{210} = \frac{W_{110} - W_{200}}{a^2 - b^2}, \quad W_{021} = \frac{W_{011} - W_{020}}{b^2 - c^2},$$

$$W_{300} = \frac{1}{5} \left(\frac{2}{\Delta(a^2 + l)^2} - W_{210} - W_{201} \right), \quad W_{030} = \frac{1}{5} \left(\frac{2}{\Delta(b^2 + l)^2} - W_{021} - W_{120} \right),$$

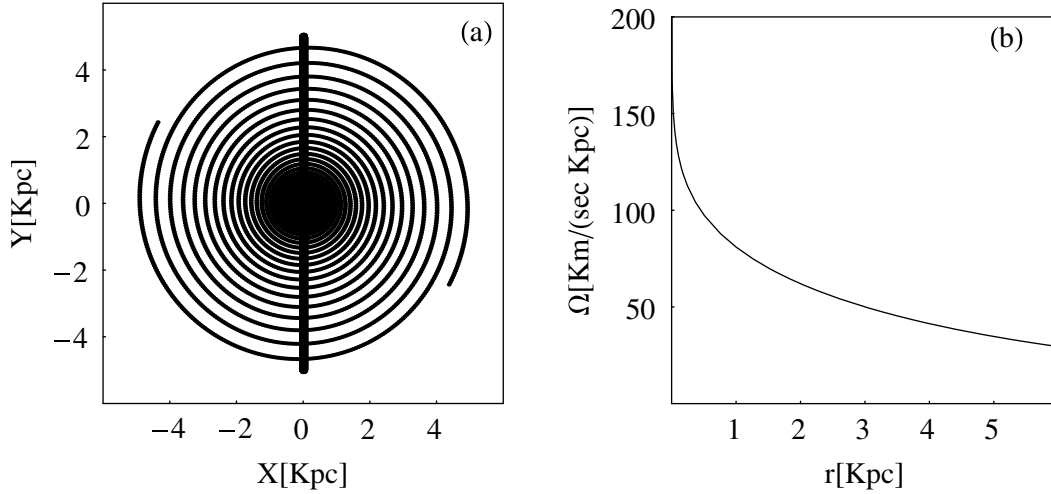


Fig. 2. (a) Fast winding of a bar-like structure embedded in an exponential disc. Thick dots represent the initial positions of stars along a bar-like structure. The tightly wound spirals are the locus of the positions of the same stars after 1Gyr. (b) The angular velocity Ω of stars in circular orbits as function of the distance r for the exponential disc model (Eq. (17)) and parameters $\Sigma_0 = 10^9 M_\odot/\text{Kpc}^2$, $\epsilon_d = 1/(3 \text{ Kpc})$.

and

$$\phi = \arcsin \left(\sqrt{\frac{a^2 - c^2}{a^2 + l}} \right), \quad \Delta = \sqrt{(a^2 + l)(b^2 + l)(c^2 + l)}, \quad k = \frac{a^2 - b^2}{a^2 - c^2}$$

$$l = 0 \quad \text{if} \quad \frac{y^2}{a^2} + \frac{x^2}{b^2} \leq 1$$

$$l = \frac{1}{2} \left(-(a^2 + b^2) + r^2 + \sqrt{[-(a^2 + b^2) + r^2]^2 + 4[-a^2 b^2 + a^2 x^2 + b^2 y^2]} \right) \quad \text{if} \quad \frac{y^2}{a^2} + \frac{x^2}{b^2} > 1.$$

The functions $F(\phi, k)$, $E(\phi, k)$ are elliptic functions of the first and second kind respectively.

Let us summarize progress so far: starting by the surface brightness profile of a disc, which is determined by photometric observations, we i) Fourier-analyzed it, and ii) transformed it into a model for the gravitational potential. This can be given also in form of a Fourier transform, or by simple fitting models describing distinct components (dark halo, bulge, spiral arms, bar). We now discuss the problem of *self-consistency*, which refers to the reproduction of the various observed patterns by a combination of orbits of stars acted upon by the forces due to their self-gravitational potential. This is also linked to the question of determination of the *pattern speed*, i.e. the speed by which a *density wave* rotates on the disc.

2.3 Pattern speed and the concept of density wave

Consider, for simplicity, an exponential disc (potential given by Eq. (17)), wherein a bar-like non-axisymmetric perturbation is embedded (Fig. 2a). We shall examine the following question: is it possible that this bar maintains its shape and rotation as a rigid body, if we know in advance that the stars composing the bar move in *nearly circular* orbits around the center of the galaxy? To answer this question, it suffices to determine the angular velocity $\Omega(r)$ of stars in circular orbits. This is given by $\Omega^2(r)r = |F_0(r)|$ or

$$\Omega(r) = \left(\frac{1}{r} \frac{\partial V_0(r)}{\partial r} \right)^{1/2}. \quad (21)$$

Fig. 2b shows the dependence of Ω on r for the exponential disk model (17) with parameters as in the caption. We see that Ω decreases with r , so that the stars closer to the center rotate faster than the stars further away. As a result, if the bar at a given snapshot was composed by stars moving in nearly-circular orbits, the structure would start immediately *winding*. Using the useful scaling $1 \text{ Km sec}^{-1} \text{ Kpc}^{-1} \simeq 1 \text{ Gyr}^{-1}$, and observing that the values of Ω are $\sim 10^2 \text{ Gyr}^{-1}$, we conclude that the stars accomplish one revolution in times of order 10^8 yr . This means that after only a small number of orbital periods (say after $\sim 1 \text{ Gyr}$) the initial bar-like structure would have become a tightly-wound structure, which could no longer represent either the initial bar or some form of spiral arms with pitch angle falling within the range of observations (a few degrees). The conclusion is that non-axisymmetric structures observed in disc galaxies are most likely not *material* structures, i.e. always composed by the same stars, but *density waves*, i.e. local enhancements of the density on the disc that rotate in general much more slowly than the stars. In fact, in the inner parts of bars the stars might always stay inside the bar provided that their orbits deviate considerably from circular. In spiral arms, however, as well as in the outer envelopes of bars, the stars go in and out of the density wave, and the ‘pattern’ is maintained due to detailed balance between the number of stars which enter or leave the pattern at any given moment of time.

In order to establish that a pattern observed in a galaxy is a density wave, it has to be demonstrated that when all stellar motions are properly taken into account, there can be a right proportion of stars whose orbits ‘collaborate’ in creating a non-axisymmetric local enhancement of the density (= pattern). Furthermore, the collective effect of all stellar motions, due to the self-gravity of stars, should be such as to induce a slow precession of the pattern on the disc, with some angular velocity Ω_p . This is called the ‘pattern speed’.

Density wave theory for disc galaxies was conceived by B. Lindblad [48, 50] and comes in two main versions. In the first version, pioneered by Lin and Shu [47] and Kalnajs [42], one considers the effect of collective motions via perturbative solutions of the collisionless Boltzmann equation (Part I; subsection 11.2.3). This is a nonlinear integro-differential equation reading:

$$\frac{\partial f}{\partial t} + \mathbf{v} \frac{\partial f}{\partial \mathbf{r}} - \frac{\partial}{\partial \mathbf{r}} \left(\int d\mathbf{r}' \frac{G \int f d^3 \mathbf{v}'}{|\mathbf{r} - \mathbf{r}'|} \right) \frac{\partial f}{\partial \mathbf{v}} = 0. \quad (22)$$

In the above equation, $f(\mathbf{r}, \mathbf{v}, t)$ is the *distribution function*, i.e. density of stars in phase space. If a plausible model f_0 is chosen for an axisymmetric disc galaxy we may consider a perturbation $f = f_0 + f_1$ describing the density wave on the disc. For f_0 we typically choose a steady-state model $f_0(\mathbf{r}, \mathbf{v})$ (see, for example, [7], Appendix K). Linearizing (22) we then have to solve

$$\frac{\partial f_1}{\partial t} + \mathbf{v} \frac{\partial f_1}{\partial \mathbf{r}} - \frac{\partial}{\partial \mathbf{r}} \left(\int d\mathbf{r}' \frac{G \int f_1 d^3 \mathbf{v}'}{|\mathbf{r} - \mathbf{r}'|} \right) \frac{\partial f_0}{\partial \mathbf{v}} - \frac{\partial}{\partial \mathbf{r}} \left(\int d\mathbf{r}' \frac{G \int f_0 d^3 \mathbf{v}'}{|\mathbf{r} - \mathbf{r}'|} \right) \frac{\partial f_1}{\partial \mathbf{v}} = 0. \quad (23)$$

The terms

$$V_0(\mathbf{r}) = - \int d\mathbf{r}' \frac{G \int f_0 d^3 \mathbf{v}'}{|\mathbf{r} - \mathbf{r}'|}, \quad V_1(\mathbf{r}) = - \int d\mathbf{r}' \frac{G \int f_1 d^3 \mathbf{v}'}{|\mathbf{r} - \mathbf{r}'|}$$

are the gravitational potential terms due to the distribution of mass in the unperturbed and in the perturbed models respectively. In the original work of Lin and Shu [47], instead of solving (23) directly the authors considered solutions of *moments* of this equation which are found by multiplying both sides of Boltzmann equation with progressively higher power combinations of the velocities, and by integrating the result in velocity space. This produces the hierarchy of so-called *Jeans equations* of stellar dynamics which allows one to approximate the form of the distribution function f_1 through its moments in position and velocity state, supplemented by a closure condition which is the stellar dynamical analog of the equation of state in fluid dynamics. However, straightforward solutions of Eq. (23) can also be found (see e.g. [42] or the Appendix K of [7]). These, in general, describe *modes* of the form

$$f_1 = F(\mathbf{r}, \mathbf{v}) e^{i(\omega t - m\theta + \Phi_m(r))} \quad (24)$$

which describe rotating m -fold spiral arms. The value $\Omega_p = \text{Re}(\omega)/m$ is the pattern speed of these arms, since the solution ‘repeats’ itself periodically in time after rotation by $2\pi/m$. However, special care has to be taken regarding whether ω contains also an imaginary part. In fact, we need $\text{Im}(\omega) < 0$ to *generate* spiral arms by instability, but if this is so for long times, the spiral arms will be disrupted. On the other hand, velocity dispersion as well as nonlinear effects may stabilize the growth of the spiral mode (see [6] for an account of the so-called ‘modal’ theory). This has to be self-consistently taken into account.

In the second version, called the *orbital approach*, the emergence of a self-consistent spiral pattern is examined as an effect of the superposition of many nearly periodic orbits with such orientations so as to produce a local enhancement of the density. The orbits have to be elliptic and their apsides have to follow the phase $\Phi(r) = \phi_2(r)/2 (+\pi)$ of the spiral arms. The ellipticity of the orbits is called ‘forced ellipticity’ by Kalnajs [43] (see also [52]). A comment is in order at this point. The famous density wave sketch by ‘precessing ellipses’ of Kalnajs [43], which has been reproduced in many textbooks and even in Wikipedia, has been called a ‘kinematic wave’ (e.g. [7], subsection 6.2.1), meaning a wave formed by elliptic orbits moving under the axisymmetric potential only, i.e. for which we ignore their self-gravity. Such periodic orbits certainly exist by definition close to the ILR, and they were originally considered by Lindblad [49]. However, these orbits *cannot* form density waves unless the combination $\Omega - \kappa/2$ (where, as defined below, Ω and κ are the azimuthal and epicyclic frequencies) is strictly constant along a disc. If, instead, $\Omega - \kappa/2$ varies, the above orbits become more elongated as we move outwards along the disc. Instead, Kalnajs considers those orbits which would be *circular* in the axisymmetric potential and acquire their (‘forced’) ellipticity only due to a triggered spiral perturbation. Thus, the density waves considered by Kalnajs are not kinematic. In fact, the reason why forced ellipticities arise, and also why the resulting periodic orbits have some preferential orientation so as to support the density wave, has been explained in the framework of resonant Hamiltonian perturbation theory [15]. This task can be viewed, in fact, as a complementary way of solving equation (23) through its *characteristics*, which are the orbits. We will examine resonant perturbation theory in some detail in subsection 3.3, after introducing first some general notions referring to orbital dynamics in disc galaxies in general.

3 Orbital dynamics

3.1 Epicyclic theory – resonances

Under the influence of only the axisymmetric potential term $V_0(r)$, all orbits in the rest frame are rosettes moving in an annulus between a pericentric and an apocentric radius (see Part I, subsection 11.4.1). The angular momentum p_θ is conserved. The Hamiltonian becomes, effectively, of one degree of freedom

$$H_0 = \frac{p_r^2}{2} + \frac{p_\theta^2}{2r^2} + V_0(r), \quad p_\theta = L = \text{const.} \quad (25)$$

The quantity $V_{\text{eff}}(r; p_\theta^2) = p_\theta^2/2r^2 + V_0(r)$ is called effective potential. For a fixed energy E , the pericenter and apocenter radii (r_p , r_a respectively) are defined by the two roots for r of $E = V_{\text{eff}}(r, p_\theta^2)$. These are joined at the radius r_c of the circular orbit which corresponds to the minimum of the effective potential, i.e.

$$-\frac{p_\theta^2}{r_c^3} + \frac{dV_0(r_c)}{dr} = 0. \quad (26)$$

The radial period (= time it takes to go from pericenter to apocenter and back to pericenter) is given by $T_r(E, p_\theta^2) = 2 \int_{r_p(E, p_\theta^2)}^{r_a(E, p_\theta^2)} [2(E - V_0(r)) - p_\theta^2/r^2]^{-1/2} dr$. The quantity $\kappa = 2\pi/T_r$ is called epicyclic frequency. For orbits not far from circular, κ is the frequency of (harmonic) radial

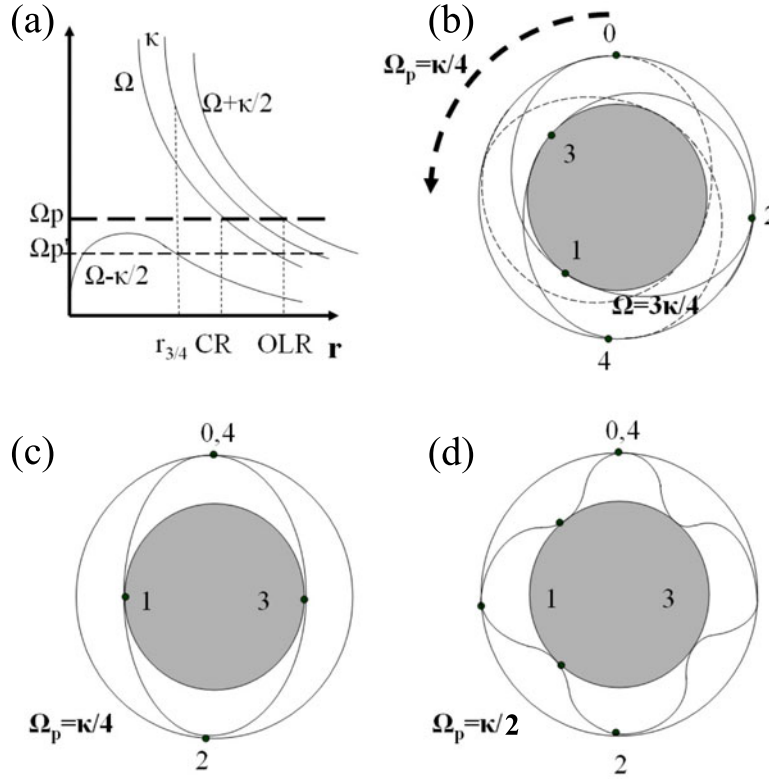


Fig. 3. (a) The form of the functions $\Omega(r)$, $\kappa(r)$ as well as their resonant combinations for ILR and OLR. Depending on the value of Ω_p , we may have none, one, or multiple ILRs. (b) A ‘rosette-like’ (but periodic) orbit in the rest frame with frequencies as indicated. (c) Same orbit as viewed in a frame rotating with $\Omega_p = \Omega/3$. (d) Same orbit as viewed in a frame rotating with a higher speed $\Omega_p = 2\Omega/3$.

oscillation close to the minimum of the effective potential at $r = r_c$. Then

$$\kappa = \sqrt{\frac{d^2 V_{eff}(r_c)}{dr_c^2}} = \sqrt{\frac{d^2 V_0(r_c)}{dr_c^2} + \frac{3}{r_c} \frac{dV_0(r_c)}{dr_c}}. \quad (27)$$

The frequencies $\Omega(r_c)$ and $\kappa(r_c)$ are the basic frequencies of the so-called ‘epicyclic theory’ of orbits. In this theory, nearly circular orbits are described as the composition of two independent motions, namely circular motion of the *guiding center*, with frequency $\Omega(r_c)$, and an oscillation in *both* the radial and angular directions, with frequency $\kappa(r_c)$.

The general form of the functions $\Omega(r)$, $\kappa(r)$ is shown schematically in Fig. 3a. Both frequencies are decreasing functions of r . Consider now, in this schematic example, the distance $r = r_{3/4}$, where we have $\Omega = 3\kappa/4$, or $T_\theta = 4T_r/3$. An epicyclic orbit whose guiding center moves on the circle $r = r_{3/4}$ accomplishes four radial oscillations by the time it completes three full revolutions (Fig. 3b). Assume, now, that we are interested in examining how would the same orbit appear when viewed in a rotating frame with constant speed Ω_p . Let us choose, for example, $\Omega_p = \Omega/3 = \kappa/4$. Fig. 3c shows the orbit (numbered points represent the position of the *same* points as in Fig. 3b, but viewed now in the rotating frame). The key remark is that in the rotating frame, where the orbit appears now closed at point 4, the orbit acquires the shape of an ellipse. In fact, the azimuthal and radial frequencies in the rotating frame, $\Omega - \Omega_p$ and κ respectively, satisfy the resonant condition $\Omega - \Omega_p = \kappa/2$. This is called *inner Lindblad resonance* (ILR). Independently of the particular values of Ω and κ (which change with r), if the ILR condition holds the orbits appear as closed and elongated ellipses in the rotating frame. If an elongated structure (pattern) is embedded in this disc, we can say that the epicyclic orbits near the ILR *support* the structure, i.e., their form is such as to contribute

to the enhancement of the density at precisely those parts of the disc where the local density maximum (due to the structure) lies.

Fig. 3d shows again the same orbit as in Fig. 3b, but viewed now in a frame rotating with a higher speed $\Omega_p = 2\Omega/3 = \kappa/2$. Again, the appearance of the orbit in this new frame has changed, and the orbit appears now more ‘rectangular’, due to the fact that there are four radial oscillations completed within one revolution with respect to this new rotating frame. Thus, we conclude that, depending on our assumption for the value of Ω_p , the same orbit can be seen to support different morphological features of the galaxy. It should be stressed again that Ω_p is neither known in advance (as e.g. the mean motion of Jupiter in Celestial Mechanics), nor possible to infer from observations without extra dynamical assumptions. Thus, ‘guessing’ the correct value of Ω_p is a highly non-trivial task involving dynamical modeling.

In general, we define resonances in a frame rotating with speed Ω_p by

$$\Omega - \Omega_p = \frac{n}{m} \kappa. \quad (28)$$

For $n = 1$, and $m = 2$ we obtain the inner Lindblad resonance (ILR), while for $n = -1$ and $m = 2$ we obtain the outer Lindblad resonance (OLR). On the other hand, for $n = 1$ and $m \rightarrow \infty$ we obtain *corotation*, defined by $\Omega = \Omega_p$. It is important to remember that since Ω and κ are functions of r , while Ω_p is a fixed number, the location of all resonances depends on Ω_p . In particular, depending on how large Ω_p is we may have zero, one or two ILRs. Since Ω_p is in fact unknown, we may think in the reverse way, i.e. identify morphological features on the disc which could be accounted for by the dynamical influence of some resonance. Then, we have a hint of where the resonance lies in the particular galaxy, and this indirectly determines Ω_p . A basic goal of dynamics thus becomes to identify ‘signatures’ of particular resonances that could lead to observational indications on the position of the resonance on the disc. This approach is analogous to the study of resonances (mean motion or secular) in celestial mechanics. A difference, however, is that in galactic dynamics the basic frequencies are comparable in size so that most resonances cannot be treated as adiabatic.

3.2 Perturbation theory at resonances

Let r_c be a fixed radius. We shall consider epicyclic motions around r_c as viewed in a rotating frame with angular speed Ω_p , when a non-axisymmetric ‘pattern’ rotates on the disc with the same speed. In the rotating frame, the pattern appears always at the same place. Thus, the time-dependence of the gravitational potential disappears at the cost of introducing centrifugal and Coriolis forces. The total Hamiltonian can be written as:

$$H(r, \theta, p_r, p_\theta) = \frac{p_r^2}{2} + \frac{p_\theta^2}{2r^2} - \Omega_p p_\theta + V_0(r) + \sum_{m=-\infty}^{\infty} V_m(r) e^{im\theta}. \quad (29)$$

In (29), θ denotes the angular position of an orbit with respect to a fixed axis ($\theta_0 = 0$) in the *rotating frame*, i.e. the axis co-rotates with the pattern. However, p_θ is the angular momentum per unit mass along an orbit (with respect to the center) as measured in the *rest frame* (this, despite the fact that (29) describes motions in the rotating frame; p_θ defined as such preserves its canonical conjugate relation with θ). Furthermore, the decomposition of the potential in complex Fourier series is equivalent to the decomposition (9) provided that $\text{Im}(V_m) = -\text{Im}(V_{-m})$.

We now define action-angle variables for the above Hamiltonian to describe local motions around r_c . For a canonical pair (q, p) to be angle and action variables, q must be an angle, i.e., the Hamiltonian must be periodic in q . According to this definition, the pair (θ, p_θ) in (29) is already an angle-action pair. A second pair (θ_r, J_r) can be defined for epicyclic oscillations via the canonical transformation $(r, p_r) \rightarrow (\theta_r, J_r)$ given by

$$r - r_c = \left(\frac{2J_r}{\kappa(r_c)} \right)^{1/2} \sin \theta_r, \quad p_r = (2\kappa(r_c)J_r)^{1/2} \cos \theta_r. \quad (30)$$

Setting $J_\theta = p_\theta - \Omega(r_c)r_c^2$, the variables (θ, J_θ) are still canonically conjugated.

Substituting the above expressions into Eq. (29), and Taylor-expanding with respect to r_c brings the Hamiltonian to the form (apart from a constant)

$$H(\theta_r, \theta, J_r, J_\theta) = \kappa_c J_r + \frac{1}{2r_c^2} J_\theta^2 + (\Omega_c - \Omega_p) J_\theta + \sum_{|m|+|n|=0}^{\infty} F_{m,n}(J_r, J_\theta) e^{i(n\theta_r + m\theta)} \quad (31)$$

where $\kappa_c \equiv \kappa(r_c)$, $\Omega_c \equiv \Omega(r_c)$. It is readily seen that the coefficients $F_{m,n}$ are of the form

$$F_{m,n}(J_r, J_\theta) = F_{0,m,n}(r_c; J_r) + F_{1,m,n}(r_c; J_r) J_\theta + F_{2,m,n}(r_c; J_r) J_\theta^2$$

where

$$F_{j,0,0} = 0, \quad j = 0, 1, 2$$

$$F_{0,0,n} = \sum_{k=0}^{\infty} \mathcal{B}_{nk} \left[\frac{d^{|n|+2k} V_0(r_c)}{dr_c^{|n|+2k}} + \frac{(-1)^{|n|+2k} (|n| + 2k + 1)! \Omega_c^2}{2r_c^{|n|+2k-2}} \right] \left(\frac{2J_r}{\kappa_c} \right)^{\frac{|n|}{2} + k}, \quad n \neq 0$$

$$F_{1,0,n} = \sum_{k=0}^{\infty} \mathcal{B}_{nk} \left[\frac{(-1)^{|n|+2k} (|n| + 2k + 1)! \Omega_c}{r_c^{|n|+2k}} \right] \left(\frac{2J_r}{\kappa_c} \right)^{\frac{|n|}{2} + k}, \quad n \neq 0$$

$$F_{2,0,n} = \sum_{k=0}^{\infty} \mathcal{B}_{nk} \left[\frac{(-1)^{|n|+2k} (|n| + 2k + 1)!}{2r_c^{|n|+2k+2}} \right] \left(\frac{2J_r}{\kappa_c} \right)^{\frac{|n|}{2} + k}, \quad n \neq 0$$

$$F_{1,m,n} = F_{2,m,n} = 0, \quad n \neq 0, \quad m \neq 0$$

$$F_{0,m,n} = \sum_{k=0}^{\infty} \mathcal{B}_{nk} \left[\frac{d^{|n|+2k} V_m(r_c)}{dr_c^{|n|+2k}} \right] \left(\frac{2J_r}{\kappa_c} \right)^{\frac{|n|}{2} + k}, \quad n \neq 0, \quad m \neq 0$$

and the coefficients \mathcal{B}_{nk} are

$$\mathcal{B}_{nk} = \frac{(-1)^k}{[sgn(n)2i]^{|n|+2k} k! (|n| + k)!}.$$

The key point to notice in Eq. (31) is that the basic frequencies κ_c and $\Omega_c - \Omega_p$ depend on the reference radius r_c in such a continuous way that, as r_c increases, an infinity of exactly resonant combinations $k_1 \kappa_c + k_2 (\Omega_c - \Omega_p) = 0$ are crossed. Even if r_c is such that no exact resonance condition is satisfied, the combination $k_1 \kappa_c + k_2 (\Omega_c - \Omega_p)$ may be of the same smallness as higher order terms in (31) so that r_c belongs to the zone of a particular resonance.

The aim of resonant perturbation theory (see [15, 16, 24], pp. 436–460) is to transform the Hamiltonian (31) into a *resonant normal form* $Z(\theta'_r, \theta', J'_r, J'_\theta)$, where the variables $(\theta'_r, \theta', J'_r, J'_\theta)$ are near-identity transformations of the original variables $(\theta_r, \theta, J_r, J_\theta)$, differing from them by quantities of the order of the amplitude of the non-axisymmetric perturbation. The new Hamiltonian $Z(\theta'_r, \theta', J'_r, J'_\theta)$ contains only terms which are either independent of the angles, or depending on them through trigonometric terms with arguments of the form $k_1 \theta_r + k_2 \theta$ (or multiples). When this is accomplished, it is possible to define a resonant formal integral for the flow of $Z(\theta'_r, \theta', J'_r, J'_\theta)$, which represents an approximate integral of motion besides the energy (Jacobi constant) which is an exact integral. As far as this approximation is valid, the properties of the orbits in the resonance domain can then be determined by analytical means.

The Hamiltonian normalization can be more efficiently carried out in practice by the method of *Lie generating functions*, and by the use of a normalization scheme which greatly simplifies the involved algebraic manipulations, that was presented elsewhere [29].

Of particular interest are those resonances satisfying $k_1 = 1$, as, for example, the inner and outer Lindblad resonances ($k_1 = 1$, $k_2 = -2$ or $k_2 = 2$ respectively, otherwise called the 2:1 and $-2:1$ resonances), as well as the resonances 3:1 ($k_1 = 1$, $k_2 = -3$), 4:1 ($k_1 = 1$, $k_2 = -4$), etc. As we shall see now, the main interest in these resonances stems from the fact that in the neighborhood of one such resonance, the circular orbit of the non-perturbed (axisymmetric) problem is continued as an ‘elongated’ orbit (called ‘ x_1 ’), whose ellipticity is, in fact, ‘forced’ by the non-axisymmetric perturbation by a mechanism formally similar to the mechanism of generation of forced eccentricities in mean motion resonances in the solar system. Namely, the existence of terms proportional to $J_r^{1/2}$ in the resonant part of the Hamiltonian induces a necessarily non-zero amplitude of the epicyclic oscillations. In fact, the terms $O(\epsilon J_r^{1/2})$, where ϵ is a measure of the amplitude of the non-axisymmetric perturbation, play here the same role as the terms $\mu e' O(\Gamma^{1/2})$ in the disturbing function of the ERTBP, where μ and e' are the mass parameter and eccentricity of the disturbing body (Jupiter) and $\Gamma = \sqrt{a}(1 - \sqrt{1 - e^2})$ is the Delaunay action of the asteroid conjugate to its argument of perihelion. As a consequence of this formal analogy, the phase portraits of mean motion resonances in the ERTBP and of epicyclic resonances in disc dynamics are also quite similar (see [63] for a detailed analysis).

3.3 Orbital density wave theory

The existence of elongated periodic orbits is of crucial importance in the theory of both bars and spiral structures. In fact, it turns out that the location of the apsides of the periodic orbits, which is also determined by resonant perturbation theory, changes in a way that the ellipses can support the spiral arms over a considerable extent. This configuration, of ‘precessing ellipses’, schematically drawn by Kalnajs [43], forms the basis of the orbital version of the density wave theory of Lin and Shu [47]. The fact that the apsides of the ellipses follow the right orientation to generate a spiral density wave is a consequence of resonant perturbation theory applied to the ILR [15]. Realistic numerical examples of the ‘precessing ellipses’ mechanism have been given in *self-consistent models* of normal spiral galaxies [20, 54]. These are produced by a variant of Schwarzschild’s method [68], where the periodic orbits responsible for the precessing ellipses flow, as well as thin tube (quasi-periodic) orbits around them, are populated by a Gaussian distribution function. Then, the response density was shown to match the imposed density, i.e. to satisfy the self-consistency criterion.

We now present a simple demonstration of how ‘precessing ellipses’ of x_1 type can be predicted by perturbation theory in the particular case of orbits near the Inner Lindblad resonance in a simple model of a normal spiral galaxy. The galaxy is modeled by a bulge and an exponential disc where a bi-symmetric logarithmic spiral is embedded. The potential is taken as

$$V(r, \theta) = V_0(r) + V_s(r, \theta) = V_b(r) + V_d(r) + V_s(r, \theta) \quad (32)$$

where i) for the bulge V_b we take the Plummer model (16), ii) for the disc we take the potential of an exponential disc (17), and for the spiral arms the model of Eq. (18). Setting the unit of length equal to 1 Kpc and the unit of velocity equal to 1 Km/sec, the unit of $G \times \text{Mass}$ is equal to $4.3 \times 10^{-6} GM_\odot$. In subsequent numerical calculations we adopt the values $M_b = 5 \cdot 10^{10} M_\odot$, $b = 2 \text{ Kpc}$, $\Sigma_0 = 10^9 M_\odot/\text{Kpc}^2$, $\epsilon_d = \epsilon_s = 1/(3 \text{ Kpc})$, and $A_0 = 100 M_\odot/\text{Kpc}^2$. Finally, we assume the spiral arms’ pattern speed to be $\Omega_p = 17 \text{ Kmsec}^{-1} \text{ Kpc}^{-1}$.

Fig. 4a shows the dependence of the azimuthal and epicyclic frequencies, as well as of $\Omega_c - \kappa_c/2$, as a function of the reference radius r_c . For this particular model and value of the pattern speed there are two radii $r_1 = 1.35 \text{ Kpc}$ and $r_2 = 4.75 \text{ Kpc}$ where an ILR occurs. Although, for simplicity, the adopted spiral potential term V_s implies that the spirals extend all the way to the center, we shall in fact be interested in the spiral structure at radii beyond the second ILR, i.e. beyond 4.7 Kpc. In fact, the curve $\Omega_c - \kappa_c/2$ varies by less than 10% in an interval from 4.7 Kpc to about 6.5 Kpc. Thus, in this whole range of radii the epicyclic orbits can be considered as lying close to a 2:1 resonance. Finally, corotation in this model is at $r_{COR} = 11.97 \text{ Kpc}$. The circular velocity under the axisymmetric potential in this distance is $v_c = 203 \text{ Km/sec}$. The rotation curve in inner parts (up to 6 Kpc) is shown in Fig. 4b.

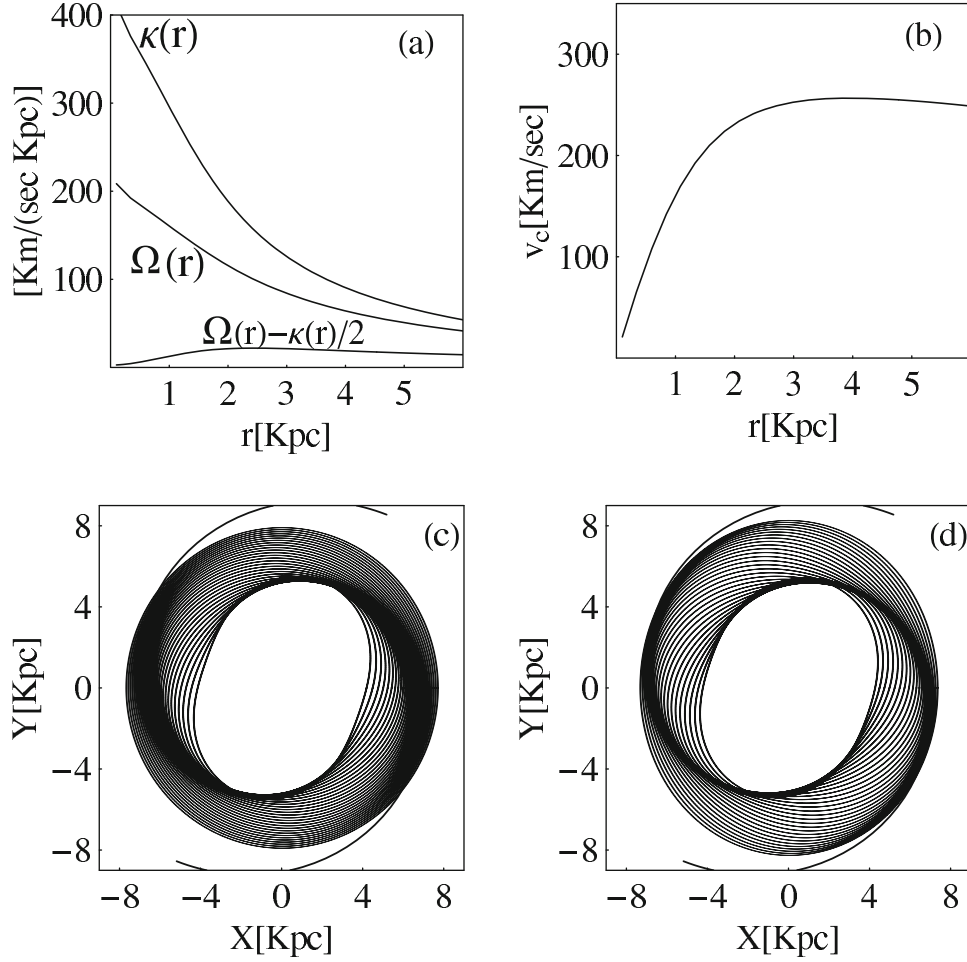


Fig. 4. (a) The azimuthal and epicyclic frequencies, as well as $\Omega - \kappa/2$, as a function of the radial distance r in a numerical model (see text). (b) The rotation curve up to 6 Kpc. (c) Spiral density wave formed by the x_1 family of periodic orbits when the amplitude of the imposed spiral perturbation is $A_0 = 100$. The spiral thin curves show the loci of maxima of the imposed spirals (d) Same as in (c) when the amplitude A_0 is five times bigger ($A_0 = 500$). The response spirals appear now more conspicuous around the locus of maximum of the density wave. This is because the forced ellipticity of the x_1 orbits increases with A_0 .

Our purpose will now be to identify *periodic orbits* in this model which are the *continuations of circular orbits* of the axisymmetric problem in the 2:1 resonant domain. To this end, we split the procedure in the following steps:

i) *Hamiltonian expansion.* We fix first a value of r_c near (and beyond) the second ILR. By Taylor-expanding the total potential $V(r, \theta)$ given by (32) around r_c up to, say, fourth degree in $r - r_c$, and by using the expressions (30), the following Hamiltonian (which is of the general form (31)) is arrived at:

$$\begin{aligned}
 H = & H_c + \omega_r J_r + \omega_\theta J_\theta + c_{20} J_r^2 + c_{11} J_r J_\theta \\
 & + c_{02} J_\theta^2 + c_{21} J_r^2 J_\theta + c_{22} J_r J_\theta^2 + \dots + H_{0F}(J_\theta; J_r, \theta_r) \\
 & + \left(d_{c1} J_r^{1/2} + d_{c3} J_r^{3/2} + \dots \right) \cos(\theta_r - 2\theta) \\
 & + \left(d_{s1} J_r^{1/2} + d_{s3} J_r^{3/2} + \dots \right) \sin(\theta_r - 2\theta) + H_{1F}(J_r; \theta, \theta_r).
 \end{aligned} \tag{33}$$

Understanding the structure of the above Hamiltonian is crucial in subsequent steps, thus let us provide some further insight:

The first two lines in (33) are produced by the ‘unperturbed’ part of the original Hamiltonian $p_r^2/2 + p_\theta^2/2r^2 - \Omega_p p_\theta + V_0(r)$. In fact, they come from the Taylor expansion of this part around r_c (see formulae after Eq. (31)). The function H_{0F} contains all terms from this Taylor expansion which depend on the epicyclic angle θ_r . The subscript ‘F’ stands for ‘fast’. This follows standard terminology of resonant perturbation theory, because the angle θ_r evolves at a frequency much faster than that of the resonant combination $\psi = \theta_r - 2\theta$, which is ‘slow’. A term of order $(r - r_c)^s$ generates terms $J_r^{s/2} e^{ik\theta_r}$ where $|k| \leq s$ and k can only be of the same parity as s . An important remark here is that the linear term $(r - r_c) \propto J_r^{1/2}$ appears in H_{0F} only as a combined product with J_θ or J_θ^2 . This is because the linear term $[-\Omega_c^2 r_c + V_0'(r_c)](r - r_c)$ in the Taylor expansion is exactly equal to zero by the circular orbit condition. This implies that no ‘low order’ $O(J_r^{1/2})$ terms can appear in the first two lines of (33).

By contrast, such terms do appear in the next two lines which are produced by Taylor expanding the spiral potential, which is of the form $V_s = A_s(r) \cos(2\theta - \phi_2(r))$, around r_c . In particular

$$\begin{aligned} d_{c1} &= \frac{1}{(2\kappa_c)^{1/2}} [A'_s(r_c) \sin \phi_2(r_c) + A_s(r_c) \phi'_2(r_c) \cos \phi_2(r_c)] \\ d_{s1} &= \frac{1}{(2\kappa_c)^{1/2}} [A'_s(r_c) \cos \phi_2(r_c) - A_s(r_c) \phi'_2(r_c) \sin \phi_2(r_c)] \end{aligned} \quad (34)$$

and we notice that the coefficients d_{c1} and d_{s1} are non-zero only if the non-axisymmetric perturbation is non-zero. For the model (18), both coefficients are $O(A_0)$, where A_0 is the amplitude of the spiral.

Similarly to H_{0F} , The function H_{1F} contains all terms due to the spiral perturbation which do not contain resonant (‘slow’) trigonometric arguments, but also some higher order resonant terms that are included in the normal form at higher normalization orders.

The presence of the $O(J_r^{1/2})e^{i(2\theta - \theta_r)}$ terms in the hamiltonian (33) produces an important physical effect, namely the fact that the periodic orbits arising as a continuation of the circular orbit in the perturbed case necessarily have a ‘forced’ ellipticity, caused by a forced epicyclic oscillation. This is because, as will be shown, the periodic orbit corresponds to a non-zero value of the epicyclic action J_r . It has been mentioned already that this mechanism of generation of ‘forced’ ellipticity is analogous to the mechanism of generation of forced eccentricities in mean motion resonances in the solar system [63]. In the latter case, the disturbing function contains terms of order $O(\mu(a_*/a')e^2)$ (μ = Jupiter’s mass, e = asteroidal eccentricity, a_* = value of the major semi-axis at the exact mean motion resonance, a' = Jupiter’s major semi-axis), but also terms $O(\mu(a_*/a')^{1/2})ee' \cos(\varpi - \varpi')$ which, in Delaunay actions read $\omega\Gamma + c\Gamma^{1/2} \cos(\varpi)$ (setting $\varpi' = 0$ for simplicity), for some constants $\omega = O(\mu(a_*/a'))$ and $c = O(\mu e'(a_*/a')^{1/2})$. Thus, if we only isolate these terms, we obtain in Poincaré variables a harmonic oscillator whose equilibrium point is not at $x_c = (2\Gamma_c/\omega)^{1/2} \cos \varpi_c = 0$, but at some non-zero value of order $x_c = O(e')$. Physically, the fact that Jupiter has a non-zero eccentricity enforces the asteroids in resonant motions to oscillate also around a non-zero value of their eccentricity.

In the case of the Hamiltonian (33), a measure of the ‘forced ellipticity’ is the value of the epicyclic action J_r for periodic orbits in the resonant domain. This is found by an analogous procedure as above, namely by the subsequent steps:

ii) *Hamiltonian normalization.* By performing a canonical transformation (using e.g. the Lie method) we eliminate in the transformed Hamiltonian all terms included in H_{0F} and H_{1F} which are not in normal form. The new canonical variables are $O(A_0)$ deformations of the old ones, where A_0 is the amplitude of the spiral perturbation. Since A_0 is assumed small, we shall refer to the new canonical variables using the same symbols as for the old ones. In numerical simulations, we often neglect the $O(A_0)$ corrections and replace numerical values of the old variables as if they were the same with those of the new variables. With these simplifications,

the resonant Hamiltonian takes the form (apart from a constant H_c)

$$H_{res} = \omega_r J_r + \omega J_\theta + c_{20} J_r^2 + c_{11} J_r J_\theta + c_{02} J_\theta^2 + c_{21} J_r^2 J_\theta + c_{22} J_r^2 J_\theta^2 + \dots \\ + \left(d_{c1} J_r^{1/2} + d_{c3} J_r^{3/2} + \dots \right) \cos(\theta_r - 2\theta) + \left(d_{s1} J_r^{1/2} + d_{s3} J_r^{3/2} + \dots \right) \sin(\theta_r - 2\theta). \quad (35)$$

Since H_c was omitted, the numerical value of H_{res} for a circular orbit when the coefficients d_c, d_s are zero, i.e., the spiral perturbation is ‘turned off’, is equal to zero. We shall look for periodic orbits in the full Hamiltonian (35) for the same value of the energy (Jacobi constant) i.e. $H_{res} = 0$. Introducing the resonant variables $\psi = \theta_r - 2\theta$, $J_F = J_\theta + 2J_r$ (the index F stands for ‘fast’), the pairs (ψ, J_r) , (θ, J_F) are canonical (it is useful to remember the correct ordering of variables in this canonical transformation, namely $(\theta_r, \theta, J_r, J) \rightarrow (\psi, \theta, J_r, J_F)$). The resonant Hamiltonian in new variables reads:

$$H_{res} = (\omega_r - 2\omega)J_r + \omega J_F + (c_{20} - 2c_{11} + 4c_{02})J_r^2 + (c_{11} - 4c_{02})J_r J_F \\ + c_{02}J_F^2 - 2c_{21}J_r^3 + c_{21}J_r^2 J_F + 4c_{22}J_r^4 + c_{22}J_r^2 J_F^2 - 4c_{22}J_r^3 J_F \\ + \left(d_{c1} J_r^{1/2} + d_{c3} J_r^{3/2} \right) \cos \psi + \left(d_{s1} J_r^{1/2} + d_{s3} J_r^{3/2} \right) \sin \psi. \quad (36)$$

iii) *Position of the periodic orbits.* Setting the numerical value $H_{res} = 0$, and considering a constant value $J_F = \text{const}$, the Hamiltonian H_{res} can be considered as describing the evolution of the one degree of freedom system of the canonical pair (ψ, J_r) . Periodic orbits correspond to the equilibria of this system, since then the motion takes place on the ‘one-torus’ (\equiv periodic orbit) defined by $J_F = \text{const}$. The periodic orbits are then found by the roots for J_F, J_r, ψ of

$$H_{res} = 0, \quad \dot{\psi} = \frac{\partial H_{res}}{\partial J_r} = 0, \quad \dot{J}_r = -\frac{\partial H_{res}}{\partial \psi} = 0. \quad (37)$$

The following simplifications allow to analytically approximate the roots (J_r^*, J_F^*, ψ^*) of the system (37):

1) to estimate J_F^* we use the lowest order terms of the first of equations (37). That is

$$(\omega_r - 2\omega)J_r^* + \omega J_F^* + J_r^{*1/2} (d_{c1} \cos \psi^* + d_{s1} \sin \psi^*) \simeq 0$$

or

$$J_F^* \simeq -\frac{1}{\omega} [(\omega_r - 2\omega)J_r^* + J_r^{*1/2} (d_{c1} \cos \psi^* + d_{s1} \sin \psi^*)]. \quad (38)$$

2) We substitute the above expression into the second and third of equations (37) and solve simultaneously for J_r^*, ψ^* . It can be shown [15] that, depending on the model and examined value of r_c , one or three roots (ψ^*, J_r^*) can be found. If one root can be found, this defines a periodic orbit called x_1 (this nomenclature follows from the Poincaré canonical coordinate $x_1^* = \sqrt{2J_r^*} \cos \psi^*$ corresponding to the unique fixed point). In this case, varying r_c so as to cross the radius of one ILR changes the number of roots from one to three. The two new roots generated at such a transition are called $x_1(2)$ (stable) and $x_1(3)$ (unstable), except in cases where there are two ILRs and r_c is in the interval between them, in which case the new periodic orbits are called x_2 (stable) and x_3 (unstable). In either case, if we neglect the terms of order $O(J_r^{3/2})$ or higher in (36), the roots for ψ^* are defined by:

$$\tan \psi^* = \frac{d_{s1}}{d_{c1}} = \frac{A'_s(r_c) \cos \phi_2(r_c) - A_s(r_c) \phi'_2(r_c) \sin \phi_2(r_c)}{A'_s(r_c) \sin \phi_2(r_c) + A_s(r_c) \phi'_2(r_c) \cos \phi_2(r_c)} \quad (39)$$

where we recall that $\phi_2(r_c) = 2 \ln(r_c/a) / \tan(i_0)$ in our model. Now, we have

$$\left| \frac{A_s(r_c) \phi'_2(r_c)}{A'_s(r_c)} \right| = \left| \frac{1}{\tan i_0 (1 - r_c \epsilon_s)} \right| \gg 1$$

for i_0 equal to a few degrees, thus the second terms in both the numerator and denominator of the r.h.s. of (39) are more important than the first order terms. Thus $\sin[\psi^* + \phi_2(r_c)] \simeq 0$. The error in this relation is, again, $O(A_0)$. The final result is that $\psi^* \simeq -\phi_2(r_c)$ or $\psi^* \simeq -\phi_2(r_c) + \pi$, i.e., as r_c increases, the angle ψ^* ‘precesses’ by closely following the same law as the phase of spiral arms $\phi_2(r_c)$. It should be stressed that this is not a precession in time of one orbit, but a geometric shift of the apsides of *different* orbits belonging to the same family, which arise by implementation of the above theory for different values of the ‘reference radii’ r_c .

To accomplish the task of determining periodic orbits, Eq. (39) has two solutions, which differ by π . Substituting one of them into the third of Eqs. (37), and ignoring terms of order $O(J_r^2)$ or higher, we obtain a cubic equation for $\sqrt{J_r}$. In our case, this equation has three real roots if r_c is beyond the second ILR. For one of the angles ψ^* , the three roots are one positive and two negative, while, for the other, two roots are positive and one negative. In total, there are three positive roots for $\sqrt{J_r}$, yielding three distinct positive values J_r^* for which the resulting orbit is periodic. The key point is that *we must always focus on the value of ψ^* for which the stable orbit generated after the outermost ILR is closest to the circular orbit of the axisymmetric case*. For this orbit the epicyclic action J_r^* is an increasing function of the spiral amplitude A_0 . For J_r^* we have the estimate $J_r^{*1/2} \sim A_0/(\omega_r - 2\omega)$. This scaling law quantifies the ‘forced ellipticity’. In fact, the ellipticity of the periodic orbit is due to the fact that its non-zero value of the epicyclic action J_r^* , which enforces a in-and-out of the orbit from the circle $r = r_c$, is induced by the spiral amplitude A_0 . Furthermore, as r_c increases, the difference $(\omega_r - 2\omega)$ also increases, thus the ellipticity of the periodic orbits decreases as we move away from the ILR (it should be stressed also that this simplified analysis breaks down also very close to the ILR, where the denominator of the previous scaling law tends to zero).

Figures 4c,d show now the combined effect of all previous phenomena, which will be called the ‘precessing ellipses’ mechanism of generation of spiral density waves. This *non-schematic* figure shows a precise calculation of periodic orbits as described in the above steps for the model of Eqs. (32) and for the adopted parameter set. The two panels refer to two different choices of spiral perturbation $A_0 = 100$ and $A_0 = 500$ respectively. In the first case the force due to the spiral in the ILR neighborhood is about 2% of the force due to the axisymmetric background, while in the second case it is of order 10%. The theoretical periodic orbits are described by the equation (in polar coordinates) $r - r_c = (2J_r^*/\kappa_c) \sin(\psi^* + 2\theta)$, where ψ^* and J_r^* are calculated on a grid of values of r_c in the range $r_{ILR} = 4.77 \text{ Kpc} < r_c \leq 7.8 \text{ Kpc}$. We see that the gradual variation of ψ^* , which, as mentioned already, essentially follows the phase of the imposed spirals, causes a gradual re-orientation of the elliptical periodic orbits in a way so as to produce a ‘response density’ which appears enhanced exactly on the locus of the imposed spiral arms. Furthermore, when A_0 becomes larger, the coordination of the orbits’ apsides on the spiral phase produces a more conspicuous response density.

How far can we extend the response spirals using the mechanism of precessing ellipses? In the above construction we have ignored higher order terms in both J_r and J_F and kept only those terms which were convenient for a qualitative description of the mechanism. In fact, however, the nonlinear effects of the spiral perturbation impose that the mechanism becomes inefficient after a certain distance beyond ILR. The main cause of this inefficiency is that, if higher order terms are taken into account in the expansions of the resonant Hamiltonian (35), it turns that the shape of the so-resulting periodic orbits is no longer elliptic, but follows essentially the shape imposed by the kind of epicyclic resonance $n : 1$, where n increases as r_c approaches corotation. The most important nonlinear effect is produced close to the 4:1 resonance [20,21,54]. A numerical calculation shows that the resulting periodic orbits should have a rectangular shape (see Figure 5). This fact prevents the response density from supporting the self-consistency of the spiral arms, and the conclusion from this type of approach is that main bisymmetric open spiral arms in normal galaxies should terminate at the 4:1 resonance (only weak extensions can survive up to corotation, see [54]).

3.4 Vertical resonances

So far, we have considered the orbits in ‘razor-thin’ disks. It is of interest to briefly discuss also what happens if vertical oscillations of the stars are also taken into account. It has already

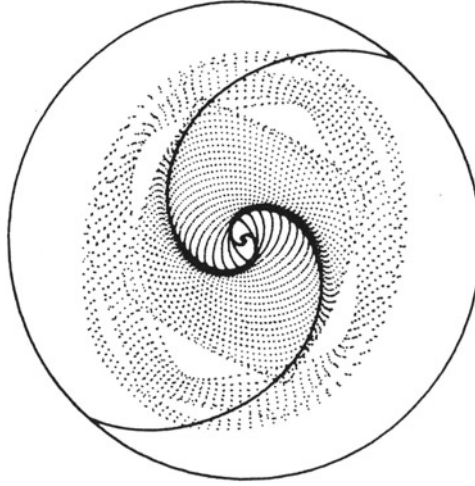


Fig. 5. Termination of spirals at the 4:1 resonance due to nonlinear effects which imply that the periodic orbits do not support the spiral arms beyond this resonance (after [20]).

been discussed in Part I (subsection 11.2.3) that, historically, the study of vertical motions in an axisymmetric galaxy by Contopoulos led to the theory of the ‘third integral’ [14], whose applicability has motivated the works of Hénon – Heiles [40] and Gustavson [38], thus opening the road to modern applications of classical perturbation theory.

The 3D motion of stars in a disc is accounted for by the Hamiltonian:

$$H(r, \theta, z, p_r, p_\theta, p_z) = \frac{p_z^2}{2} + \frac{p_r^2}{2} + \frac{p_\theta^2}{2r^2} - \Omega_p p_\theta + V_0(r, z) + \sum_{m=-\infty}^{\infty} V_m(r, z) e^{im\theta}. \quad (40)$$

We assume the potential to have an even symmetry above and below the disc plane and to be smooth at $z = 0$.

If an orbit lies on the plane $z = 0$ with initial momentum $p_z = 0$, the orbit remains always on the same plane. This is called an equatorial orbit (which, in our context, means an orbit confined on the disc plane).

The starting point for the study of 3D orbits is the set of circular equatorial orbits under the axisymmetric potential $V_0(r, z)$. Let r_c be the radius of a circular orbit. Proceeding as in the 2D case, we Taylor-expand (40) around $r = r_c$, $z = 0$. Defining

$$\kappa_z(r_c) = \left(\frac{\partial^2 V_0(r_c, z=0)}{\partial z^2} \right)^{1/2} \quad (41)$$

as well as an action-angle pair of variables for vertical oscillations via

$$p_z = (2\kappa_z(r_c)J_z)^{1/2} \cos \theta_z, \quad z = \left(\frac{2J_z}{\kappa_z(r_c)} \right)^{1/2} \sin \theta_z, \quad (42)$$

the Hamiltonian (40) resumes the form

$$\begin{aligned} H(\theta_r, \theta, \theta_z, J_r, J_\theta, J_z) = & \kappa_c J_r + (\Omega_c - \Omega_p) J_\theta + \kappa_{zc} J_z + \frac{1}{2r_c^2} J_\theta^2 \\ & + \sum_{|m|+|n|+|q|=0}^{\infty} F_{m,n,q}(J_r, J_\theta, J_z) e^{i(n\theta_r+m\theta+q\theta_z)} \end{aligned} \quad (43)$$

where $\kappa_{zc} \equiv \kappa_z(r_c)$.

Similarly to the 2D case, the coefficients $F_{m,n,q}$ are polynomial up to second degree in J_θ , while they are semi-polynomial, i.e. of half-integer powers $s/2$, $s'/2$ in J_r and J_z respectively. Expressions similar to Eq. (31) hold.

All three basic frequencies κ_c , $\Omega_c - \Omega_p$ and κ_{zc} depend on the reference radius r_c . Thus, as r_c increases outwards, new resonant combinations of the form $k_1\kappa_c + k_2(\Omega_c - \Omega_p) + k_3\kappa_{zc} = 0$ arise, which, if $k_3 \neq 0$, are different from the disc resonances ($k_3 = 0$), and are called *vertical resonances*.

Far from vertical resonances, the quantity J_z is an approximate integral of motion. If r_c is chosen so that the system is far from disc resonances as well, non-resonant perturbation theory can be applied to the Hamiltonian (43) so as to produce a normal form independent on all three angles $\theta, \theta_r, \theta_z$. Then the (canonically transformed) new action variables J'_θ, J'_r are two approximate integrals of motion, which correspond to nearly constant angular momentum and epicyclic action respectively, while the third transformed action J_z is also an approximate ‘third integral’. In applications we often consider the motion as decoupled, i.e. as the product of the 2D motion on the disc, which was analyzed in previous subsections, times a vertical oscillation with frequency κ_{zc} .

If r_c is close to a disc resonance, it is still possible to define three approximate integrals, which are the energy (Jacobi constant = exact integral), the fast action J_F (see above), and J_z .

If, however, r_c is such that the system is in vertical resonance, then the number of approximate integrals depends on the *multiplicity* M of the resonance, i.e., the maximum number of linearly independent resonant conditions of the form $k_1\kappa_c + k_2(\Omega_c - \Omega_p) + k_3\kappa_{zc} = 0$ which hold. The number of integrals (including the energy) is equal to $N = 4 - M$, where M can take the values 1 or 2. Thus, if the resonance is ‘simple’ $M = 1$, there are three formal integrals, while if $M = 2$ there are two such integrals. The last case is of particular interest, since the normalized Hamiltonian is non-integrable, thus it can generate a large degree of chaos.

Three dimensional *periodic orbits* at vertical resonances, which bifurcate from the main disc families of periodic orbits (e.g. x_1) have been discussed extensively in recent years, the most useful classification being given in the works [56,75]. It has been found that an appropriate combination of such orbits can reproduce most features seen in *edge on* profiles of disc galaxies, such as, for example, *peanut-shaped edge-on profiles*. An interesting remark is that the 3D periodic orbits have four rather than two distinct types of stability, namely they can be i) stable (four complex eigenvalues of the monodromy matrix on the unit circle), ii) simply unstable (two complex eigenvalues on the unit circle and two reciprocal on the real axis), iii) doubly unstable (two pairs of real and reciprocal eigenvalues), or iv) *complex unstable* (all four eigenvalues complex, and reciprocal by pairs, but not on the unit circle). The role of complex instability in the vertical structure of galactic discs is largely unknown. It has been suggested that the bifurcation of complex unstable 3D periodic orbits, as we move along the disc, may account for substructures seen in edge-on profiles, namely the appearance of gaps succeeded by sharp maxima of the edge-on projected surface density profile [56]. Observational evidence on whether such a pattern indeed exists is, however, still inconclusive.

3.5 Orbits in strongly perturbed potentials. The role of chaos

In normal spiral galaxies, the non-axisymmetric perturbation is rather small (<10%). This implies that chaos is a very local phenomenon, i.e. localized in the close neighborhood of unstable periodic orbits at disc resonances. As a result, chaos plays no important role in the disc dynamics. Instead, the main nonlinear effects are seen in the *form* of the periodic orbits, which deviate more and more from circular as the amplitude of the non-axisymmetric perturbation increases. In fact, it was mentioned already that this phenomenon is responsible for the termination of the main spiral arms at the 4/1 resonance (Fig. 5).

In barred galaxies, on the other hand, the amplitude of the non-axisymmetric perturbation is significantly larger than in normal spiral galaxies. Figure 6 shows the so-called Q -values for the bars and spiral arms of a recent sample of galaxies where photometric observations in the near infrared yield a reliable estimation of the mass distribution, and hence the gravitational

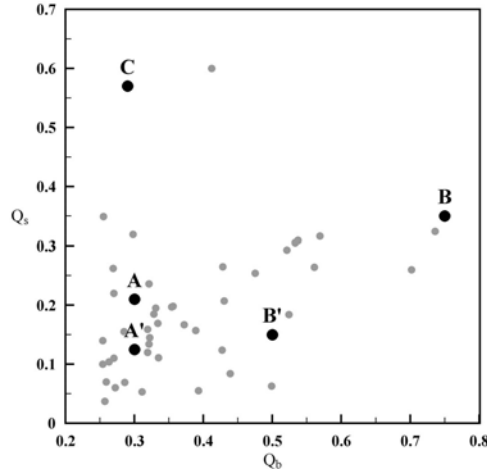


Fig. 6. Q_b versus Q_s values (ratio of non-axisymmetric over axisymmetric forcing for the bar and spiral arms respectively) for the data given in [10]. The thick dots represent the position in this diagram of five models [81], where the theory of invariant manifolds was implemented (section 4).

potential of each galaxy [10, 46]. Following an analysis of the potential in Fourier components, the Q -value is found as the maximum value of the ratio of the tangential force versus the radial force for the bar and spiral components (Q_b and Q_s respectively). The reported average values for SB galaxies in the sample of Buta et al. [10] are $\langle Q_b \rangle = 0.29$ and $\langle Q_s \rangle = 0.17$. In fact, as seen in Fig. 6, for particular galaxies the Q -values of both components can be substantially larger than the average values. In any case, these amplitudes are much stronger than in the case of normal galaxies. This implies that chaos becomes now of crucial importance. The bold points in the same figure denote the Q -values for an ensemble of models used in a recent study [81], where the role of the *invariant manifolds* of unstable periodic orbits near the unstable Lagrangian points L_1 and L_2 was investigated. This topic will be discussed in the next subsection. In our models, we use a combination of a Plummer sphere for the dark halo, an exponential disc, and a Ferrer bar. The spiral arms start at the end of bars and may extend up to the outer Lindblad resonance. A rough model for bi-symmetric spiral arms is

$$V_s(r, \theta) = A(r)r \exp(-\epsilon_s r) \cos 2\Phi \quad (44)$$

where

$$\Phi = \frac{1}{2} [1 + \tanh(2(r - a))] \frac{\ln(r/a)}{\tan i} - \theta$$

and

$$A(r) = \left(\frac{A - A_r}{4} \right) (1 + \tanh[\kappa_1(r - r_1)]) (1 + \tanh[\kappa_2(r_2 - r)]) + A_r.$$

In the above formulae, Φ is taken so that the logarithmic dependence of the phase on r has a smooth cut-off for radii smaller than the bar's major semi-axis a . Furthermore, the spiral amplitude $A(r)$ is considered to be the sum of a constant term A_r and of a term varying with the distance from the center. This latter term becomes important between an inner and an outer radius (r_1, r_2 respectively). This allows for a flexible modeling of the spiral arms according to where the observations indicate that the spiral arms terminate.

In understanding the orbital structure of barred-spiral galaxies, it is useful to recall the form of the zero velocity curves, which are the contours of the effective potential

$$V_e(r, \theta) = V(r, \theta) - \frac{1}{2} \Omega_p^2 r^2 = E_J \quad (45)$$

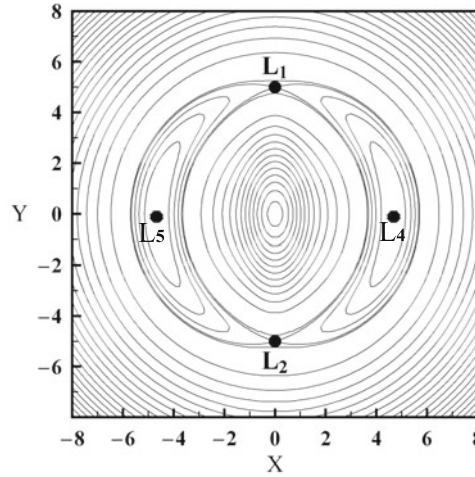


Fig. 7. Zero Velocity Curves of the effective potential in a barred galaxy.

where E_J is the numerical value of the Jacobi constant at points where the velocity of an orbit is equal to zero. The structure of this diagram (Fig. 7) has some similarities but also some differences with the corresponding diagram in the RTBP. In particular, there are three stable and two unstable Lagrangian equilibrium points. In galactic dynamics, we call L_1 , L_2 the *unstable* points, L_3 the stable point at the center and L_4 , L_5 the stable points (same for this pair as in the RTBP, but lying at 90° degrees from the bar's major axis, instead of 60° from Jupiter). The corotation circle is defined by $\Omega(r_{COR}) = \Omega_p$. If there are no spiral arms, the $m = 2$ Fourier component of the mass distribution inside the bar implies that L_1 , L_2 are aligned with the bar and lie outside r_{COR} , while L_4 , L_5 are in the axis perpendicular to the bar and lie inside r_{COR} . This can be easily seen if we examine the simple case of a potential field $V = V_0(r) + A_2(r) \cos(2\theta - \phi_{2bar})$, where we take $A_2(r) > 0$ (this is equivalent to taking $A_2(r) < 0$ and shifting ϕ_{2bar} by π). The bar density is maximum at the minima of the potential, namely $\theta = (\phi_{2bar} \pm \pi)/2$. These are the angles of (opposite) orientations of the two bar's major semi-axes. The amplitude $A_2(r)$ must decrease outwards. If r_* is any fixed distance, the force at $r = r_*$ due to the bar varies with θ . Thus, along the major axis the force is $|F(r_*, \theta = (\phi_{2bar} \pm \pi)/2)| = \frac{dV_0}{dr_*} + \left| \frac{dA_2}{dr_*} \right|$, while along the minor axis it is $|F(r_*, \theta = (\phi_{2bar} \pm \pi)/2)| = \frac{dV_0}{dr_*} - \left| \frac{dA_2}{dr_*} \right|$. Thus the force at the same distance is larger along the major than along the minor axis. At equilibrium points this has to be balanced by the centrifugal force $\Omega^2 r_*$, which increases with distance. Thus, the balancing takes place at larger r_* along the major axis (L_1 or L_2), than along the minor axis (L_4 or L_5).

ii) The main families of periodic orbits are reviewed in [22]. We distinguish:

a) The x_1 family [18] (Fig. 8) which is the main stable continuation of the circular orbit beyond the ILR. It has already been discussed that close to ILR this family yields elliptic orbits. However, at larger distances, i.e., closer to higher order resonances, the orbit tends to acquire a polygonal form so as to mimic the kind of epicyclic oscillation that corresponds to the particular resonance. For example, close to the 4:1 resonance the x_1 orbit tends to a rectangular shape (Fig. 9a), which means that the orbit makes four epicyclic oscillations ('in and out') in one azimuthal period. The orbit may also develop loops at the 'corners'. Similarly, close to the 6:1 resonance the orbit tends to a hexagonal shape (Fig. 9b). The x_1 family continues also beyond corotation, where it becomes retrograde.

b) the x_2 , x_3 and x_4 families. The families x_2, x_3 lie between the two ILRs (if these exist) and they are elongated perpendicularly to the bar. The family x_2 is stable while the family x_3 is unstable. The x_4 family is retrograde, and plays a role only close to the center.

c) Resonant periodic orbits 2:1, 3:1, 4:1 etc., which are the survived periodic orbits of the corresponding orbits in the axisymmetric background arising by application of the Poincaré-Birkhoff theorem.

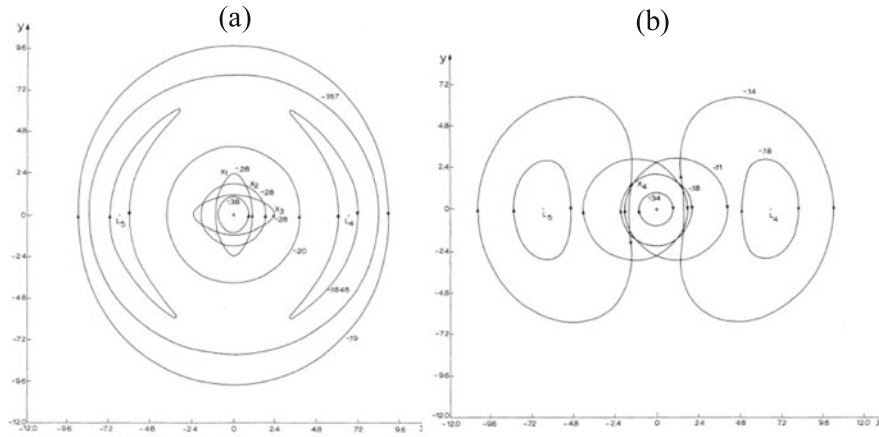


Fig. 8. (a) Main families of periodic orbits in a barred galaxy (after [18,22]). We distinguish the x_1 , x_2 , x_3 , families as well as the long period ('banana') orbits around L_4 and L_5 . (b) Short period orbits around L_4 and L_5 as well as x_4 orbits.

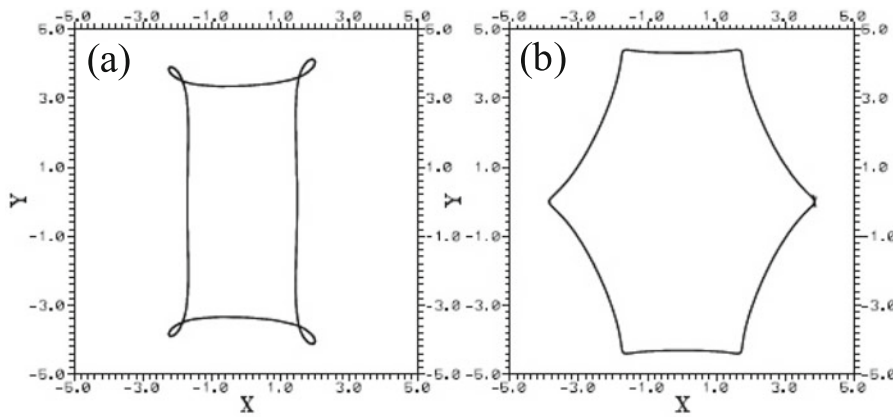


Fig. 9. (a) A x_1 orbit resembling a parallelogram with loops at the corners (after [55]). (b) Hexagonal-like orbit from the same family.

d) Short and long period orbits around L_4 and L_5 (Fig. 8). These are quite similar as the short and long period orbits at the 1:1 resonance of the ERTBP. In particular, the short period orbits form small loops around the Lagrangian points.

e) Short period orbits around L_1 and L_2 . As we shall see in the next subsection, these orbits appear to play a key role in the generation of spiral structure beyond corotation, via their unstable *invariant manifolds*.

The most important question regards what sort of structures are supported by the various types of orbits. The stable periodic orbits are surrounded by quasi-periodic orbits forming 'tubes' that correspond to motions on invariant tori in phase space. Inside corotation, the quasi-periodic orbits follow in configuration space a similar pattern as described by the periodic orbits. Thus, in general they support the bar. On the other hand, outside corotation the regular orbits may support parts of the spiral arms [44,58].

The chaotic orbits are important mostly in the interval of values of the Jacobi constant $E_{J,L4} < E_J < E_{J,L1}$. The structure of the zero velocity curves (Fig. 7) implies that in this interval an orbit is energetically allowed to wander partly inside and partly outside the corotation region. In this way, it was shown [44] that the chaotic orbits can partly support the outer envelope of the bar and partly the spiral arms. This fact is important, because it implies that chaos is not always opposed to, but in many cases may actually enhance self-consistency. Chaotic orbits of this type were likewise found in N-body simulations of barred galaxies [76],

and they have been called ‘hot population’, because of the large velocity dispersion produced by stars in such orbits.

A combination of the above potential terms and of two more terms for the halo (Plummer sphere) and for the disc (exponential) yields the total potential that was used in the production of *self-consistent* models of three real galaxies [44]. The method used was a variant of Schwarzschild’s method [68] (see Part I; subsection 11.4.6). Successful models were found when the pattern speed Ω_p was fixed at such a value that the bar in all three models is a little inside corotation. The fact that bars terminate near corotation [17] is mainly due to the large degree of chaos near corotation, due to the destabilization of an ensemble of periodic orbits which accumulate there, generating a large chaotic domain which has been found in virtually all numerical studies [1, 19, 35, 44, 45, 55, 59–61, 76]. Stochasticity of the orbits also results in a considerable deformation of the shape of outer isophotes of the bar from pure ellipses [2]. An additional factor is the change of orientation of the main periodic orbits outside corotation, which become elongated perpendicularly to the bar, thus no longer supporting self-consistency.

4 Invariant manifolds and spiral structure

We will now focus on a theoretical model for the chaotic nature of spiral arms in barred galaxies, which has been recently proposed and numerically explored in two versions ([82, 83] and [65, 66]). According to this model, the spiral arms (or even ring structures) are supported by the *unstable invariant manifolds* of the two short period families of unstable periodic orbits around the unstable Lagrangian equilibria L_1 and L_2 . These are called hereafter the PL1 and PL2 families respectively. This theory was extended in [80], by examining the contribution of the unstable manifolds of other families, besides PL1 or PL2, to the same phenomenon. It was found that all the manifolds form together a ‘coalescence’ which produces a strong amplification of the spiral arms beyond corotation. The importance of the chaotic orbits in supporting the spiral structure of barred galaxies has also been emphasized in [57], where it was found that chaotic orbits emanating mainly from the neighborhood of the 4:1 resonance spend a long time supporting the spiral arms of a realistic model fitting the galaxy NGC4314 (whose analytic form of the potential, following an analysis similar to that presented in section 2, has been derived on the basis of IR photometric observations [62]). Further examples of the applicability of the invariant manifold theory in models of real galaxies were given in [81]. Finally, in [4] the question was explored of how can various morphological types, from closed rings to open spirals, be supported by the invariant manifolds under different parameters of the non-axisymmetric amplitude and/or the pattern speed.

We proceed now in a brief account of the theory of the invariant manifolds. Returning to the 2D Hamiltonian

$$H(r, \theta, p_r, p_\theta) \equiv \frac{1}{2} \left(p_r^2 + \frac{p_\theta^2}{r^2} \right) - \Omega_p p_\theta + V_0(r) + V_1(r, \theta) = E_J \quad (46)$$

the unstable manifold $\mathcal{W}_{L_1}^U$ of the equilibrium point L_1 is defined as the set of all initial conditions $(r_0, \theta_0, p_{r0}, p_{\theta0})$ in phase space for which the resulting orbit tends asymptotically to L_1 in the backward sense of time, namely

$$\mathcal{W}_{L_1}^U = \left\{ \bigcup (r_0, \theta_0, p_{r0}, p_{\theta0}) : \lim_{t \rightarrow -\infty} \|Q(t; r_0, \theta_0, p_{r0}, p_{\theta0}) - (r_{L_1}, \theta_{L_1}, 0, \Omega_p)\| = 0 \right\} \quad (47)$$

where $Q(t; r_0, \theta_0, p_{r0}, p_{\theta0})$ denotes the position (point in phase space) at time t of a particle along an orbit starting with the above initial conditions, and the norm $\|\cdot\|$ means the Euclidean distance between this point and the phase space point $(r_{L_1}, \theta_{L_1}, 0, \Omega_p)$, corresponding to L_1 . Since the equilibrium point L_1 is simply unstable, $\mathcal{W}_{L_1}^U$ is a two-dimensional manifold embedded in the three-dimensional hypersurface of the phase space corresponding to a fixed Jacobi constant $E_J = E_{J, L_1}$. Similar definitions and properties hold for L_2 and $\mathcal{W}_{L_2}^U$ and for the stable manifolds $\mathcal{W}_{L_1}^S$, $\mathcal{W}_{L_2}^S$, i.e. the sets of initial conditions tending asymptotically to L_1 , or L_2 in the forward sense of time, as $t \rightarrow \infty$.

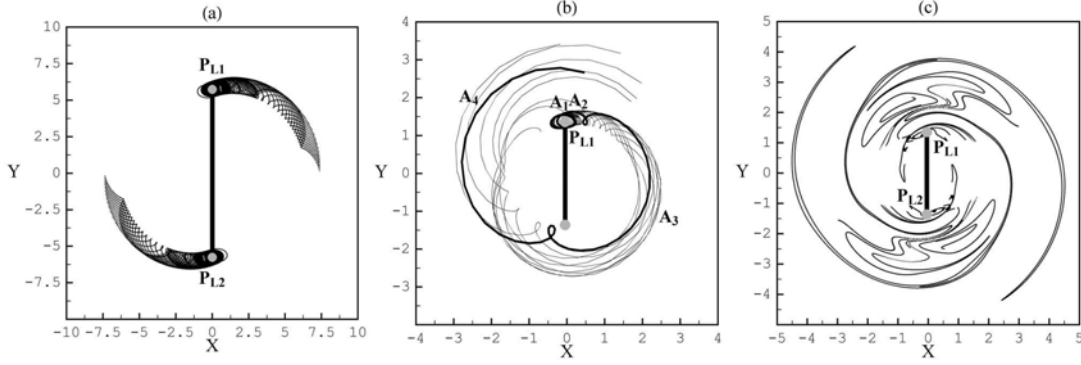


Fig. 10. (a) Projections of the invariant manifolds \mathcal{W}_{PL1}^U and \mathcal{W}_{PL2}^U on the configuration space, approximated by an ensemble of 20 orbits with initial conditions on the unstable direction of the tangent space to the periodic orbits PL1 and PL2 (small bold circles) in the neighborhood of L_1 and L_2 . The potential used is model A of [80]. The bar is aligned to the y-axis and rotates counterclockwise. (b) Same as in (a) but in a strongly non-linear model (in this case we use the potential of the N -Body simulation analyzed in [82,83]. Only the manifold \mathcal{W}_{PL1}^U is plotted). (c) The projection of the intersection of the manifolds \mathcal{W}_{PL1}^U and \mathcal{W}_{PL2}^U with the apocentric surface of section $p_r = 0$, $\dot{p}_r < 0$ on the configuration space of the model of (b) (after [81]).

It has already been mentioned that, for $E_J > E_{J,L1}$, a short-period unstable periodic orbit (PL1) bifurcates from L_1 (and the symmetric orbit PL2 from L_2). This orbit forms a small loop around L_1 (Fig. 10a). The unstable manifold of PL1 is defined as

$$\mathcal{W}_{PL1}^U = \left\{ \bigcup (r_0, \theta_0, p_{r0}, p_{\theta0}) : \lim_{t \rightarrow -\infty} \|Q(t; r_0, \theta_0, p_{r0}, p_{\theta0}) - PL1\| = 0 \right\} \quad (48)$$

where the notation $\|\cdot\|$ refers now to the minimum of the distances of $Q(t)$ from the locus of all the phase space points of the orbit PL1. For any fixed value of $E_J > E_{J,L1}$, \mathcal{W}_{PL1}^U is also a two-dimensional manifold embedded in the three-dimensional hypersurface of constant E_J .

The manifolds \mathcal{W}_{L1}^U or \mathcal{W}_{PL1}^U can be numerically computed as follows:

1) We define an *apocentric* surface of section \mathcal{S} , given by the points (θ, p_θ) where an orbit satisfies $p_r = \dot{r} = 0$, $\ddot{r} < 0$. A point (θ, p_θ) on \mathcal{S} defines a unique initial condition of an orbit, whose remaining initial conditions are $p_r = 0$ and r is one root of $H(r, \theta, p_r, p_\theta) = E_J$. In fact, the latter equation has two roots and the apocentric condition implies to choose the largest of these roots.

2) By numerical integration we can compute the images under Poincaré mapping of all initial conditions (θ, p_θ) on \mathcal{S} . We thus determine the unstable fixed points of PL1 or PL2 on \mathcal{S} . Computing also the monodromy matrix of the tangent map, we calculate its eigenvalues and eigenvectors. The latter determine two opposite directions on a straight line for the unstable and two for the stable eigenvalue.

3) Taking now a small segment dS on one unstable direction, and calculating many orbits with initial conditions on dS , we form a representation of the unstable manifold \mathcal{W}_{PL1}^U . This exploits the fact that the manifold \mathcal{W}_{PL1}^U is tangent to the unstable manifold of the linearized Hamiltonian flow near PL1 (the so-called Grobman – Hartman theorem [36,39]).

The unstable manifold of the equilibrium points themselves, e.g. \mathcal{W}_{L1}^U , can be calculated by a similar manner, since these points are also fixed points of the Poincaré mapping on \mathcal{S} for $E_J = E_{J,L1}$. However, the latter manifolds can be computed also analytically up to a meaningful extent, using a *hyperbolic normal form* method introduced by Moser (e.g. [74]). Following this method, we first expand all canonical variables in the Hamiltonian (46) around their values at L_1 (or L_2). Since L_1 is simply unstable, the linearized form of the equations of motion imply that one degree of freedom is elliptic (this yields the small loops around L_1 , which are described with some frequency ω), while the second degree of freedom is hyperbolic, i.e. it expands or contracts as $\sim e^{\pm \lambda t}$ for some characteristic exponents $\pm \lambda$ which are eigenvalues of the matrix of

the linearized equations of motion around L_1 . After a diagonalizing canonical transformation

$$(r - r_{L1}, \theta - \theta_{L1}, p_r, p_\theta - \Omega_p r_{L1}^2) \rightarrow (q, p, \xi, \eta)$$

the Hamiltonian locally resumes the form

$$H = i\omega qp + \lambda \xi \eta + \sum_{s=3}^{\infty} P_s(q, \xi, p, \eta) \quad (49)$$

where P_s are polynomials of degree s in the canonical variables (q, ξ, p, η) . This hamiltonian can now be brought into normal form by the usual normalization procedure, using e.g. composition of series of Lie generating functions. At any order s , the homological equation to solve in order to determine the generating function χ_s is of the form

$$\{\chi_s, i\omega qp + \lambda \xi \eta\} = \tilde{H}^{(s)} \quad (50)$$

where $\tilde{H}^{(s)}$ contains all terms of degree s of the transformed Hamiltonian function (after $s - 2$ normalization steps) not belonging to the kernel of the Poisson bracket operator $\{\chi_s, \cdot\}$. In fact, since $i\omega$ is imaginary while λ is real, it follows that no other kernel terms exist than those, at even orders, of the form $q^{s_1} p^{s_1} \xi^{s_2} \eta^{s_2}$ with $s_1 + s_2 = s/2$. This simple remark forms the basis of the proof that the hyperbolic normal form of Moser is actually *convergent* (in contrast to the doubly elliptic case of the Birkhoff normal form). Namely, the divisor structure at any order s is of the form $k_1 \omega + i k_2 \lambda$, where $|k_1| + |k_2| \leq s$. A careful inspection of how divisors accumulate then shows that the worst accumulation at order s produces terms of size $\sim sA/(s \min(\omega, \lambda)) = A/\min(\omega, \lambda)$ times the size of terms at order $s - 1$, where A is a measure of the size of non-linear terms (cubic and beyond) in (49). It follows that the growth of the size of series terms is essentially geometric, with ratio $1/\min(\omega, \lambda)$. This also sets the scale of the domain of convergence, which in turn determines up to what length the invariant manifolds can be calculated using this method. The latter are given by the straight lines $\xi' = 0$ and $\eta' = 0$, where primes denote transformed variables, and the above lines pass through L_1 . Back transforming to the original variables yields the form of the invariant manifolds in the usual configuration space as well.

Fig. 10a shows the initial part of the unstable manifold \mathcal{W}_{PL1}^U for a particular model of barred galaxy ([81], model A). In the above reference, the dynamical role of the invariant manifolds was investigated in the case of models referring to three real galaxies, that were also the subject of an earlier investigation on the existence of self-consistent models [44]. In Fig. 10a we can distinguish the epicyclic loops of size nearly equal to the PL1 loop, formed by orbits belonging to \mathcal{W}_{PL1}^U . In the same time, the guiding center recedes from PL1 along a path which yields a *trailing spiral arm*.

In fact, the deviation of the guiding center from PL1 is exponential in time $\sim e^{\lambda t}$. Since the value of λ depends on the strength of the non-axisymmetric perturbation, it follows that in strongly nonlinear models the size of the epicycles further away from PL1 becomes great, and it may exceed even the size of the bar. Such an example is shown in Fig. 10b, referring to the orbits of the \mathcal{W}_{PL1}^U family in a N -Body model of a barred galaxy [82]. We see that one asymptotic orbit (bold) forms two relatively small loops near PL1, reaching the apocentric positions A_1 and A_2 , but the exponential recession of the guiding center from L_1 is so fast that there is no loop formed between the second and third (A_3) apocentric positions. Furthermore, the fourth apocentric position is at a distance about twice the bar's major semi-axis. Further integration beyond that of Fig. 10b shows that, in fact, all these orbits belong to the so-called 'hot population' [76] that has already been mentioned. In this case, the manifolds viewed as the loci of a continuous flow of orbits do not produce a spiral structure. However, if recurrent motions are taken into account, by considering the locus of all apocentric positions of the orbits on \mathcal{W}_{PL1}^U , we find again that the manifolds support a well-developed spiral structure. In the N -body model, this structure is found to coincide with the maxima of the $m = 2$ component of the particles' density. Also, in the models of [81], we find that the manifolds nicely reproduce the imposed spiral arms, which represent a model for the observed spiral arms in the examined

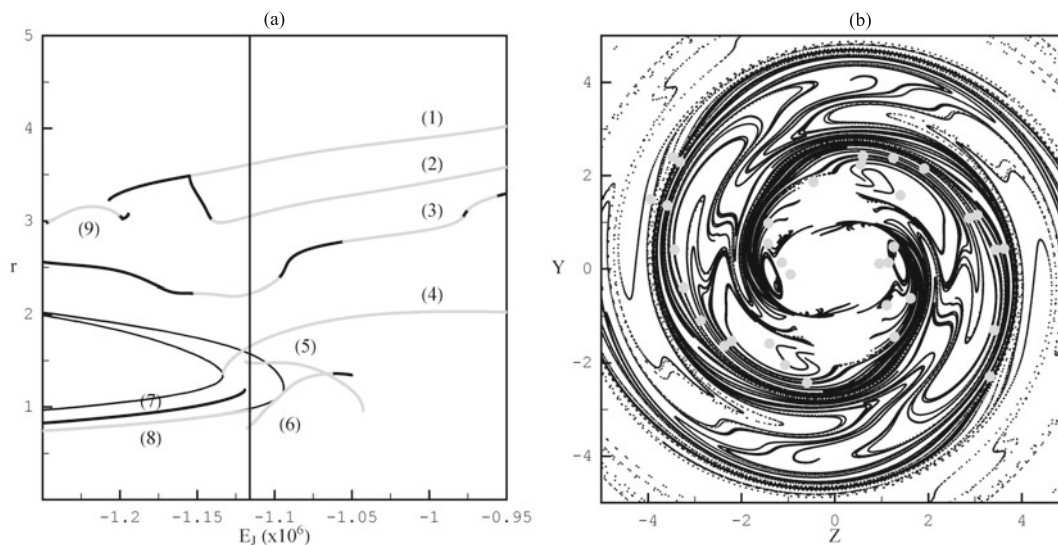


Fig. 11. (a) Characteristic curves of different periodic orbits near and beyond corotation in an N-body model (after [80]). A gray line indicates that the orbit is unstable. The families are (1) $-1 : 1$, (2) $-1 : 1$ of multiplicity two (bifurcating from the $-1 : 1$ family), (3) $-2 : 1$, (4) PL1, (5) $-4 : 1$, (6) $4 : 1$, (7) and (8) $3 : 1$, and (9) $-1 : 1$. The vertical line marks a value of the Jacobi constant where seven out of the nine examined families are unstable. (b) The ‘coalescence of invariant manifolds’, formed by the superposition of the unstable manifolds of all unstable orbits on the vertical line of (a).

galaxies. In fact, we find that the most crucial factor in order that the manifolds produce a spiral structure is *not* the amplitude of the non-axisymmetric perturbation, but its *azimuthal tilt*. This can be concluded by comparing the phase portraits arising in an original model with azimuthal tilt, and one where we artificially align the $\Phi_2(r)$ phase along the bar for all radii r . Namely, while both cases are characterized by the same perturbation strength, in the case when $\Phi_2(r)$ is constant the invariant manifolds can only yield a ring rather than spiral structure beyond corotation.

In summary, the invariant manifold theories of [82] and [66] start with the same principle, namely the possibility to build spiral arms with chaotic orbits exploiting the unstable invariant manifolds emanating from the short period orbits around L_1 and L_2 . However, the two approaches have also one important difference. Namely, in the approach of [66], the spirals are generated due to a continuous flow along the manifolds, and the associated orbits have essentially no recurrence to the neighborhood of the unstable Lagrangian points. The so-resulting spiral arms can be quite open and extend far from corotation, but the stars on them do not return inside the galaxy and a mechanism of replenishment of this material must be called upon to justify whether these spiral arms can be long-lived. On the other hand, if we only consider the apsidal (apocentric or pericentric) sections of the manifolds [82] we find that these support spirals which are a morphological continuation of rings. In the same time, these manifolds have many recurrences in and out of corotation, thus they can produce long-lived spiral arms.

Finally, when chaos is pronounced, we expect that there are heteroclinic connections between the invariant manifolds of many *different families* of periodic orbits which are all embedded in a connected chaotic domain. This implies the possibility of chaotic transfer of material from inner regions (inside the bar) to outer regions (outside the bar) and vice versa. Such connections are possible between the unstable manifolds of one family and the stable manifolds of another. However, the unstable manifolds of all families do not intersect one another, thus they all have to move in parallel and produce quite similar structures. An example is given in Fig. 11. The left diagram shows the characteristic curves (value of the apocentric radius r_a of a periodic orbit as a function of the Jacobi constant E_J) of nine different families of periodic orbits in the model of [80]. The left vertical line indicates a value of the Jacobi constant where 7 families are unstable. If the unstable manifolds of all these families

are superposed (Fig. 11b), they generate a ‘coalescence’ of invariant manifolds, which enhances even more the spiral structure. This phenomenon leads to the conclusion that no family should be assigned a preferential role in the creation of the spiral pattern, since the topological properties of the invariant manifolds of all families are quite similar in outer radii (beyond corotation) where the manifolds form the ‘coalescence’, by entering one into the other.

We conclude this section by briefly discussing a different methodology [57], which has been developed primarily in order to determine plausible values of the pattern speed Ω_p in rotating galaxies of any type, on the basis of a simple and robust dynamical criterion. According to this method, of *response models*, after selecting a value for Ω_p , we take the initial conditions of a number (say 10^5) particles all lying on circular orbits of the axisymmetric part of the potential and on part of a uniformly populated disk, with distances in a range corresponding to the whole range of Jacobi constant values up to $E_{J,L4}$ of the full potential. By softly introducing the non-axisymmetric part of the potential (in a small number of pattern periods), we integrate the particles’ orbits and allow for them to settle to a mass distribution on the disk. If Ω_p is properly chosen, we observe that this distribution remains nearly constant for many periods. If the response distribution matches the mass distribution corresponding to the used potential, we say to have a successful response model of the galaxy.

The theory of invariant manifolds is relevant to some features of the theory of response models. This is because any open set of initial conditions in a connected chaotic domain tends to follow, after a small number of periods, preferential directions in phase space which are the ones imposed by the unstable invariant manifolds. This phenomenon is exemplified in Fig. 12 [81]. In panels b–e, we see the response of particles with initial conditions (θ, p_θ) on a locus of \mathcal{S} forming a zone around a straight line $p_\theta = \text{const}$, which corresponds to the angular momentum for a circular orbit under the axisymmetric potential only for a fixed value of the Jacobi constant (Fig. 12a). Running these initial conditions for only one period (Fig. 12b) or twelve periods (Fig. 12c, d), we see that, already after one iteration, the initially straight zone of initial conditions in the surface of section is deformed so as to closely follow a pattern induced essentially by the invariant manifolds $\mathcal{W}_{PL1,2}^U$. This phenomenon is repeated at subsequent iterations, and it persists after a substantial number of iterations. Figure 12f, now, shows a full response model at a time corresponding to the 25th pattern period. Clearly, the particles have settled to orbits supporting both the bar and the imposed spiral arms. But the most interesting feature of this distribution is that the maxima of the density of the response spirals depart from the maxima of the imposed spiral at nearly the same points (points A and A') where this happens for the invariant manifolds.

We thus see that the invariant manifolds play for chaotic orbits a role similar to that of stable resonances for regular orbits, i.e. the manifolds are able to capture the chaotic orbits in their neighborhood and to create response spiral arms. This has been explained in [80, 81] as a manifestation of the phenomenon of *stickiness in chaos* [25, 31]. Namely, even if the invariant manifolds are embedded in a strongly chaotic domain, the manifolds create partial barriers to diffusion inside their lobes, whose boundaries (for the manifolds of one unstable orbit) cannot be intersected by the unstable manifolds of other orbits. In this way, we may maintain patterns induced by the invariant manifolds over a considerable percentage of the Hubble time, even if chaotic orbits far from the manifolds tend to escape from the system in only a few dynamical times.

5 Additional remarks

We close this brief introduction by a quick reference to a number of topics that have raised great current interest concerning disc dynamics (this is not an exclusive list; for a recent review of additional topics see [72]):

i) *Observational determination of the pattern speed*. It has been repeatedly stressed above that the pattern speed is a quantity of central interest in disc dynamics, which, however, cannot be determined from observations without some dynamical assumptions on the observed galaxy. There are, nevertheless, proposals on how to minimize the input from dynamics in this process. The most promising appears to be the *Tremaine-Weinberg* method [78, 79].

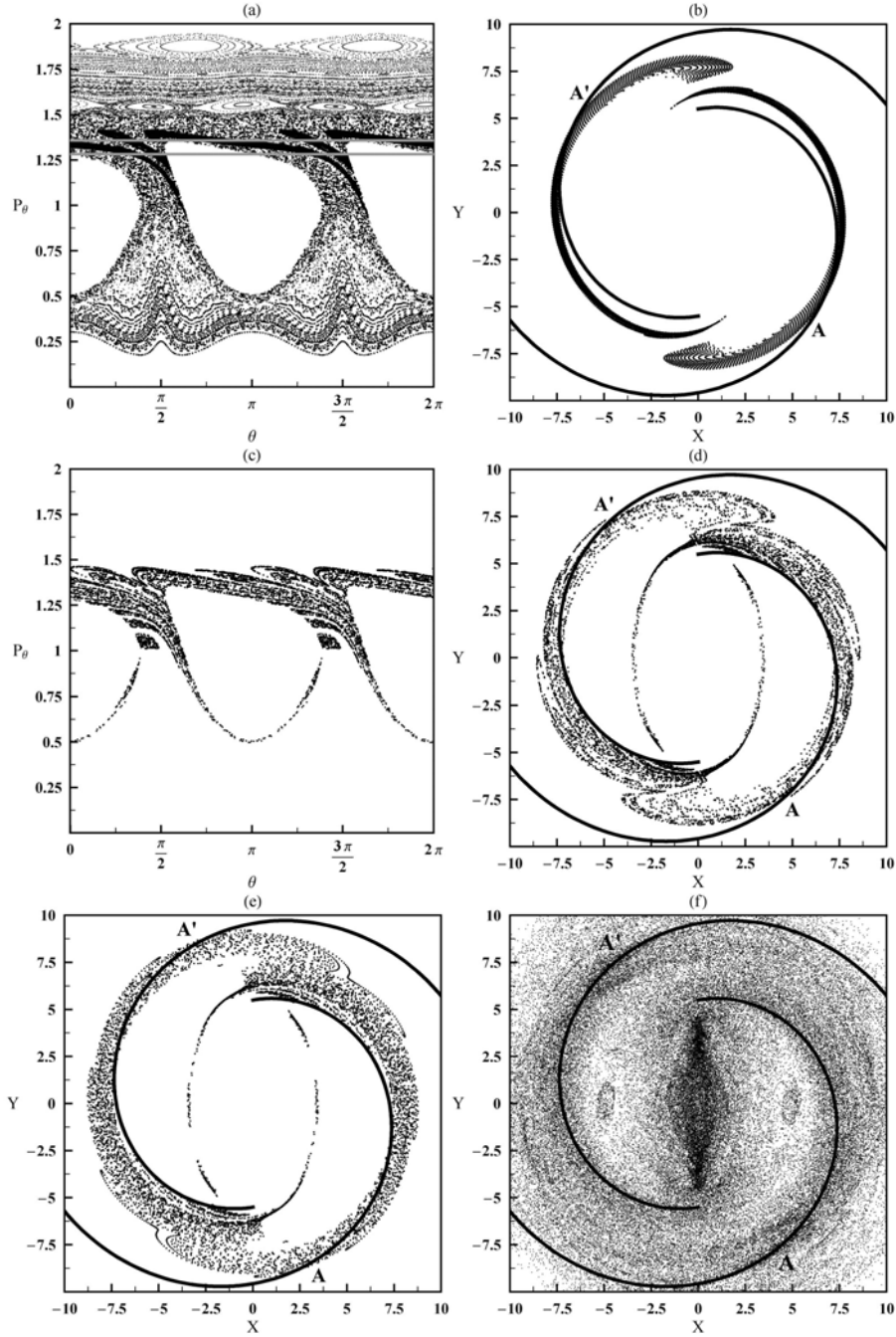


Fig. 12. (After [81]) (a) The two gray horizontal lines define a zone of initial conditions around the circular orbit of the unperturbed (axisymmetric) potential in the model A of [81]. The thick black dots are the images of these initial conditions after one iteration of the Poincaré map. The background thin points show the underlying phase space structure. (b) The first image of the Poincaré map of the initial conditions specified in (a) as it appears in the configuration space. After only one iteration this set of points already acquires the form of the invariant manifolds, i.e., it yields trailing spiral arms. The projection of the tenth iterates of the same points on the surface of section and on the configuration space are shown in (c) and (d) respectively. (e) Same as in (d) when the initial conditions are taken in a zone four times as wide as the zone of (a). (f) A full ‘response model’ for the same galaxy produced according to the methodology suggested in [57]. The black curves represent the minima of the imposed spiral potential.

In this method, which is based on considerations of *gas* rather than stellar dynamics, the only dynamical assumption made is that the gas satisfies a mass-preservation condition equivalent to the ‘surface continuity equation’:

$$\frac{\partial \Sigma(\mathbf{r}, t)}{\partial t} + \nabla \cdot (\Sigma(\mathbf{r}, t) \mathbf{v}) = 0 \quad (51)$$

where coordinates and velocities are taken with respect to the rotating frame. Starting from this assumption, it is possible to show that the pattern speed is given by

$$\Omega_p(y) = \frac{\int_{-\infty}^{\infty} \Sigma v_y dx}{\int_{-\infty}^{\infty} \Sigma dx} \quad (52)$$

where x is the axis coinciding with the line of nodes of the galaxy (intersection of the disc plane with the sky plane), and $v_y = v_{LOS}/\sin i$, where v_{LOS} are measured velocities along the line of sight as a function of x and $\sin i$ is the inclination of the galaxy, which can be inferred by independent observations. This method (or variants of) has given some promising results in the case of galactic bars (see e.g. [5]). Alternatively, it has been proposed that Ω_p could be determined by locating the distance of corotation via Canzian’s method [11, 12]. In this method, the dynamical assumption is that the density wave has a ‘grand design’ $m = 2$ component. Then, it is argued that starting from a map of line-of-sight velocities on the disc, corotation lies at the circle where the leading component of the Fourier analysis of this map changes from $m = 1$ to $m = 3$. The validity of assumptions like the above, and the extent to which we can rely on such methods for the determination of Ω_p , has to be judged by comparison with independent estimates of Ω_p obtained by dynamical modeling. This is a very active and interesting current area of research.

ii) *Live haloes*. In our analysis so far it has been tacitly assumed that the halo influences the disc through a rigid potential. However, it has been demonstrated, both theoretically and by numerical simulations, that haloes are *responsive* to discs, the main interaction being due to a transfer of angular momentum from the disc to the halo (e.g. [3, 34, 53, 70, 71]). This transfer is most effective at resonances. In the case of bars, this leads to a slowing down of the bar accompanied by a gain of halo rotation, until the disc-halo system is locked to a unique angular velocity. Due to the change in Ω_p by the slowing down of the bar, all resonances on the disc are gradually shifted outwards. This phenomenon induces significant *secular evolution* of the system, while it may also cause an evolution in the morphology and kinematics of the halo, which in general is a triaxial body subject to the dynamical processes discussed in Part I. This type of evolution is non-periodic, but it can be described as an adiabatic process, at least in its later stages. Dynamical studies showing how one should implement resonant perturbation theory in this case are missing, and would also constitute a promising area of research.

iii) Finally, we have only examined the case where the whole density wave on the disc rotates with a unique pattern speed. However, there are indications both from observations, and dynamical or N-Body models, of the existence of *multiple pattern speeds* [9, 26, 45, 64, 67, 69]. In the most usual case, an inner structure (bar) rotates quickly, and an outer structure (spiral arms) rotates more slowly. This effect is linked to a scenario in which the spiral structure is described as a *recurrent* phenomenon (see [73] for a review). Namely, spiral arms can follow consecutive phases of disappearance and reappearance, whose sequence can be resonant with the bar. When the spiral arms are present, the ratio of the bar and spiral pattern speeds can be close to resonant values, e.g. 2:1, 3:1, etc. In the case of two pattern speeds, assuming that one of the two rotating components is much weaker than the other, the Hamiltonian (31) is generalized to:

$$H(\theta_r, \theta, J_r, J_\theta, I) = \kappa_c J_r + \frac{1}{2r_c^2} J_\theta^2 + (\Omega_c - \Omega_p) J_\theta + \omega_3 I_3 \sum_{|m|+|n|+|l|=0}^{\infty} F_{m,n}(J_r, J_\theta) e^{i(n\theta_r + m\theta + l\theta_3)} \quad (53)$$

where I_3 is a dummy action, while $\theta_3 = \omega_3 t$ is the phase of time-dependent terms in the potential due to the angular change of the position of the weak pattern in the frame co-rotating with the

speed of the strong pattern. Under this form of the Hamiltonian, the problem can be viewed as the analog of the problem of *indirect effects* (e.g. of Jupiter) in mean motion resonances. Analytical studies of this problem, aiming to locate the fundamental invariant objects, their stability, and phase space structure in general would constitute a significant advancement of galactic disc dynamics.

Fruitful discussions with Profs. G. Contopoulos, P. Patsis and Drs. C. Kalapotharakos and M. Harsoula are gratefully acknowledged. I thank P. Tsoutsis for much help in calculations and in the production of some figures, and Liana Tsigarida for a careful reading of the manuscript.

References

1. E. Athanassoula, New York Acad. Sci. Ann. **596**, 181 (1990)
2. E. Athanassoula, S. Morin, H. Wozniak, D. Puy, M.J. Pierce, J. Lombard, A. Bosma, Mon. Not. R. Astron. Soc. **245**, 130 (1990)
3. E. Athanassoula, Astrophys. J. **569**, 83 (2002)
4. E. Athanassoula, M. Romero-Gomez, A. Bosma, J.J. Masdemont, Mon. Not. R. Astron. Soc. **400**, 1706 (2009)
5. J.E. Beckman, K. Fathi, N. Piñol, S. Toonen, O. Hernandez, C. Carignan, Astron. Soc. Pacific Conf. Proc. **396**, 401 (2008)
6. G. Bertin, *Dynamics of Galaxies* (Cambridge University Press, Cambridge, 2000)
7. J. Binney, S. Tremaine, *Galactic Dynamics*, 2nd edn. (Princeton University Press, New Jersey, 2008)
8. D. Boccaletti, G. Pucacco, *Theory of Orbits* (Springer, Berlin, 1996)
9. V. Boonyasait, P.A. Patsis, S.T. Gottesman, New York Acad. Sci. Ann. **1045**, 203 (2005)
10. R. Buta, S. Vasylyev, H. Salo, E. Laurikainen, Astron. J. **130**, 506 (2005)
11. B. Canzian, Astrophys. J. **414**, 487 (1993)
12. B. Canzian, R.J. Allen, Astrophys. J. **479**, 723 (1993)
13. S. Chandrasekhar, *Ellipsoidal Figures of Equilibrium* (Yale University Press, N. Haven, 1969)
14. G. Contopoulos, Z. Astrophys. **49**, 273 (1960)
15. G. Contopoulos, Astrophys. J. **201**, 566 (1975)
16. G. Contopoulos, Astron. Astrophys. **64**, 323 (1978)
17. G. Contopoulos, Astron. Astrophys. **81**, 198 (1980)
18. G. Contopoulos, Th. Papayiannopoulos, Astron. Astrophys. **92**, 33 (1980)
19. G. Contopoulos, Astron. Astrophys. **102**, 265 (1981)
20. G. Contopoulos, P. Grosbol, Astron. Astrophys. **155**, 11 (1986)
21. G. Contopoulos, P. Grosbol, Astron. Astrophys. **197**, 83 (1988)
22. G. Contopoulos, P. Grosbol, Ann. Rev. Astron. Astrophys. **1**, 261 (1989)
23. G. Contopoulos, M. Harsoula, N. Voglis, R. Dvorak, J. Phys. A: Math. Gen. **32**, 5213 (1999)
24. G. Contopoulos, *Order and Chaos in Dynamical Astronomy* (Springer, New York, 2004)
25. G. Contopoulos, M. Harsoula, Int. J. Bifur. Chaos **18**, 1 (2008)
26. E.M. Corsini, V.P. Debattista, J.A.L. Aguerri, Astrophys. J. **599**, 29 (2003)
27. D. Earn, Astrophys. J. **465**, 91 (1996)
28. C. Efthymiopoulos, G. Contopoulos, R. Dvorak, N. Voglis, J. Phys. A **30**, 8167 (1997)
29. C. Efthymiopoulos, Cel. Mech. Dyn. Astron. **102**, 49 (2008)
30. C. Efthymiopoulos, N. Voglis, C., in *Topics in Gravitational Dynamics*, edited by D. Benest, C. Froeschlé, E. Lega, Lect. Notes Phys. **729** (Springer, Berlin, 2008), p. 297
31. C. Efthymiopoulos, G. Contopoulos, N. Voglis, R. Dvorak, J. Phys. A: Math. Gen. **30**, 8167 (1997)
32. N.M. Ferrers, Quart. J. Pure Appl. Math. **14**, 1 (1877)
33. K.C. Freeman, Astrophys. J. **160**, 811 (1970)
34. B. Fuchs, E. Athanassoula, Astron. Astrophys. **444**, 455 (2005)
35. R. Fux, Astron. Astrophys. **373**, 511 (2001)
36. D.M. Grobman, Dokl. Akad. Nauk SSSR **128**, 880 (1959)
37. P. Grosbøl, in *Galaxies and Chaos*, edited by G. Contopoulos, N. Voglis, Lect. Notes Phys. **626**, 201 (2003)
38. F. Gustavson, Astron. Astrophys. **71**, 670 (1966)
39. P. Hartman, Proc. Amer. Math. Soc. **11**, 610 (1960)

40. M. Hénon, C. Heiles, *Astron. Astrophys.* **69**, 73 (1964)
41. C. Kalapotharakos, P.A. Patsis, P. Grosbol, *Mon. Not. R. Astron. Soc.* **403**, 83 (2010)
42. A. Kalnajs, *Astrophys. J.* **166**, 275 (1971)
43. A. Kalnajs, *Proc. Astron. Soc. Australia* **2**, 174 (1973)
44. D.E. Kaufmann, G. Contopoulos, *Astron. Astrophys.* **309**, 381 (1996)
45. D.E. Kaufmann, P. Patsis, *Astrophys. J.* **624**, 693 (2005)
46. E. Laurikainen, H. Salo, *Mon. Not. R. Astron. Soc.* **337**, 1118 (2002)
47. C.C. Lin, F.H. Shu, *Astrophys. J.* **140**, 646 (1964)
48. B. Lindblad, *Astrophys. J.* **92**, 1 (1940)
49. B. Lindblad, *Stockholm Obs. Ann.* **19**, 1 (1956)
50. B. Lindblad, *Stockholm Obs. Ann.* **21**, 8 (1961)
51. P.A.B. Lindblad, P.O. Lindblad, E. Athanassoula, *Astron. Astrophys.* **313**, 65 (1996)
52. D. Lynden-Bell, A. Kalnajs, *Mon. Not. R. Astron. Soc.* **157**, 1 (1972)
53. I. Martinez-Valpuesta, I. Shlosman, C. Heller, *Astrophys. J.* **637**, 214 (2006)
54. P.A. Patsis, G. Contopoulos, P. Grosbol, *Astron. Astrophys.* **243**, 373 (1991)
55. P.A. Patsis, C. Efthymiopoulos, G. Contopoulos, N. Voglis, *Astron. Astrophys.* **326**, 493 (1997)
56. P. Patsis, Ch. Skokos, E. Athanassoula, *Mon. Not. R. Astron. Soc.* **337**, 578 (2002)
57. P.A. Patsis, *Mon. Not. R. Astron. Soc.* **369**, 56 (2006)
58. P.A. Patsis, D. Kaufmann, S.T. Gottesman, V. Bonnyasait, *Mon. Not. R. Astron. Soc.* **394**, 142 (2009)
59. D. Pfenniger, *Astron. Astrophys.* **134**, 373 (1984)
60. D. Pfenniger, D. Friedli, *Astron. Astrophys.* **252**, 75 (1991)
61. B. Pichard, M. Martos, E. Moreno, *Astrophys. J.* **609**, 144 (2004)
62. A.C. Quillen, J.A. Frogel, R.A. Gonzalez, *Astrophys. J.* **437**, 162 (1994)
63. A.C. Quillen, *Astrophys. Sp. Sci. Proc.* **9**, 191 (2009)
64. P. Rautiainen, H. Salo, *Astron. Astrophys.* **348**, 747 (1999)
65. M. Romero-Gomez, J.J. Masdemont, E.M. Athanassoula, C. Garcia-Gomez, *Astron. Astrophys.* **453**, 39 (2006)
66. M. Romero-Gomez, E.M. Athanassoula, J.J. Masdemont, C. Garcia-Gomez, *Astron. Astrophys.* **472**, 63 (2007)
67. M. Rozas, M.J. Sempere, *Astron. Soc. Pacific Conf. Proc.* **215**, 117 (2000)
68. Schwarzschild, M., *Astrophys. J.* **232**, 236 (1979)
69. J.A. Sellwood, L.S. Sparke, *Mon. Not. R. Astron. Soc.* **231**, 25
70. J.A. Sellwood, V.P. Debattista, *Lect. Notes Phys.* **474**, 43 (1996)
71. J.A. Sellwood, V.P. Debattista, *Astrophys. J.* **639**, 868 (2006)
72. J.A. Sellwood, *Dynamics of Disks and Warps* [ArXiv1006.4855] (2010)
73. J.A. Sellwood, *New Developments in Spiral Structure theory* [ArXiv1001.5430] (2010)
74. C.L. Siegel, J. Moser, *Lectures on Celestial Mechanics* (Springer, Heidelberg, 1991)
75. Ch. Skokos, P. Patsis, E. Athanassoula, *Mon. Not. R. Astron. Soc.* **333**, 847 (2002)
76. L.S. Sparke, J.A. Sellwood, *Mon. Not. R. Astron. Soc.* **225**, 653 (1987)
77. A. Toomre, *Astrophys. J.* **138**, 385 (1962)
78. S. Tremaine, M. Weinberg, *Astrophys. J.* **282**, L5 (1984)
79. S. Tremaine, M. Weinberg, *Mon. Not. R. Astron. Soc.* **209**, 729 (1984)
80. P. Tsoutsis, C. Efthymiopoulos, N. Voglis, *Mon. Not. R. Astron. Soc.* **387**, 1264 (2008)
81. P. Tsoutsis, C. Kalapotharakos, Efthymiopoulos, G. Contopoulos, *Astron. Astrophys.* **495**, 743 (2009)
82. N. Voglis, P. Tsoutsis, C. Efthymiopoulos, *Mon. Not. R. Astron. Soc.* **373**, 280 (2006)
83. N. Voglis, I. Stavropoulos, C. Kalapotharakos, *Mon. Not. R. Astron. Soc.* **372**, 901 (2006)