AAA Workshop Series
Asociación Argentina de Astronomía

*3rd La Plata International
School on Astronomy and Geophysics*

**Chaos, Diffusion and Non-integrability in
Hamiltonian Systems – Applications to Astronomy**

La Plata Observatory, July 11 - 15, 2011

AAA Workshop Series
Asociación Argentina de Astronomía

3rd *La Plata International School on Astronomy and Geophysics*

Chaos, Diffusion and Non-integrability in Hamiltonian Systems - Applications to Astronomy

La Plata Observatory, July 11 - 15, 2011

Preface

The present volume contains the lectures and oral contributions of the participants of the Third La Plata International School on Astronomy and Geophysics, which was held at the Facultad de Ciencias Astronómicas y Geofísicas de la Universidad Nacional de La Plata (FCAGLP) in July 2011, with the participation of ten lecturers and about forty researchers, graduate and advanced students from several countries.

The School was devoted to the subject of *Chaos, diffusion and non-integrability in Hamiltonian systems – Applications to Astronomy.* The professors were known authorities in the field. The lectures covered a variety of topics ranging from fundamental concepts, theory and mathematical tools, integrability and non-integrability, chaos and resonances in the Solar System and dynamics of exoplanets, chaos in elliptical galaxies and diffusion theories in multidimensional Hamiltonian systems. Also, an open session took place devoted to *The gravitational effects of faraway matter: from galaxy clusters to the rotation curves in spiral galaxies.*

Already in the early sixties, pioneering works such as those of M. Hénon and C. Heiles, as well as G. Contopoulos, revealed the importance of chaos in a variety of phenomena encountered in Dynamical Astronomy. In the present days, however, several aspects concerning how to cope with chaos and global instabilities are still largely open. All these issues are essential to understand or explain much of, for instance, Solar System dynamics or the dynamics of galaxies.

This book is organized in twelve Chapters. The first part corresponds to the lectures delivered by the professors of the School, and the second one to the oral contributions of some of the participants. All the material published in the present volume was subject to expert review. We acknowledge all the reviewers for their kind disposition.

This School was made possible only thanks to the efforts of the professors and the participants, as well as the members of the Local Organizing Committee: Luciano Darriba, Nicolás Maffione, Martín Mestre, Octavio Miloni, Rocío Paez and María Paula Ronco. The valuable assistance of Federico Bareilles in the edition of the present volume is acknowledged.

Additional thanks are due to the authorities of the FCAGLP, the Dean Prof. Adrián Brunini and his staff, and the Government Council. We would like to thank them all for rendering this event an academic success as well as a delightful experience.

We are also indebted to the different institutions that provided support to make this School a reality: the Facultad de Ciencias Astronómicas y Geofísicas (UNLP), the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), the Comisión de Investigaciones Científicas (CIC) and the Centro Latinoamericano de Física (CLAF).

Pablo Cincotta, Claudia Giordano & Christos Efthymiopoulos, Eds.
La Plata, July 2012.

ii

Index

I Lectures

II Oral contributions

Part I

Lectures

LECTURE

Canonical perturbation theory, stability and diffusion in Hamiltonian systems: applications in dynamical astronomy

C. Efthymiopoulos[1]

(1) Research Center for Astronomy and Applied Mathematics, Academy of Athens

Abstract.   We present some basic methods and techniques of canonical perturbation theory, as well as some of its applications in problems of stability and/or diffusion in dynamical astronomy. The methods presented are: i) the Birkhoff normal form, ii) the Kolmogorov normal form, iii) the resonant normal form, and iv) the hyperbolic normal form used in the computation of invariant manifolds of unstable periodic orbits in the chaotic regime. For each method we give concrete examples presented in some detail in order to facilitate study. In particular, we discuss a step by step implementation of a so-called 'book-keeping' algorithm by which all quantities (i.e. Hamiltonian, generating functions etc.) can be split in groups of terms of similar order of smallness. We explain why the book-keeping schemes presently suggested are particularly suitable in computer-algebraic implementations of normal forms. Also, for each method we explain the pattern by which small divisors are accumulated in the series terms at successive normalization steps, outlining why such accumulation leads to a divergent normalization process in the case of the Birkhoff normal form (both non-resonant or resonant), while it leads to a convergent normalization process in the case of the Kolmogorov normal form or the hyperbolic normal form. After these formal aspects, we present applications of canonical perturbation theory in concrete Hamiltonian dynamical systems appearing in problems of dynamical astronomy. In particular, we explain how resonant normal form theory is connected to the phenomenon of Arnold diffusion, as well as to estimates of the diffusion rate in the action space in systems of three (or more) degrees of freedom. We discuss how is 'book-keeping' implemented in paradigmatic cases, like the treatment of mean motion resonances in solar system dynamics, and the study of orbits in axisymmetric galaxies or in barred-spiral rotating galaxies. Finally, we give an example of implementation of normal form theory in the orbital version of the so-called 'density wave theory' of spiral arms in galaxies.

CONTENTS

1. INTRODUCTION

This is a set of lecture notes on methods and techniques of *canonical perturbation theory*, as well as on the latter's applications in the study of *diffusion processes and chaos* in physical systems related to dynamical astronomy.

The text represents, to a considerable extent, an elaborated transcript of the lectures given by the author during the third La Plata school on Astronomy and Geophysics, in July 2011. However, the structure of the text, and in particular, the sequence of presentation of the topics, has been substantially modified with respect to the structure of the lectures. Also, the text has been enriched by a number of additional topics, reference to which was judged necessary for reasons of completeness.

For a more complete study of the subject, the reader cannot but have recourse to excellent existing textbooks in the literature. As an indicative list of references for one or more of the topics treated below we mention: Ozorio de Almeida (1988), Siegel and Moser (1991), Arnold et al. (1993), Bocaletti and Pucacco (1996), Contopoulos (2002), Morbidelli (2002), and Ferraz-Mello (2007). Quite useful are also reviews and sets of lecture notes as, for example: Delshams and Gutiérrez (1996), Morbidelli and Guzzo (1997), Benettin (1999), Giorgilli and Locatelli (1999), Giorgilli (2002), Cincotta (2002), Chierchia (2008), as well as the review articles included in Benest et al. (2008) and (2010).

My own effort, in writing the present notes, has been to abandon a formal style of presentation in favor of an approach that could be helpful *in practice*. The goal is to show how canonical perturbation theory can be implemented, with much profit, in concrete computations referring to the study of stability, and/or diffusion in systems appearing mainly (but not exclusively) in celestial mechanics, and stellar and galactic dynamics. In fact, with the advent of modern computers (and computer-algebraic techniques), in recent years it has become possible to implement canonical perturbation theory at quite high expansion orders. This fact allows one to sustain that the theory's present usefulness renders it competitive to purely numerical methods, provided that one is able to write a program implementing some perturbative technique of interest at high order in the computer. This topic is discussed also by A. Giorgilli and M. Sansottera in the present volume of proceedings.

On the other hand, the introduction of computer-algebraic methods in the computation of normal forms has rendered clear that, in nearly every form of canonical perturbation theory, a computation can proceed in principle by more than one distinct algorithmic approaches. However, it is a fact that only a subset of such approaches prove to be practical and useful in the applications. Furthermore, a presentation of such practical approaches is often absent in literature reviews of the subject. A well known example refers to the so-called splitting of an analytic Hamiltonian function in parts of 'different order of smallness'. One can show (see Giorgilli (2002), pp.90–91) that the most practical splitting is in powers of a quantity $\sim e^{-\sigma}$, where σ is a positive constant characterizing the analyticity domain of the Hamiltonian in the space of angles. For reasons explained in section 3 below, such a splitting is clearly advantageous (and of nearly exclusive interest in practice) with respect to a more traditional splitting in powers of apparently more 'natural' small parameters, as for example the small parameter ϵ in a Hamiltonian of the form $H = H_0 + \epsilon H_1$, where H_0 is

the integrable part. In fact, the preference to a scheme based on powers of $e^{-\sigma}$, which explicitly appears in the works of Poincaré (1892) and Arnold (1963, see also Arnold et al. (1993)), is not so often discussed in modern references to perturbation theory.

In sections 2 and 3 below, we emphasize the practical aspects of normal form computations by introducing a formal procedure, called 'book-keeping'[1], by which we can systemize the most practical approach to such computations. We give specific examples of the book-keeping process in four different normal form methods: i) the non-resonant normal form of Birkhoff, ii) the normal form of Kolmogorov, employed in the proof of the KAM theorem, iii) the resonant (Birkhoff) normal form used in the study of diffusion along resonances, and iv) the hyperbolic normal form of Moser (1958), by which we can compute unstable periodic orbits and their asymptotic manifolds in the chaotic regime.

In each case, we combine the discussion on 'book-keeping' with a discussion of how *small divisors* are accumulated in the series terms at successive normalization orders, the latter being presently defined as successive orders in a suitably defined 'book-keeping parameter'. This discussion allows to demonstrate in a heuristic way how the pattern of accumulation of small divisors explains the (non-)convergence properties of the various methods, i.e., why the Birkhoff normal form, in both its non-resonant or resonant form, diverges, while the Kolmogorov and the hyperbolic normal forms are convergent within some domain of the phase space.[2].

After discussing the above formal issues, in section 4 we discuss one more issue related to diffusion in the weakly chaotic regime, namely the connection between resonant normal form theory and the example given by Arnold (1964) largely referred to as 'Arnold diffusion'. The main result shown here is that a practical implementation of normal form theory in the computer at a high order allows one to construct a set of good canonical variables (suggested first in the theoretical work of Benettin and Gallavotti (1986)), in which the phenomenon of Arnold diffusion can be visualized in a way particular convenient for quantitative study. In fact, we were recently able to construct such an example (Efthymiopoulos and Harsoula (2012)), briefly presented in the end of section 4. We should emphasize at this point that a study of Arnold diffusion using normal forms proves to be an approach complementary to another approach extensively discussed in the present school, namely the theory of diffusion developed by Chirikov (1979) using tools like the Melnikov - Arnold integral (see the review by Cincotta (2002)). In fact, the combination of the two methods serves as a bridge connecting the 'two sides of the river', i.e. normal form expansions and the use of Arnold-Melnikov techniques. Some discussion of this subject is made in subsection 4.3 below.

The present notes end with a reference to some basic Hamiltonian models encountered in dynamical astronomy, as for example, in the restricted three body problem, in axisymmetric galaxies, and in rotating barred-spiral galaxies. The variety of applications of canonical perturbation theory in dynamical astronomy

---

[1]Term suggested by T. Bountis (private communication).

[2]Quite pedagogic expositions of this subject can be found in various lectures by A. Giorgilli, see reference list.

renders hardly tractable any attempt to review them. Instead, we present be-low only the Hamiltonian formalisms associated with these specific examples, as well as the way by which book-keeping, and other formal aspects of normal form theory, can be implemented in each case. We hope that the collection of some basic formulae could motivate readers to develop their own implementations of the methods and principles discussed below. This is actually the overall purpose of the present lectures.

A final word: These notes are my effort to respond to the many inspiring (and on-going) interactions with a group of very motivated young astronomers that attended the school. I would like to mention in particular the LOC members Lu-ciano Darriba, Nicolás Maffione, Martín Mestre, Octavio Miloni, Rocío Paez and Paula Ronco. Let me hope that the present notes could fulfil in part their ex-pectations for a text providing practical help in their and their colleagues' study of a fascinating subject, namely Hamiltonian dynamical systems as applied to Dynamical Astronomy.

2. A 'WARM-UP' EXAMPLE: THE PERIODICALLY DRIVEN PENDULUM

2.1. Phenomenology

Let us consider a periodically driven pendulum model given by the Hamiltonian

$$H = \frac{p^2}{2} - \omega_0^2(1 + \epsilon(1 + p)\cos\omega t)\cos\psi \quad . \tag{1}$$

We postpone until section 5 a detailed discussion of the physical context in which a model like (1) could arise in systems of interest in dynamical astronomy. We

mention only that a model like (1) could represent a simplified version of the local form of Hamiltonian dynamics in cases of *resonances*, as, for example, the mean motion resonances in solar system dynamics or the disc resonances in spiral and barred galaxies.

The model (1) can be considered as a time-dependent system of one degree of freedom (called sometimes 'one and a half' degrees of freedom). In practice, it often proves helpful to eliminate the formal dependence of a Hamiltonian function like (1) on time. When this dependence is through trigonometric terms of the form $\cos[(n_1\omega_1 + n_2\omega_2 + ... + n_m\omega_m)t]$, i.e. with one or more (incommensurable) frequencies $\omega_1, \ldots, \omega_m$, a usual procedure by which we do the formal elimination is to 'upgrade' each of the quantities $\omega_i t$ to an angular variable $\phi_i = \omega_i t$, to which we associate a conjugate 'dummy' action I_i. It is then straightforward to show that Hamilton's equations under the original Hamiltonian $H(\psi, p; \omega_1 t, \omega_2 t, \ldots, \omega_m t)$ are equivalent to Hamilton's equations under a new Hamiltonian, called the extended Hamiltonian:

$$H' = H(\psi, p, \phi_1, \phi_2 \ldots, \phi_m) + \omega_1 I_1 + \ldots + \omega_m I_m \quad . \tag{2}$$

Implementing the above procedure in the Hamiltonian (1) (where we only have one external frequency, i.e. ω), we arrive at the extended hamiltonian:

$$H'(\psi, \phi, p, I) = \frac{p^2}{2} + \omega I - \omega_0^2 (1 + \epsilon(1 + p)\cos\phi)\cos\psi \quad . \tag{3}$$

In numerical studies, an advantage of employing the Hamiltonian (3) over (1) is that the energy $E = H'(\psi, \phi, p, I)$ is a preserved quantity along the orbital flow of the hamiltonian (3), while this is not so for the hamiltonian (1). Thus, when we solve numerically the equations of motion under the Hamiltonian (3), we can test the accuracy of numerical integrations by checking how well is the energy E preserved along the numerically computed orbits. This, at the expense of some extra time required to compute the time evolution of the dummy action I.

We proceed now in a rough investigation of the dynamical features of motion under the Hamiltonian (3) via a numerical example. To this end, we first fix numerical values for the frequencies, say, $\omega_0 = 0.2\sqrt{2}$ and $\omega = 1$. The motivation for such choice will be gradually recognized as we proceed.

In order to numerically visualize the properties of motion, a basic first step is to plot *phase portraits*, as we vary the perturbation parameter ϵ. In this it is helpful to introduce a convenient *surface of section*, i.e. a surface crossed by all trajectories. By plotting only the crossing points of the trajectories with the surface of section, we obtain two-dimensional plots easy to visualize and interpret. In systems with a periodic-in-time perturbation, a convenient choice of surface of section stems from noticing that the angle ϕ grows linearly in time (i.e. $\phi = \omega t$, setting $\phi = 0$ at $t = 0$), independently of the time evolution of the remaining variables ψ, p, I. Thus, in a system like (3) we can choose as surface of section the surface given by the condition $\phi \, mod \, 2\pi = 0$. Substituting this condition in the Hamiltonian (3), for fixed energy E, we find:

$$E = \frac{p^2}{2} + \omega I - \omega_0^2 (1 + \epsilon(1 + p))\cos\psi \quad . \tag{4}$$

This is a condition binding together the variables (ψ, p, I). Solving, for example, for I, we find:

$$I = \frac{1}{\omega}\left(E - \frac{p^2}{2} + \omega_0^2(1 + \epsilon(1 + p))\cos\psi\right) \quad . \tag{5}$$

Eq.(5) implies that in order to visualize dynamics, it suffices to plot the points (ψ, p) on a two-dimensional plot every time when ϕ becomes equal to a multiple of 2π, since, then, the value of I is specified completely. Equivalently, if we give some initial condition $\psi(0) = \psi_0$, $p(0) = p_0$, and an auxiliary initial condition for the dummy action I, e.g. $I(0) = 0$, we can compute numerically the time evolution $\psi(t), p(t)$ (and $I(t)$), and plot one point on the plane (ψ, p) whenever the time t is equal to a multiple of the perturber's period, i.e. at the times $t_1 = 2\pi/\omega$, $t_2 = 4\pi/\omega$, etc. The set of points $(\psi_i, p_i) = (\psi(t_i), p(t_i))$, $t_i = i(2\pi/\omega)$, $i = 1, 2, ...$ are called surface of section iterates. As a rule, computing a few thousand iterates per trajectory is enough to obtain a clear visual picture of the dynamics.[3]

Figure 1 shows the phase portraits (surfaces of section) of the Hamiltonian model (3) for two different values of ϵ, namely (a) $\epsilon = 0.04$, and (b) $\epsilon = 1$. These are chosen so as to represent two well distinct regimes characterizing hamiltonian dynamics. In particular:

(a) Figure 1a shows a typical phase portrait in the *nearly-integrable*, or *weakly chaotic* case. Its main feature is that the phase space is filled for the most by invariant tori. In fact, we distinguish three types of such tori in Fig.1a: i) librational tori (closed curves), ii) rotational tori (open curves extending from $-\pi < \psi \le \pi$), and iii) tori around higher order resonances (islands of stability). As we will see, we can establish that the trajectories giving rise to such invariant curves are *quasi-periodic*, i.e. they can be represented by a sum of trigonometric terms with one or more frequencies. Also, these orbits are called regular, i.e. they introduce no chaos, and they have zero Lyapunov characteristic exponents. However, Fig.1a shows also some local chaos around the pendulum separatrix layer. But a main feature of the orbits in the chaotic layer is that these orbits cannot exhibit macroscopic transport beyond a domain confined by the presence of invariant tori. This is a property of systems of at most two degrees of freedom,

---

[3]We note, in passing, that the sequence $(\psi_0, p_0) \to (\psi_1, p_1) \to (\psi_2, p_2) \to \ldots$ can be viewed also as a *mapping*, i.e. the point (ψ_0, p_0) is mapped to (ψ_1, p_1), then (ψ_2, p_2) etc. In fact, in the study of hamiltonian dynamics, we often use as models explicit mappings, like, for example, a 2D mapping of the form

$$\psi_{i+1} = f(\psi_i, p_i), \quad p_{i+1} = g(\psi_i, p_i) \tag{6}$$

where the functions f, g are explicitly given, instead of being computed as the surface of section iterates of some Hamiltonian model. Such is the case of the celebrated Chirikov's (1979) *standard mapping*:

$$\begin{aligned}\psi_{i+1} &= \psi_i + p_i + K\sin\psi_i \quad (mod 2\pi)\\ p_{i+1} &= p_i + K\sin\psi_i\end{aligned} \tag{7}$$

where K is a constant 'non-linearity parameter'. The mapping (7), or variants of it in two or more dimensions, have been used as prototypes for many studies of chaotic diffusion in conservative systems.
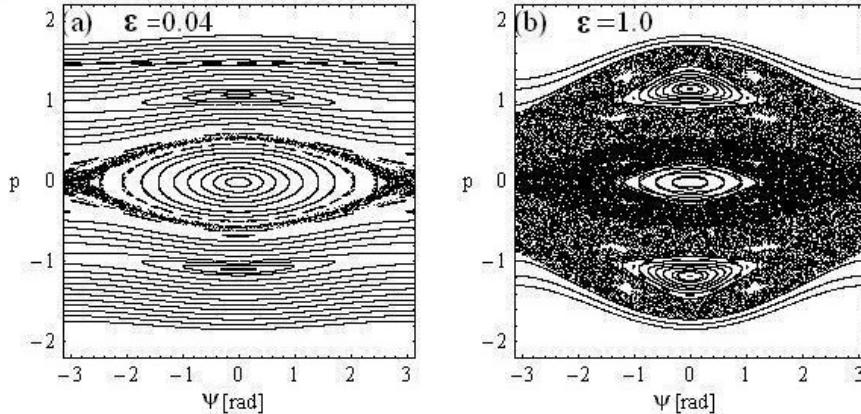
Figure 1.   Surfaces of section of the perturbed pendulum model (Hamiltonian (3)) for (a) $\epsilon = 0.04$, and (b) $\epsilon = 1$.

while, as we will see in section 4, in systems of three or more degrees of freedom such transport is allowable via the so-called mechanism of *Arnold diffusion*.

(b) Figure 1b exemplifies the so-called *strong chaos* regime, where most trajectories are chaotic. In the surface of section, chaos shows up as an apparently random distribution of points on the surface, called a 'chaotic layer' or 'chaotic sea'. It should be noted, however, that the motion in a chaotic layer like in Fig.1b has also some underlying structure and obeys laws, which render chaotic motions qualitatively very different from random motions (see Contopoulos (2002)).

In the remaining part of section 2, we will exploit the numerical example of figure 1 as a basis in order to introduce some basic forms and techniques of canonical perturbation theory. Most of the methods presented below deal with a description of the regime of *regular dynamics*. These are useful when the perturbation parameter ϵ is relatively small, for example $\epsilon < 0.1$ in the case of the Hamiltonian (3). Are such perturbation values relevant to the size of various perturbations encountered in systems of interest in dynamical astronomy? We will see that, while we can definitely identify cases of strong chaos in astronomical systems, the case of systems approximated by nearly-integrable (or weakly chaotic) dynamics appears also quite frequently. Table I shows different types of astronomical systems, and the sort of perturbations that can appear in them along with a rough estimate of the expected size of various perturbations. We observe that typical perturbations in dynamical astronomy are in the range $10^{-4} \leq \epsilon \leq 1$. In this regime, most forms of canonical perturbation theory are applicable to some extent. The domain where this is most evident is *solar system dynamics*, which is historically the domain which has lead to most developments in canonical perturbation theory. The importance of nearly-integrable systems in the study of solar system dynamics has been emphasized by H. Poincaré (1892) in his seminal *"Méthodes Nouvelles de la Mécanique Celeste"*. In fact, the study of motions under a Hamiltonian of the form

$$H = H_0 + \epsilon H_1$$

where H_0 is an integrable model and ϵH_1 is a function analytic in ϵ, (i.e. developable as a convergent series in powers of ϵ), is called by Poincaré the "fundamental problem of dynamics".

In subsection 2.10, however, we will see that canonical perturbation theory can be used also with profit in describing even some features of chaotic dynamics, as in Fig.1b. In particular, canonical perturbation theory can be used in order to compute the asymptotic *invariant manifolds* emanating from unstable periodic orbits located in the chaotic subset of the phase space. The invariant manifolds are objects of great complexity, whose form, nevertheless, can be computed by the method of hyperbolic normal forms, discussed first by Moser (1958), and implemented in the canonical framework by Giorgilli (2001). The study of manifold dynamics is a quite modern subject that has led to many applications in celestial mechanics and galactic dynamics (Voglis et al. (2006), Romero-Gomez et al. (2006, 2007), Tsoutsis et al. (2008, 2009)), and even space-flight dynamics (see Perozzi and Ferraz-Mello (2010) and references there in, and Gómez and Barrabés (2011)).

Table 1: Perturbations in astronomical systems

System	Type of perturbation	perturbation value
Earth sattelite dynamics	Earth's oblateness	$\sim 10^{-2}$
Solar system dynamics	Mass of Jupiter	10^{-3}
	Mass of other giant planets	10^{-4}
	Eccentricities	$10^{-2} - 10^{-1}$
	Inclinations	$10^{-2} - 10^{-1}$
Extrasolar planetary systems	Mass of giant planets	$10^{-3} - 10^{-2}$
	Eccentricities	$10^{-1} - 1$
Spiral - Barred galaxies	Spiral perturbation	$10^{-2} - 10^{-1}$
	Bar perturbation	$10^{-1} - 1$
Elliptical galaxies	Ellipticity	~ 0.5, but equipotential surfaces rounder $\sim 0.1 - 0.3$
	Triaxiality	$0.1 - 0.5$
	Ratio of central mass to galaxy mass	$10^{-5} - 10^{-2}$
Charged particle motions in magnetic fields	$B_\perp/B_0$	$10^{-3} - 10^{-1}$

2.2. The concept of normal form

We start our discussion of how to implement canonical perturbation theory in the study of regular motions in a hamiltonian system like (3) by recalling first some basic notions related to the theory of *normal forms* in Hamiltonian systems. This theory will be implemented in the case of the extended Hamiltonian H' of

Eq.(3) (in the sequel, for simplicity we drop the prime symbol and refer to the latter simply as H).

In the context of canonical perturbation theory, a *normal form* can be defined as *a Hamiltonian function yielding a simple-to-analyze dynamics*.

It should be made clear at once that, in the above definition, 'simple-to-analyze' does *not* necessarily mean 'integrable'. In fact, normal forms that are integrable Hamiltonians appear only in particular cases of non-resonant, or at most simply-resonant, models. In general, however, a multiply-resonant normal form is a non-integrable Hamiltonian, and in fact, the dynamics under such a normal form could imply even a large degree of chaos (see section 4). However, normal forms are Hamiltonians having some particular properties to be discussed below, which render their study simpler than the study of the original Hamiltonian from which they arise.

What is precisely the relation between a normal form and the original Hamiltonian of interest? The normal form arises after implementing a *canonical transformation* to the variables appearing in the original Hamiltonian. In the example of the hamiltonian (3), the overall procedure is the following:

i) We are interested in modeling, e.g., the regular trajectories appearing in Fig.1a, by computing one or more normal form models associated with the Hamiltonian (3).

ii) To this end, we introduce, by a suitable method, a canonical transformation allowing to pass from the old canonical variables (ψ, ϕ, p, I) to new canonical variables (ψ', ϕ', p', I').

iii) We substitute this transformation in the original Hamiltonian, and find the form of the Hamiltonian as expressed in the new variables. We then find that, after the substitution, the new Hamiltonian takes typically the form of a sum of two parts

$$H_{new}(\psi', \phi', p', I') =$$
$$H(\psi(\psi', \phi', p', I'), \phi(\psi', \phi', p', I'), p(\psi', \phi', p', I'), I(\psi', \phi', p', I'))$$
$$= Z(\psi', \phi', p', I') + R(\psi', \phi', p', I') \tag{8}$$

where the terms $Z(\psi', \phi', p', I')$ and $R(\psi', \phi', p', I')$ are called *normal form* and *remainder* respectively. The normal form term Z is the one whose significance has been discussed already. However, the transformed hamiltonian contains also the remainder term R. This term is important, because it tells us how much the dynamics of the original Hamiltonian really differs from the dynamics of the normal form. In fact, most schemes of perturbation theory are formulated upon the requirement to have proper control over the growth of the *size of the remainder* (see below).

The study of the influence of the remainder on dynamics requires new techniques, going beyond the formal aspects of canonical perturbation theory. A basic example of such techniques is provided by the theory of diffusion developed by Chirikov (1979, see the review by Cincotta (2002) and section 4 below).

We will now examine in detail the technical aspects of how to construct a normal form in practice for the Hamiltonian (3). We first observe that, in order to deal with canonical transformations, we are in need of a formal apparatus performing such transformations in a way convenient to the derivation of a normal form. The technique of *Lie series transformations* is widely used to this

end, and easily transferable to a computer-algebraic program. To this we now turn our attention.

2.3. Canonical transformations by Lie series

The idea of a *Lie series canonical transformation* is very simple: just consider an arbitrary function $\chi(\psi, \phi, p, I)$, and compute the flow produced under Hamilton's equations of motion if, instead of H, χ was supposed to be the Hamiltonian. The equations of motion would then be:

$$\dot{\psi} = \frac{\partial \chi}{\partial p}, \quad \dot{\phi} = \frac{\partial \chi}{\partial I}, \quad \dot{p} = -\frac{\partial \chi}{\partial \psi}, \quad \dot{I} = -\frac{\partial \chi}{\partial \phi} \quad . \tag{9}$$

Let $\psi(t), \phi(t), p(t), I(t)$ be a solution of Eqs.(9) for some choice of initial conditions $\psi(0) = \psi_0$, $\phi(0) = \phi_0$, $p(0) = p_0$, and $I(0) = I_0$. The key remark, now, is the following. For any time t, the mapping of the variables in time, namely

$$(\psi_0, \phi_0, p_0, I_0) \to (\psi_t, \phi_t, p_t, I_t)$$

can be proven to be a canonical transformation (see, for example, Arnold (1978)). In that sense, any arbitrary function $\chi(\psi, \phi, p, I)$ can be thought of as a function which can generate an infinity of different canonical transformations, via its Hamilton equations of motion solved for infinitely many different values of the time t. In fact, from this viewpoint t can be considered as a parameter which defines, according to its value, the whole family of the canonical transformations generated by χ.

How practical is this method in defining canonical transformations? Clearly, for an arbitrary function χ, the task of solving Eqs.(9) for every value of t is hardly tractable. However, we can note that for t small enough, a solution of the initial value problem of Eqs.(9) is always possible via *Taylor series*. This, because by knowing explicitly how the first derivatives $\dot{\psi}, \dot{\phi}, \dot{p}, \dot{I}$ are expressed as functions of the canonical variables (ψ, ϕ, p, I) (via Eqs.(9)), we can find similar expressions for the time derivatives of *all orders*, depending on the same quantities. We have, for example:

$$
\begin{aligned}
\frac{d^2\psi}{dt^2} &= \frac{d}{dt}\left(\frac{\partial \chi}{\partial p}\right) = \frac{\partial}{\partial \psi}\left(\frac{\partial \chi}{\partial p}\right)\dot{\psi} + \frac{\partial}{\partial \phi}\left(\frac{\partial \chi}{\partial p}\right)\dot{\phi} + \frac{\partial}{\partial p}\left(\frac{\partial \chi}{\partial p}\right)\dot{p} + \frac{\partial}{\partial I}\left(\frac{\partial \chi}{\partial p}\right)\dot{I} \\
&= \frac{\partial^2 \chi}{\partial \psi \partial p}\frac{\partial \chi}{\partial p} + \frac{\partial^2 \chi}{\partial \phi \partial p}\frac{\partial \chi}{\partial I} - \frac{\partial^2 \chi}{\partial^2 p}\frac{\partial \chi}{\partial \psi} - \frac{\partial^2 \chi}{\partial I \partial p}\frac{\partial \chi}{\partial \phi} \quad .
\end{aligned}
\tag{10}
$$

We note that, after all operations, the second derivative $d^2\psi/dt^2$ has been expressed in Eq.(10) in terms of the function $\chi(\psi, \phi, p, I)$ and some partial derivatives of it. This implies that Eq.(10) resumes the form $d^2\psi/dt^2 = F_{\psi,2}(\psi, \phi, p, I)$, i.e. the second derivative $d^2\psi/dt^2$ can be expressed as a function of the canonical variables (ψ, ϕ, p, I). By repeating the above procedure, it is straightforward to see that the same holds true for the derivatives of all orders, i.e. we can derive expressions of the form $d^n\psi/dt^n = F_{\psi,n}(\psi, \phi, p, I)$ for all $n = 1, 2, \ldots$.

Why is this required in practice? If we express the solution $\psi(t)$ as a Taylor series

$$\psi(t) = \psi(0) + \frac{d\psi(0)}{dt}t + \frac{1}{2}\frac{d^2\psi(0)}{dt^2}t^2 + \ldots \tag{11}$$

and substitute in (11) the expressions for all functions $d^n\psi/dt^n$ by their equivalent functions $F_{\psi,n}(\psi,\phi,p,I)$, we find an expression of the form

$$\psi_t = \psi_0 + F_{\psi,1}(\psi_0,\phi_0,p_0,I_0)t + \frac{1}{2}F_{\psi,2}(\psi_0,\phi_0,p_0,I_0)t^2 + \dots \qquad (12)$$

where $\psi_t \equiv \psi(t)$, $\psi_0 \equiv \psi(0)$, etc. This procedure can be repeated for all remaining variables, yielding finally expressions of the form

$$
\begin{aligned}
\psi_t &= \psi_0 + F_{\psi,1}(\psi_0,\phi_0,p_0,I_0)t + \frac{1}{2}F_{\psi,2}(\psi_0,\phi_0,p_0,I_0)t^2 + \dots \\
\phi_t &= \phi_0 + F_{\phi,1}(\psi_0,\phi_0,p_0,I_0)t + \frac{1}{2}F_{\phi,2}(\psi_0,\phi_0,p_0,I_0)t^2 + \dots \qquad (13) \\
p_t &= p_0 + F_{p,1}(\psi_0,\phi_0,p_0,I_0)t + \frac{1}{2}F_{p,2}(\psi_0,\phi_0,p_0,I_0)t^2 + \dots \\
I_t &= I_0 + F_{I,1}(\psi_0,\phi_0,p_0,I_0)t + \frac{1}{2}F_{I,2}(\psi_0,\phi_0,p_0,I_0)t^2 + \dots
\end{aligned}
$$

Expressions like (13) are called *Lie series*, and they provide a family of formal canonical transformations for any value of the time variable t, which, for this reason, can be considered as a parameter characterizing this family. Of course, in order that the expressions (13) have meaning, the series in the r.h.s. of all equations must be convergent. We will examine the question of convergence in some detail below, but for the moment let us assume that the function χ, and its derivatives, are small enough so that the series (13) are convergent when the time is $t = 1$. We then obtain a *Lie canonical transformation*, from (ψ_0,ϕ_0,p_0,I_0) to (ψ_1,ϕ_1,p_1,I_1), i.e.:

$$
\begin{aligned}
\psi_1 &= \psi_0 + F_{\psi,1}(\psi_0,\phi_0,p_0,I_0) + \frac{1}{2}F_{\psi,2}(\psi_0,\phi_0,p_0,I_0) + \dots \\
\phi_1 &= \phi_0 + F_{\phi,1}(\psi_0,\phi_0,p_0,I_0) + \frac{1}{2}F_{\phi,2}(\psi_0,\phi_0,p_0,I_0) + \dots \qquad (14) \\
p_1 &= p_0 + F_{p,1}(\psi_0,\phi_0,p_0,I_0) + \frac{1}{2}F_{p,2}(\psi_0,\phi_0,p_0,I_0) + \dots \\
I_1 &= I_0 + F_{I,1}(\psi_0,\phi_0,p_0,I_0) + \frac{1}{2}F_{I,2}(\psi_0,\phi_0,p_0,I_0) + \dots
\end{aligned}
$$

The function χ is called a *Lie generating function*, i.e. it "generates" a canonical transformation via Eqs.(14).

The canonical transformation (13) can be written in a concise form if we introduce the Poisson bracket operator $L_\chi \equiv \{\cdot,\chi\}$, whose action on functions $f(\psi,\phi,p,I)$ is defined by:

$$L_\chi f = \{f,\chi\} = \frac{\partial f}{\partial \psi}\frac{\partial \chi}{\partial p} + \frac{\partial f}{\partial \phi}\frac{\partial \chi}{\partial I} - \frac{\partial f}{\partial p}\frac{\partial \chi}{\partial \psi} - \frac{\partial f}{\partial I}\frac{\partial \chi}{\partial \phi} \quad . \qquad (15)$$

The time derivative of any function $f(\psi,\phi,p,I)$ along a hamiltonian flow defined by the function χ is given by:

$$\frac{df}{dt} = \frac{\partial f}{\partial \psi}\dot{\psi} + \frac{\partial f}{\partial \phi}\dot{\phi} + \frac{\partial f}{\partial p}\dot{p} + \frac{\partial f}{\partial I}\dot{I} = \frac{\partial f}{\partial \psi}\frac{\partial \chi}{\partial p} + \frac{\partial f}{\partial \phi}\frac{\partial \chi}{\partial I} - \frac{\partial f}{\partial p}\frac{\partial \chi}{\partial \psi} - \frac{\partial f}{\partial I}\frac{\partial \chi}{\partial \phi}$$

that is

$$\frac{df}{dt} = \{f, \chi\} = L_\chi f \quad . \tag{16}$$

Extending this to higher order derivatives, we have

$$\frac{d^n f}{dt^n} = \{\ldots\{\{f, \chi\}, \chi\} \cdots \chi\} = L_\chi^n f \quad . \tag{17}$$

The Taylor series (13) can now be written as

$$\psi_t = \psi_0 + \frac{d\psi_0}{dt}t + \frac{d^2\psi_0}{dt^2}t^2 + \ldots = \sum_{n=0}^{\infty} \frac{1}{n!}\frac{d^n\psi_0}{dt^n}t^n \quad . \tag{18}$$

However, taking into account that the Taylor expansion of the exponential around the origin is given by

$$\exp(x) = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \ldots = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

we can see that the Taylor expansion (18) is formally given by the following exponential operator

$$\exp\frac{d}{dt} = 1 + \frac{d}{dt} + \frac{1}{2}\frac{d^2}{dt^2} + \ldots$$

Finally, taking into account Eqs.(16) and (17), we are lead to the formal definition of the *Lie series*

$$\psi_t = \psi_0 + (L_\chi\psi_0)t + \frac{1}{2}(L_\chi^2\psi_0)t^2 + \ldots \tag{19}$$

Setting, again, in (19) the time as $t = 1$, we define a canonical transformation using Lie series by recasting equations (14) in the form:

$$\psi_1 = \exp(L_\chi)\psi_0, \quad \phi_1 = \exp(L_\chi)\phi_0, \quad p_1 = \exp(L_\chi)p_0, \quad I_1 = \exp(L_\chi)I_0 \quad . \tag{20}$$

We summarize here the following basic properties of Lie series:

i) *Any* arbitrary choice of function χ produces a canonical transformation $(\psi_0, \phi_0, p_0, I_0) \rightarrow (\psi_1, \phi_1, p_1, I_1)$ via Eqs.(20). In normal form theory, however, we will see how to make a proper choice of one or more Lie generating functions χ so as to ensure that, after accomplishing a sequence of canonical transformations, the Hamiltonian in the new variables obtains a form representing, indeed, the kind of properties we want it to represent.

ii) The operations involved in Eqs.(20) are just calculations of derivatives. This renders the method quite easy to implement and to adapt to a computer-algebraic program.

iii) The fact that the time derivative of any function f under the Hamiltonian flow of another function χ is given by Eq.(16) implies that:

$$f_1(\psi_1, \phi_1, p_1, I_1) = \tag{21}$$
$$f(\psi(\psi_1, \phi_1, p_1, I_1), \phi(\psi_1, \phi_1, p_1, I_1), p(\psi_1, \phi_1, p_1, I_1), I(\psi_1, \phi_1, p_1, I_1))$$
$$= \exp(L_\chi)f(\psi_1, \phi_1, p_1, I_1) \quad .$$

In other words, in order to find what form does a function f of the canonical variables (ψ, ϕ, p, I) take, if, in the place of (ψ, ϕ, p, I) we substitute the new variables $(\psi_1, \phi_1, p_1, I_1)$, we do not really need to compute the transformation and substitute afterwards in f, we can simply implement the Lie operator $\exp(L_\chi)$ directly on f. This is algebraically very convenient, because it means we never have to make function compositions, which is a cumbersome algebraic procedure, but we only have to deal with computing derivatives of functions, which is always a straightforward procedure.

iv) If χ is a *small* quantity, the Lie series transformation (20) can be considered as a *near-identity* transformation, since, for example,

$$\exp(L_\chi)\psi = \psi + \{\psi, \chi\} + \dots$$

All the above properties are relevant in the implementation of normal form theory as will be clear by the examples below.

2.4.   Application: Birkhoff normal form for rotational tori

We are now ready to see how normal form theory can be implemented in the Hamiltonian (3), in order to help answering some practical questions. The first question that we will address is the following: in the phase portrait of Fig.1a, in the domain beyond the separatrix, we observe the existence of many invariant curves corresponding to rotational tori (roughly at values of p in the domain $|p| > 0.5$). The motion on such tori appears to be *quasi-periodic* [4]. Can we find a useful normal form by which to represent the motions in this domain as quasi-periodic? Even more, can we use normal form theory in order to prove the existence of quasi-periodic motions in the system (3)?

It is always important to recall the following fact: despite that normal form theory is a mathematical theory, our chance to implement it successfully in order to answer such questions depends crucially on our correct understanding of the *physical* properties of the motions we try to represent. In the domain $|p| > 0.5$ of Fig.1a, these motions are rotations, i.e. a test particle (or a real pendulum) subscribes a rotation with a value of the angle ψ monotonically increasing in time (above the separatrix), or decreasing in time (below the separatrix). In fact, it is evident that for p large in measure, the rotational motion of the pendulum approaches more and more uniform rotation. In the phase portrait of Fig.1a, this means that for higher values of p the invariant curves on the top side of the plot take closer and closer the form of straight lines. Thus, it seems reasonable to look for a normal form representing the motions in this domain, constructed in such a way that its lowest order approximation represents uniform rotation.

On the basis of the above argument, let us focus on *one* value of p, to be hereafter denoted p_*, for example $p_* = 2(\sqrt{3} - 1) \simeq 1.46410$, corresponding to the value of the ordinate of a dashed line shown in the upper part of Fig.1a. The reason for choosing a value of this type (which is incommensurable with $\omega = 1$) will become evident in a while. We now clearly see that the dashed line

---

[4]A dynamical variable $x(t)$ is said to undergo a quasi-periodic time evolution with n incommensurable frequencies $\omega_1, ..., \omega_n$ if its dependence on time is via one or more trigonometric terms of the form $\genfrac{}{}{0pt}{}{\cos}{\sin}((m_1\omega_1 + m_2\omega_2 + \dots + m_n\omega_n)t)$, with $m_1, m_2, ..., m_n$ integer.

in Fig.1a intersects some invariant curves in such a way, that the variations of p along one curve above or below the dashed line are relatively small (of order ~ 0.1) around the fixed value $p_* = 1.46410$. Let us then introduce a local action variable I_ψ via

$$I_\psi = p - p_* \tag{22}$$

i.e. by simply shifting the center at $p = p_*$. A change of variables like (22) is a trivial form of canonical transformation, thus it can be substituted in the Hamiltonian function (3). Setting $p = 1.46410 + I_\psi$, and also $\omega_0 = 0.2\sqrt{2}$, $\omega = 1$, $\epsilon = 0.04$, we arrive at the Hamiltonian [5]

$$
\begin{aligned}
H &= 1.07179 + 1.46410 I_\psi + I + \frac{I_\psi^2}{2} \\
&\quad - 0.08 \cos\psi - (0.0078851 + 0.0032 I_\psi) \cos\psi \cos\phi \ .
\end{aligned}
\tag{23}
$$

The constant 1.07179 in the hamiltonian (23) does not affect the dynamics, thus we will omit it in all subsequent steps. It is interesting to note that, already at this stage, the Hamiltonian (23) provides us with some information about the character of the rotational motion. In fact, if we take Hamilton's equation for $\dot\psi$ we find $\dot\psi = \partial H/\partial I_\psi = 1.46410 + ...$, where the three dots denote here all remaining terms, whose size, however, is considerably smaller than the size of the first term. We thus see that the Hamiltonian (23) describes rotational motions with an angular frequency that varies in time a little around the value $\omega_* = 1.46410$.

In order to make this statement more precise, we need to characterize quantitatively how does the presence of the terms $-I_\psi^2/2$, $-0.08 \cos\psi$ and $-(0.0078851 +0.0032 I_\psi) \cos\psi \cos\phi$ perturb the motion with respect to a uniform rotation. A quick visual look to Fig.1a shows that the variable I_ψ is expected to hold variations of about ~ 0.2, or $I_\psi^2/2 \sim 0.02$. Thus, the two terms $I_\psi^2/2$ and $0.08 \cos\psi$ have a rather comparable size, of the order of a few times 10^{-2}, while the third term can be considered of the same order, or one order of magnitude smaller. For simplicity, we will consider all three terms of similar size [6]. We will introduce a formal notation to account for this consideration: in front of every term in (23), we introduce a factor λ^s, where λ, called hereafter the 'book-keeping parameter', is a constant with numerical value equal to $\lambda = 1$, while s is a positive integer exponent whose value, for every term in (23), is selected so as to

---

[5]The reader is asked to tolerate the fact that we will continue the example hereafter with numerical coefficients. In fact, this is exactly what happens in practice, i.e., in practical computations we rarely choose to carry along symbols like $\omega_0, \omega, \epsilon$ etc. With a little effort to acquaint him/herself with this notation, the reader will realize that the use of numerical coefficients hereafter renders the example easier to study. Furthermore, there seems to be no better way for one to acquire a 'feeling' of how a method works, than by seeing the real numbers the method produces in a practical example. At any rate, we will truncate all figures at five digits, so as not to produce very lengthy formulae. In the computer, we retain of course many more significant digits.

[6]This is really a convention; with a little a posteriori experience, one understands that a difference of one order of magnitude does not really matter very much in whether or not we should diversify the third term from the first two terms.

reflect our consideration regarding what order of smallness we estimate a term to be of in the Hamiltonian. Thus, considering the leading terms $1.46410I_\psi$ and I as of order zero, we put λ^0 in front of them. On the other hand, considering the remaining three terms as of a similar order of smallness, i.e. first order, we put a factor λ^1 in front of them. The Hamiltonian now reads:

$$
\begin{aligned}
H \;=\;& \lambda^0 (1.46410 I_\psi + I) \\
+\;& \lambda^1 \left(\frac{I_\psi^2}{2} - 0.08\cos\psi - (0.0078851 + 0.0032 I_\psi)\cos\psi\cos\phi \right) \;\;.
\end{aligned}
\tag{24}
$$

Since $\lambda = 1$, nothing has really changed. However, the appearance of the symbol λ allows one to clearly identify one's own perceptions about the 'hierarchy' by which various terms in the Hamiltonian affect the dynamics. These perceptions are not completely objective. Let us discuss some alternative choices with respect to the book-keeping introduced in (24). We could have written:

$$
\begin{aligned}
H \;=\;& \lambda^0 \left(1.46410 I_\psi + I + \frac{I_\psi^2}{2} \right) \\
-\;& \lambda^1 \left(0.08\cos\psi + (0.0078851 + 0.0032 I_\psi)\cos\psi\cos\phi \right) \;\;.
\end{aligned}
\tag{25}
$$

Such a choice would imply that we consider the free rotator model as the basic approximation, i.e. a model in which the frequency of rotation varies with the action. In fact, this is perfectly true for the pendulum. Thus, one may incorporate this property directly from the start in the 'zeroth-order Hamiltonian', a process which, as we will see, is indispensable in the construction of the so-called Kolmogorov normal form, by which we rigorously show the existence of rotational tori in the Hamiltonian (3). If, on the other hand, we ignore this property at zeroth order, and consider it a higher order effect (like in the book-keeping of Eq.(24)), we will 'recover' the dependence of the frequency on the action I_ψ after working out some steps of perturbation theory. Let us mention still a third option:

$$
\begin{aligned}
H \;=\;& \lambda^0 \left(1.46410 I_\psi + I + \frac{I_\psi^2}{2} - 0.08\cos\psi \right) \\
-\;& \lambda^1 (0.0078851 + 0.0032 I_\psi)\cos\psi\cos\phi \;\;.
\end{aligned}
\tag{26}
$$

Choosing this book-keeping option is equivalent to saying that, at zero order, we consider the dynamics being that of the pendulum, so that the only perturbing term is provided by the external driving.

It must be emphasized that all these choices are nothing but different representations of what we consider to be the essential dynamics, although, in reality, the true dynamics is one and the same. Thus, in a certain sense a choice of book-keeping is only a representation of *our own* viewpoint of what is a useful basic approximation to the dynamics. This 'freedom' notwithstanding, it should be stressed that having a correct viewpoint guided by physical principles is essential, because the choice of book-keeping affects in a crucial way all subsequent

steps of perturbation theory, thus determining how successful will the approximation prove to be in the end. As we will see, this reflects both in formal, as well as physical properties of the solutions found via any particular perturbative scheme. [7]

We now return to our original book-keeping scheme, i.e. Eq.(24). In this scheme, the zeroth order Hamiltonian is just $H_0 = 1.46410 I_\psi + I$. Thus, under the Hamiltonian flow of H_0, both quantities I_ψ, I are integrals of motion, i.e. one has $\dot{I}_\psi = \dot{I} = 0$, since in the Hamiltonian H_0 both angles ψ, ϕ are ignorable. This property of H_0 suggests looking for the possibility to construct, via a sequence of canonical transformations, a normal form which, in the new variables, is independent of the angles. This guarantees that the new action variables, after the transformation, will be integrals of motion of the normal form canonical flow.

Let us see in detail how to compute the normal form at first order in the book-keeping parameter λ. We implement the following steps[8]:

i) *notation:* We denote by $H^{(0)}$ the original hamiltonian and by $H^{(1)}$ the Hamiltonian after implementing a canonical transformation from the original variables, denoted hereafter by $(\psi^{(0)}, \phi^{(0)}, I_\psi^{(0)}, I^{(0)})$, to new canonical variables, denoted by $(\psi^{(1)}, \phi^{(1)}, I_\psi^{(1)}, I^{(1)})$. Furthermore, we denote by $Z_0 = H_0^{(0)}$ the terms of $H^{(0)}$ of order zero in λ, and by $H_1^{(0)}$ the terms of order one in λ, that is

$$H^{(0)} = Z_0 + \lambda H_1^{(0)} \tag{27}$$

where

$$Z_0 = 1.46410 I_\psi^{(0)} + I^{(0)},$$

$$H_1^{(0)} = \frac{(I_\psi^{(0)})^2}{2} - 0.08 \cos \psi^{(0)} - (0.0078851 + 0.0032 I_\psi^{(0)}) \cos \psi^{(0)} \cos \phi^{(0)} \quad .$$

ii) We look for a Lie generating function bringing the Hamiltonian in normal form up to terms of first order in λ. The generating function, denoted by χ_1, is a function of the *new* canonical variables, i.e. $\chi_1 \equiv \chi_1(\psi^{(1)}, \phi^{(1)}, I_\psi^{(1)}, I^{(1)})$.

---

[7] We will see below that the terms of order zero in λ appear in the so-called 'homological equation', i.e. a partial differential equation by which we compute generating functions in canonical perturbation theory. These terms are also called the 'kernel' of the homological equation, or the 'Hori kernel' (Hori (1966), Deprit (1969), see Ferraz-Mello (2007)). In summary, the choice of Hori kernel implies a choice of basic dynamical model whose qualitative behavior we expect to approximate the true dynamics in the domain of the phase space considered. Four basic choices of possible Hori kernel are: i) the oscillator (e.g., in two degrees of freedom, $\omega I + \omega_* I_\psi$, ii) the rotator ($\omega I + \omega_* I_1 + \beta I_\psi^2/2$), iii) the pendulum (or 'first fundamental resonance model' (Henrard and Lemaitre (1983)) $\omega I + \omega_* I_\psi + \beta I_\psi^2/2 + B \cos(k\psi)$), and iv) the Andoyer Hamiltonian (or 'second fundamental resonance model' (Henrard and Lemaitre 1983), $\omega I + \omega_* I_\psi + \beta I_\psi^2/2 + B I_\psi^{1/2} \cos(k\psi)$). The latter kernel appears quite often in problems or resonance in solar system and in galactic dynamics (see e.g. subsection 5.4).

[8] The reader is prompted not to be discouraged by the apparent complexity of subsequent formulae. But this is really the crucial point of the whole method, so study with paper, pencil, and a lot of patience, is at this point indispensable.

iii) The canonical transformation induced by χ_1 yields the *old* canonical variables *in terms of the new* canonical variables (not the other way around). This is explicitly given by:

$$\psi^{(0)} = \exp(L_{\chi_1})\psi^{(1)}, \quad \phi^{(0)} = \exp(L_{\chi_1})\phi^{(1)}, \tag{28}$$
$$I_\psi^{(0)} = \exp(L_{\chi_1})I_\psi^{(1)}, \quad I^{(0)} = \exp(L_{\chi_1})I^{(1)} \ .$$

We now specify what is an appropriate form χ_1 should have, to render $H^{(1)}$ in normal form up to terms $O(\lambda)$. To accomplish this task, we note that if we knew χ_1, then, according to Eq.(21), the Hamiltonian after the transformation (28) would be given by:

$$H^{(1)}(\psi^{(1)}, \phi^{(1)}, I_\psi^{(1)}, I^{(1)}) = \exp(L_{\chi_1})H^{(0)}(\psi^{(1)}, \phi^{(1)}, I_\psi^{(1)}, I^{(1)}) \ . \tag{29}$$

The lowest order terms of the Lie operation in the r.h.s. of Eq.(29) are:

$$H^{(1)} = \exp(L_{\chi_1})H^{(0)} = H^{(0)} + L_{\chi_1}H^{(0)} + \frac{1}{2}L_{\chi_1}^2 H^{(0)} + ...$$

In view of (27) this takes the form

$$\begin{aligned} H^{(1)} \ &= \ Z_0 + \lambda H_1^{(0)} + \{Z_0, \chi_1\} + \lambda\{H_1^{(0)}, \chi_1\} \\ &+ \ \frac{1}{2}\{\{Z_0, \chi_1\}, \chi_1\} + \lambda\frac{1}{2}\{\{H_1^{(0)}, \chi_1\}, \chi_1\} + ... \end{aligned} \tag{30}$$

The key remark, now, is the following: in the Hamiltonian (27), the 'unwanted' terms (containing angles) are all found in the $H_1^{(0)}$ term. We denote by $h_1^{(0)}$ these terms, and in order to simplify notations, we will drop superscripts from the notation of the canonical variables (ψ, ϕ, I_ψ, I), keeping always in mind that before and after the canonical transformation there is an omitted superscript (0) and (1) respectively in these variables. With these conventions, we have:

$$h_1^{(0)} = -0.08\cos\psi - (0.0078851 + 0.0032 I_\psi)\cos\psi\cos\phi \ .$$

We then note the following: if we choose χ_1 to be a quantity of first order $(O(\lambda))$ in the book-keeping parameter, then, *only the second and third terms in Eq.(30) are of first order in λ*. Indeed, Z_0 is of order zero, while (assuming $\chi_1 = O(\lambda)$)

$$\{Z_0, \chi_1\} + \lambda H_1^{(0)} = O(\lambda),$$

$$\lambda\{H_1^{(0)}, \chi_1\} = O(\lambda^2),$$

$$\frac{1}{2}\{\{Z_0, \chi_1\}, \chi_1\} = O(\lambda^2),$$

$$\lambda\frac{1}{2}\{\{H_1^{(0)}, \chi_1\}, \chi_1\} = O(\lambda^3)$$

etc. We then require that χ_1 be chosen in such a way that the combination of the only two possible terms of order $O(\lambda)$ eliminates the unwanted terms $h_1^{(0)}$ from the new Hamiltonian $H^{(1)}$. That is, we require that χ_1 be such that

$$\lambda(\text{unwanted terms in } H_1^{(0)}) + \{Z_0, \chi_1\} = 0$$

or

$$\{Z_0, \chi_1\} + \lambda h_1^{(0)} = 0 \quad . \tag{31}$$

An equation of the form (31) is called *homological equation*. It is the most basic equation of canonical perturbation theory, since it is the one by which we specify the various generating functions appearing in a theory, like χ_1. Also, we will see that the solution of a homological equation introduces *divisors*, whose accumulation after subsequent steps is responsible for the convergence (or non-convergence) properties of the series under study.

The solution of Eq.(31) is found in a straightforward manner, if we use the exponential notation $\cos\psi = (e^{i\psi} + e^{-i\psi})/2$ (and similarly for $\cos\phi$), in terms of which we have:

$$h_1^{(0)} = -0.04(e^{i\psi} + e^{-i\psi})$$

$$-(0.0019713 + 0.0008 I_\psi)(e^{i(\phi+\psi)} + e^{i(\phi-\psi)} + e^{i(-\phi+\psi)} + e^{i(-\phi-\psi)}) \quad .$$

The solution of (31) is then found by noting that a trigonometric term of the form $a(I_\psi, I)e^{i(k_1\psi+k_2\phi)}$, when acted upon by the operator $\{Z_0, \cdot\} = \{\omega_* I_\psi + \omega I, \cdot\}$ (where, in our case, $\omega_* = 1.46410$, $\omega = 1$), yields

$$\begin{aligned}\{Z_0, a(I_\psi, I)e^{i(k_1\psi+k_2\phi)}\} &= \{\omega_* I_\psi + \omega I, a(I_\psi, I)e^{i(k_1\psi+k_2\phi)}\} \\ &= -i(k_1\omega_* + k_2\omega)a(I_\psi, I)e^{i(k_1\psi+k_2\phi)} \quad .\end{aligned} \tag{32}$$

We then readily verify that if $h_1^{(0)}$ is written under the form of a sum of Fourier terms

$$h_1^{(0)} = \sum_{k_1,k_2,|k_1|+|k_2|\neq 0} b_{k_1,k_2}(I_\psi, I)e^{i(k_1\psi+k_2\phi)}$$

the homological equation (31) is satisfied by setting χ_1 equal to:

$$\chi_1 = \lambda \sum_{k_1,k_2,|k_1|+|k_2|\neq 0} \frac{b_{k_1,k_2}(I_\psi, I)}{i(k_1\omega_* + k_2\omega)} e^{i(k_1\psi+k_2\phi)} \quad . \tag{33}$$

In the specific example of $h_1^{(0)}$ given as above, we find:

$$\chi_1 = \lambda i \left(0.027320(e^{i\psi} - e^{-i\psi}) + (0.0008 + 0.00032466 I_\psi)(e^{i(\psi+\phi)} - e^{-i(\psi+\phi)}) \right.$$

$$\left. + (0.0042475 + 0.0017238 I_\psi)(e^{i(\psi-\phi)} - e^{-i(\psi-\phi)}) \right) \quad .$$

We are essentially done. We now only need to find the form of the Hamiltonian in the new canonical variables, by computing

$$H^{(1)} = \exp(L_{\chi_1}) H^{(0)} \quad . \tag{34}$$

Up to second order in λ we find[9]:

$$H^{(1)} = Z_0 + \lambda Z_1 + \lambda^2 H_2^{(1)} + ...$$

where

$$Z_1 = \frac{I_\psi^2}{2}$$

$$H_2^{(1)} = -8.07603 \times 10^{-6} - 3.27748 \times 10^{-6} I_\psi - 4.43496 \times 10^{-6} (e^{i(2\psi+\phi)} + e^{-i(2\psi+\phi)})$$

$$+ 2.35470 \times 10^{-5} (e^{i(2\psi-\phi)} + e^{-i(2\psi-\phi)}) - 6.28249 \times 10^{-5} (e^{i\phi} + e^{-i\phi})$$

$$+ 0.027320 I_\psi (e^{i\psi} + e^{-i\psi})$$

$$+ (8 I_\psi + 3.24662 I_\psi^2) \times 10^{-4} (e^{i(\psi+\phi)} + e^{-i(\psi+\phi)})$$

$$+ (0.0042475 I_\psi + 0.0017238 I_\psi^2)(e^{i(\psi-\phi)} + e^{-i(\psi-\phi)})$$

$$- (4.03802 + 1.63874 I_\psi) \times 10^{-6} (e^{2i\phi} + e^{-2i\phi}) \quad .$$

This resumes one complete step of the normalization algorithm.

The following are some remarks regarding the form of the Hamiltonian after the first normalization step:

i) Up to first order, the Hamiltonian $H^{(1)}$ has been 'brought into normal form', i.e. we see the appearance of terms depending only on the actions I_ψ and I.

ii) At second and subsequent orders, on the other hand, the Hamiltonian $H^{(1)}$ contains new terms, depending on the angles. These terms were not present in the initial Hamiltonian, but they were produced by the Lie operation of Eq.(34). In fact, the quantity $R^{(1)} = \lambda^2 H_2^{(1)} + \lambda^3 H_3^{(1)} + ...$ constitutes the *remainder* function at this order of normalization. The appearance of new, 'unwanted', terms, at second and higher orders, implies that, in order to eliminate these terms, we need further canonical transformations. Thus, using a Lie generating function χ_2 (of order $O(\lambda^2)$), we eliminate the unwanted terms of order $O(\lambda^2)$ in the Hamiltonian. This, in turn, generates new unwanted terms to be eliminated at subsequent steps, etc.

iii) Related to the previous remark, we also observe that the action of multiple Poisson brackets in the implementation of the Lie operation of Eq.(34) generates new trigonometric terms, some of which are of *higher Fourier order* than those of the original Hamiltonian. Such are the terms $\exp(i(2\psi \pm \phi))$. This last property is important, because, as shown below, the generation of terms of

---

[9]Again, possibility to reproduce these results by paper and pencil implies that the method has been understood.

higher and higher Fourier order in the course of normalization implies also the appearance of *new divisors* (of higher order) in the series.[10]

We finally recall again that, despite the absence of superscripts on the canonical variables (ψ, ϕ, I_ψ, I), all variables appearing in the above expressions for $H^{(1)}$ are the new variables, i.e. the one following the Lie series canonical transformation with χ_1.

2.5. General normalization algorithm

So far we demonstrated in detail the computation of a Lie canonical transformation and associated normal form at first order of perturbation theory. We can readily generalize this computation and formulate a recurrent algorithm for Hamiltonian normalization at an arbitrary order r. To this end, we assume that r normalization steps were accomplished, and give the formulae for the $r + 1$ step. After r steps, the Hamiltonian has the form:

$$H^{(r)} = Z_0 + \lambda Z_1 + ... + \lambda^r Z_r + \lambda^{r+1} H_{r+1}^{(r)} + \lambda^{r+2} H_{r+2}^{(r)} + ... \qquad (35)$$

The Hamiltonian term $H_{r+1}^{(r)}$ contains some terms that we want to eliminate, denoted by $h_{r+1}^{(r)}$, as well as some terms that we do not want to eliminate, denoted by Z_{r+1} (in the present example, these are the terms containing only the action variables). The question is to specify a generating function, of order $O(\lambda^{r+1})$, which accomplishes the normalizing transformation. As in the case $r = 0$, we observe that in the Lie series $\exp(L_{\chi_{r+1}})H^{(r)}$, the only terms of order λ^{r+1} are $\lambda^{r+1} H_{r+1}^{(r)}$ and $L_{\chi_{r+1}} Z_0$. This implies that the generating function χ_{r+1} can be specified by solving the homological equation:

$$\{Z_0, \chi_{r+1}\} + \lambda^{r+1} h_{r+1}^{(r)} = 0 \quad . \qquad (36)$$

This equation can be solved in exactly the same manner as Eq.(33). After the generating function χ_{r+1} has been specified, we can compute the new, transformed Hamiltonian

$$H^{(r+1)} = \exp(L_{\chi_{r+1}})H^{(r)} \quad . \qquad (37)$$

By construction, this is in normal form up to terms of order $r + 1$, namely:

$$H^{(r+1)} = Z_0 + \lambda Z_1 + ... + \lambda^r Z_r + \lambda^{r+1} Z_{r+1} + \lambda^{r+2} H_{r+2}^{(r+1)} + ... \qquad (38)$$

This completes the $r + 1$ step of the normalization algorithm. The entire recursive algorithm thus takes the following form:

---

[10]On the other hand, in section 3 we will consider cases where Fourier terms of all orders are present in the series from the start. In such cases, however, the analyticity properties of the original Hamiltonian determine the size of its various Fourier terms as a function of the Fourier order in the original Hamiltonian. Then, it turns out that the process of generation of new terms provides again the leading contribution to the growth of the size of the series terms at subsequent normalization steps (see Morbidelli and Giorgilli (1997), and Morbidelli (2002), for a detailed discussion of this process). Thus, despite some formal differences, the outcomes of the normalization of Hamiltonians with either a finite, or a non-finite number of Fourier terms, are not so different in the end. More on this topic is discussed in section 3.

Recursive normalization algorithm with Lie series

Assuming r algorithm steps have been completed:

1) Isolate from the Hamiltonian $H^{(r)}$ (Eq.(35)) the terms in $H_{r+1}^{(r)}$ to be eliminated at the present step, i.e. $h_{r+1}^{(r)}$.

2) Solve the homological equation (36) and define χ_{r+1}.

3) Implement Eq.(37) and compute $H^{(r+1)}$. This contains a normal form part:

$$Z^{(r+1)} = Z_0 + \lambda Z_1 + ... + \lambda^r Z_r + \lambda^{r+1} Z_{r+1}$$

and a remainder part

$$R^{(r+1)} = \lambda^{r+2} H_{r+2}^{(r+1)} + \lambda^{r+3} H_{r+3}^{(r+1)} + \dots$$

The reader is invited to accomplish, for practice, the second normalization step in our example treated in subsection 2.4. We give, for verification, the form of the generating function χ_2:

$$\chi_2 = \lambda^2 i \bigg(1.12901 \times 10^{-6} (e^{i(2\psi+\phi)} - e^{-i(2\psi+\phi)})$$

$$-1.22119 \times 10^{-5} (e^{i(2\psi-\phi)} - e^{-i(2\psi-\phi)}) + 6.28249 \times 10^{-5} (e^{i\phi} - e^{-i\phi})$$

$$-0.018660 I_\psi (e^{i\psi} - e^{-i\psi})$$

$$-(3.24662 I_\psi + 1.31757 I_\psi^2) \times 10^{-4} (e^{i(\psi+\phi)} - e^{-i(\psi+\phi)})$$

$$-(0.00915214 I_\psi + 0.00371419 I_\psi^2)(e^{i(\psi-\phi)} - e^{-i(\psi-\phi)})$$

$$+(2.01901 + 0.81937 I_\psi) \times 10^{-6} (e^{2i\phi} - e^{-2i\phi}) \bigg) \ .$$

The new Hamiltonian, after the second normalization, is given by:

$$H^{(2)} = \exp(L_{\chi_2}) H^{(1)} \ . \tag{39}$$

Restoring the numerical value of λ, i.e. $\lambda = 1$, the Hamiltonian $H^{(2)}$ is in normal form up to terms of order 2, namely (omitting a constant):

$$H^{(2)} = (1.4641016 - 3.27748 \times 10^{-6}) I_\psi + I + \frac{I_\psi^2}{2} + O(\lambda^3) \ .$$

We observe that the normalization procedure generated a small frequency correction even to the term linear in the action I_ψ.

On the other hand, the remainder is

$$R^{(2)} = \lambda^3 H_3^{(2)} + \lambda^4 H_4^{(2)} + \dots \tag{40}$$

The leading term in the remainder is $H_3^{(2)}$. This contains 75 terms, thus it is unpractical to reproduce here. In fact, beyond the second order, calculations

are all together hard to do without use of a computer-algebraic program. However, we can predict some features of the term $\lambda^3 H_3^{(2)}$. In the series (39), the terms contributing to $H_3^{(2)}$ are just $H_3^{(1)}$ and $\{Z_1, \chi_2\}$. The term $H_3^{(1)}$ can be analyzed itself in terms of Eq.(34), and it is found to contain contributions from $(1/2)\{\{H_1^{(0)}, \chi_1\}, \chi_1\}$, and $(1/6)\{\{\{Z_0, \chi_1\}, \chi_1\}, \chi_1\}$. We thus start seeing, now, that the overall effect of the normalization procedure is to 'propagate' the initial Hamiltonian terms at higher and higher orders via multiple Poisson brackets with one or more of the generating functions χ_1, χ_2, etc. As we have seen already in subsection 2.4, a main effect of this process is the generation of new harmonics, as the normalization proceeds. Namely, taking the Poisson bracket between any two terms of the form $ae^{\pm i(k_1\psi + k_2\phi)}$ and $be^{\pm i(m_1\psi + m_2\phi)}$, with at least one of the coefficients a, or b, depending on the action I_ψ, we find new terms according to

$$\{ae^{\pm i(k_1\psi + k_2\phi)}, be^{\pm i(m_1\psi + m_2\phi)}\} \to \text{New Fourier terms} \quad e^{\pm i((k_1 \pm m_1)\psi + (k_2 \pm m_2)\phi)} \quad .$$

Since two out of the four possible (plus or minus) combinations of the expressions $|k_1 \pm m_1| + |k_2 \pm m_2|$ are larger from both $|k_1| + |k_2|$, and $|m_1| + |m_2|$, we see that the result of the Poisson brackets acting via the Lie operator is to generate new Fourier harmonics, of higher and higher order, at consecutive normalization steps. Thus, for example, we can check that the term $H_3^{(2)}$ contains harmonics beyond the order 3, namely the harmonics:

$$H_3^{(2)} \to e^{i(2\psi \pm 2\phi)}, \; e^{i(\psi \pm 3\phi)}, \; e^{i(3\psi \pm 2\phi)} \quad .$$

In subsection 2.7 we analyze how the appearance of harmonics of increasing order affects the *convergence properties* of the whole normalization process presently examined. We will see that these harmonics result in the appearance of new divisors, which, in turn, affect the growth rate of the series terms at successive normalization steps.

2.6. Practical benefits from the normal form computation

The practical question now is: what is our benefit from computing a normal form as above? In particular, can we advance our understanding of the dynamics by a normal form computation in a system like (3) with respect to a purely numerical investigation of the orbits?

Two main ways to benefit from computing a normal form are related to explicitly computing the *normalizing transformation* by which we pass from the old to the new canonical variables, and vice versa. Consider first the transformation yielding the old variables in terms of the new variables. After r normalization steps, this is given by a *composition of Lie series*:

$$
\begin{aligned}
\psi &= \exp(L_{\chi_r})\exp(L_{\chi_{r-1}})...\exp(L_{\chi_1})\psi^{(r)} \\
\phi &= \exp(L_{\chi_r})\exp(L_{\chi_{r-1}})...\exp(L_{\chi_1})\phi^{(r)} \\
I_\psi &= \exp(L_{\chi_r})\exp(L_{\chi_{r-1}})...\exp(L_{\chi_1})I_\psi^{(r)} \\
I &= \exp(L_{\chi_r})\exp(L_{\chi_{r-1}})...\exp(L_{\chi_1})I_\psi^{(r)} \quad .
\end{aligned}
\tag{41}
$$

The net result of Eqs.(41) is to find series expressions of the form

$$
\begin{aligned}
\psi &= F_\psi(\psi^{(r)}, \phi^{(r)}, I_\psi^{(r)}, I^{(r)}) \\
\phi &= F_\phi(\psi^{(r)}, \phi^{(r)}, I_\psi^{(r)}, I^{(r)}) \\
I_\psi &= F_{I_\psi}(\psi^{(r)}, \phi^{(r)}, I_\psi, I^{(r)}) \\
I &= F_I(\psi^{(r)}, \phi^{(r)}, I_\psi^{(r)}, I^{(r)})
\end{aligned}
\tag{42}
$$

in which the old canonical functions are expressed in terms of the new canonical variables[11]. However, in our example of subsections 2.4 and 2.5, after a normal form has been computed, the time evolution of the new canonical variables under the normal form dynamics alone (i.e. ignoring the effect of the remainder) can itself be easily computed. In fact, since our normal form depends only on the actions $I_\psi^{(r)}, I^{(r)}$, both $I_\psi^{(r)}$ and $I^{(r)}$ are *integrals* of the normal form dynamics. This also determines the frequencies by which the angles $\psi^{(r)}$ and $\phi^{(r)}$ evolve. In summary, we have the following time evolution of all new canonical variables:

$$
\begin{aligned}
I_\psi^{(r)}(t) &= I_{\psi,0}^{(r)} \\
I^{(r)}(t) &= I_0^{(r)} \\
\psi^{(r)}(t) &= \left(\psi_0^{(r)} + \left. \frac{\partial Z^{(r)}}{\partial I_\psi^{(r)}} \right|_{I_{\psi,0}^{(r)}, I_0^{(r)}} \right) t \\
\phi^{(r)}(t) &= \left(\phi_0^{(r)} + \left. \frac{\partial Z^{(r)}}{\partial I^{(r)}} \right|_{I_{\psi,0}^{(r)}, I_0^{(r)}} \right) t
\end{aligned}
\tag{43}
$$

where the constants $(\psi_0^{(r)}, \phi_0^{(r)}, I_{\psi,0}^{(r)}, I_0^{(r)})$ mark the initial conditions of an orbit computed in the new variables. Since in a numerical calculation we can usually know the initial conditions only in the old variables $\psi_0, \phi_0, I_{\psi,0}, I_0$, we need also the inverse canonical transformation (from old to new variables) to compute the constants $(\psi_0^{(r)}, \phi_0^{(r)}, I_{\psi,0}^{(r)}, I_0^{(r)})$. It is easy to verify that the latter is given by:

$$
\begin{aligned}
\psi^{(r)} &= \exp(-L_{\chi_1}) \exp(-L_{\chi_2}) ... \exp(-L_{\chi_r}) \psi \\
\phi^{(r)} &= \exp(-L_{\chi_1}) \exp(-L_{\chi_2}) ... \exp(-L_{\chi_r}) \phi \\
I_\psi^{(r)} &= \exp(-L_{\chi_1}) \exp(-L_{\chi_2}) ... \exp(-L_{\chi_r}) I_\psi \\
I^{(r)} &= \exp(-L_{\chi_1}) \exp(-L_{\chi_2}) ... \exp(-L_{\chi_r}) I \ .
\end{aligned}
\tag{44}
$$

Substituting the values of the initial conditions $(\psi_0, \phi_0, I_{\psi,0}, I_0)$ in Eqs.(44) we then find the constants $(\psi_0^{(r)}, \phi_0^{(r)}, I_{\psi,0}^{(r)}, I_0^{(r)})$, and, thereby, the whole time evolution of an orbit in the new variables via Eqs.(43). But then, substituting the

---

[11]Of course, in the computer we can only store a finite truncation of such series.

expressions for $\psi^{(r)}(t)$, $\phi^{(r)}(t)$, $I_\psi^{(r)}(t)$, and $I^{(r)}(t)$ in the transformation (41), we can obtain an analytical formula for $\psi(t)$, $\phi(t)$, $I_\psi(t)$, and $I(t)$, i.e. for the time evolution of the old variables as well. That is, using the normal form we can obtain an analytical *quasi-periodic representation of the time evolution of all canonical variables for regular orbits.*

What is the precision level of such a representation? This is of the same order as the difference between the normal form dynamics and the true dynamics, implying that *the precision level of normal form calculations is of the order of the size of the remainder $R^{(r)}$.* In a realistic computation, aimed, for example to represent the orbit of a planet or an asteroid in the solar system, or the motion of a satellite around a planet, the size of the remainder can be used in the above sense in order to estimate the timescale up to which a normal form computation yields a useful prediction. This timescale is essentially given by the inverse of the size of the remainder function.

The second way in which Eqs.(41) are useful regards the possibility to parametrically represent the invariant surfaces on which lie the regular orbits, i.e., in our example, the rotational tori on which the orbits evolve quasi-periodically. In fact, the parametrization is provided, precisely, by Eqs.(41). This is trivial to see, since the values of $I_\psi^{(r)}$ and $I^{(r)}$, which are integrals of motion, can be replaced by the constants $(I_{\psi,0}^{(r)}, I_0^{(r)})$, which act as labels for invariant tori. After this replacement, all the old variables $(\psi_0, \phi_0, I_{\psi,0}, I_0)$ are given by equations depending on two parameters, namely $\psi^{(r)}$ and $\phi^{(r)}$. Thus, we have the definition of a surface topologically equivalent to a two-torus, which we can actually explicitly compute by taking many values of both angles $\psi^{(r)}$ and $\phi^{(r)}$ in the interval $[0, 2\pi)$.

Let us see the above properties in practice, by explicitly performing the associated calculations in our working example of subsections 2.4 and 2.5, up to a normalization order $r = 2$. Since χ_1 and χ_2 are given, it is straightforward to compute the transformations (41), expanding both exponentials up to order 2 in the book-keeping parameter λ. Recalling that $\chi_1 = O(\lambda)$ and $\chi_2 = O(\lambda^2)$, we have:

$$\exp(L_{\chi_2})\exp(L_{\chi_1}) = (1 + L_{\chi_2} + \frac{1}{2}L_{\chi_2}^2 + ...)(1 + L_{\chi_1} + \frac{1}{2}L_{\chi_1}^2 + ...)$$

$$= 1 + L_{\chi_1} + \frac{1}{2}L_{\chi_1}^2 + L_{\chi_2} + O(\lambda^3) \ .$$

Passing from exponentials back to trigonometric expressions, and setting $\lambda = 1$, we then find (using the expressions found in previous pages for χ_1 and χ_2):

$$\psi = \psi^{(2)} + 0.0373205\sin(\psi^{(2)}) + 1.11928 \times 10^{-6}\sin(2\psi^{(2)})$$

$$+0.01486\left(1 + I_\psi^{(2)}\right)\sin(\psi^{(2)} - \phi^{(2)}) - 1.63874 \times 10^{-6}\sin(2\phi^{(2)})$$

$$+5.27027 \times 10^{-4}I_\psi^{(2)}\sin(\psi^{(2)} + \phi^{(2)}) + 2.97135 \times 10^{-6}\sin(2\psi^{(2)} - 2\phi^{(2)})$$

$$+1.05405 \times 10^{-7}\sin(2\psi^{(2)} + 2\phi^{(2)}) + \ldots$$

$$I_\psi = 1.51629 \times 10^{-5} + 1.00001 I_\psi^{(2)} + \left(0.054641 - 0.0.037320 I_\psi^{(2)}\right) \cos(\psi^{(2)})$$

$$+1.11928 \times 10^{-4} \cos(\phi^{(2)}) + \left(5.51603 + 2.23856 I_\psi^{(2)}\right) \times 10^{-6} \cos(2\phi^{(2)})$$

$$+ \left(0.00849504 - 0.014857 I_\psi^{(2)} - 0.0074284 (I_\psi^{(2)})^2\right) \cos(\psi^{(2)} - \phi^{(2)})$$

$$-4.88476 \times 10^{-5} \cos(2\psi^{(2)} - \phi^{(2)})$$

$$+ \left(0.0016 - 2.63513 \times 10^{-4}(I_\psi^{(2)})^2\right) \cos(\psi^{(2)} + \phi^{(2)})$$

$$+4.51602 \times 10^{-6} \cos(2\psi^{(2)} + \phi^{(2)}) + \dots$$

We also have

$$\phi = \phi^{(2)} \quad,$$

i.e. the angle ϕ obeys the identity transformation. This is a particular feature of Hamiltonian systems like (3), i.e. where some angles (and their conjugate dummy actions) were artificially introduced to account for the time-dependent trigonometric terms. In fact, we observe that the dummy action I does not appear in any of the above expressions, or the expressions found for the Lie generating functions χ_r.
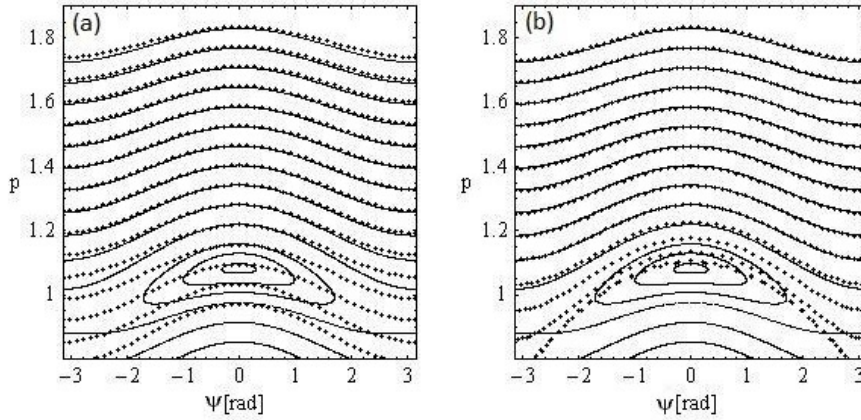
Figure 2. Numerical (solid) and theoretical (dotted) invariant curves after a hamiltonian normalization (see text) up to the maximum normalization order $r_{max} = 2$ (left panel), or $r_{max} = 8$ (right panel). There are 15 theoretical invariant curves shown, corresponding to the constant 'label' values of $I_\psi^{(2)}$ (in panel (a), or $I_\psi^{(8)}$, respectively, in panel (b)) given by $I_\psi^{(2)} = 0.065n - 0.008$, $n = -9, -8, \dots, 5$.

How well can the above transformations account for a precise analytical description of motions on rotational tori? Figure 2 shows a comparison between the theoretical invariant curves arising from the above expressions and the true invariant curves corresponding to the intersection of the rotational invariant tori

in the domain $0.5 \leq p \leq 1.3$, with the surface of section $\phi(mod 2\pi) = 0$. The theoretical invariant curves are computed as follows: due to the surface of section condition, we first set $\phi^{(2)} = \phi = 0$. Then, we are left with expressions yielding ψ and I_ψ each in terms of $\psi^{(2)}$ and $I_\psi^{(2)}$. However, $I_\psi^{(2)}$ is an integral under the normal form dynamics. Thus, each (constant) value of $I_\psi^{(2)}$ represents a label value for one rotational torus. In particular, the value $I_\psi^{(2)} = 0$ represents a torus around the value $p = p_* = 1.46410$ in the original variables. By fixing a value for $I_\psi^{(2)}$, and giving several values to $\psi^{(2)}$ in the interval $0 \leq \psi^{(2)} < 2\pi$, we compute both ψ and I_ψ (and hence $p = p_* + I_\psi$) via the transformation equations. This yields theoretical invariant curves that can be compared to the numerical invariant curves on the surface of section.

Figure 2a shows this comparison using the normal form computation up to second order in λ, for which the explicit parametric formulae corresponding to invariant curves are given above. The theoretical invariant curves are shown by thick dotted lines, superposed to the true invariant curves. We see that, already at this order of approximation, the theoretical invariant curves explain the shape of the true invariant curves in a domain above $p = 1.2$, although they have visible differences from them (of order 10^{-2}). The approximation becomes worse as we approach closer to the separatrix of an island of stability located in the lower part of the figure. In fact, we observe that the theoretical invariant curves cannot represent the shape of the invariant curves in *resonant domains*, i.e. where islands of stability are formed. This is due to the fact that the basic frequencies ω_* and ω employed in the current normal form construction are non resonant. We will see, however (subsection 2.9) that it is possible to make a so-called resonant normal form construction, accounting locally for the dynamics within resonant domains of the model (3) that give rise to island chains as the one shown in Fig.2.

On the other hand, far from resonant domains the local approximation by the theoretical invariant curves becomes better, at least for low orders, as we increase the order of normalization. Thus, Fig.2b shows the comparison between theoretical and numerical invariant curves when the maximum normalization order is equal to $r = 8$. We now see that in the domain $p > 1.2$ the theoretical invariant curves nearly coincide with the true ones, as there are no visible differences in the scale of Fig.2b. In fact, the accuracy of the approximation at this order can be checked to be of the order of 10^{-4}, and the same result is reached by checking the time variations of the quantity $I_\psi^{(8)}(t)$, which is an integral of the normal form dynamics, when computed along the numerical orbits lying on some invariant curves of Fig.2.

2.7. (Non-)Convergence properties. Small divisors

So far, we have seen that a normal form computation already at an order as low as two may provide a useful practical representation of the regular motions of a system under study. Furthermore, calculations like the one of Fig.2 yield the impression that by going to higher and higher normalization order, we can approximate regular motions up to any desired level of accuracy. However, one can see that by the method of Birkhoff normalization exposed in subsections

2.4 and 2.5, this is *not* possible, i.e., there is a *finite precision level* reached at an optimal order, beyond which the above method does not converge. This is because the normalization algorithm of subsection 2.5 leads to a sequence of remainder values, as a function of the normalization order r, which is not a convergent, but only an *asymptotic* sequence.

Let $||R^{(r)}||_{W^{(r)}}$ denote a suitably defined norm for the remainder series $R^{(r)}$ at the r-th normalization step, that gives the size of the remainder in a subdomain $W^{(r)}$ of the phase space contained in the domain of analyticity of the function $R^{(r)}$.[12] Our statement on non-convergence of the Birkhoff normalization process can be formulated as follows:

i) If the initial Hamiltonian is an analytic function in an open domain $W^{(0)}$ of the phase space, then, for arbitrarily high normalization order r, there is a domain $W^{(r)}$, which is a restriction of the initial analyticity domain $W^{(0)}$, in which the remainder $R^{(r)}$ is an analytic function with bounded norm, i.e. $||R^{(r)}||_{W^{(r)}} < \infty$.

ii) The sequence $||R^{(r)}||_{W^{(r)}}$, for $r = 1, 2, 3, \ldots$ has an asymptotic behavior. Thus, initially (at low orders) $||R^{(r)}||_{W^{(r)}}$ decreases as r increases, up to an optimal order r_{opt} at which $||R^{(r_{opt})}||_{W^{(r_{opt})}}$ becomes minimum. Then, for $r > r_{opt}$, we find that $||R^{(r)}||_{W^{(r)}}$ increases with r, and we have $||R^{(r)}||_{W^{(r)}} \to \infty$ as $r \to \infty$.

iii) *Exponential stability:* under particular assumptions for the original Hamiltonian, we find that the optimal remainder is exponentially small in a small parameter μ, namely

$$||R^{(r_{opt})}||_{W^{(r_{opt})}} \sim \exp\left(-\left(\frac{\mu_0}{\mu} \right)^b \right) \qquad (45)$$

where μ_0 and b are positive constants and μ is a parameter related to small quantities appearing in the problem under study. Thus, in the Hamiltonian (3) we can construct exponential estimates where μ coincides with ϵ, but also, for fixed ϵ and p_*, estimates where μ gives the distance from p_* (e.g. coinciding with the variable I_ψ).

Let us make some additional comments on statements (i) to (iii) above.

Regarding (i), a common misunderstanding is that by computing normal forms at higher and higher order r, there is a finite order r beyond which the Hamiltonian $H^{(r)} = Z^{(r)} + R^{(r)}$ becomes a divergent series for any possible datum

---

[12]The author recognizes that too much is said at this point in only one sentence, without detailed explanation. But my purpose is to avoid at this point a detailed reference on how we deal with domains and norms in the functional space of interest in normal form theory, because such reference would distract us considerably from the main line of thought, concerning the practical implementation of normal forms in concrete problems. But we will return to the technical aspects on this issue in subsections 3.3 and 4.2, giving, in 4.2, a more precise definition of the so-called *Fourier-weighted norm*, convenient in the study of the analyticity properties of the various functions appearing in the implementation of normal form theory.

in the phase space. In fact, we can show that for an arbitrarily high value of r, there is always a certain domain $W^{(r)}$ where we have convergence of $H^{(r)}$. However, our interest is in characterizing particular orbits lying in a certain domain of the phase space, denoted, say, by W_{orbits}, whose size depends only on the initial conditions and the time evolution of the orbits themselves, and not on the normal form that we compute for their study. On the other hand, as r increases, the analyticity domain $W^{(r)}$ of the normal form series becomes smaller and smaller. Thus, there comes a critical order r_c, beyond which the domain W_{orbits} is no longer included in $W^{(r)}$. Then, further normalization becomes pointless. It is in this sense that we speak about "the divergence of the Birkhoff series". [13]

(ii) Not every form of perturbation theory leads to a divergent sequence of remainders. In fact, by changing our normalization strategy we can be lead to a convergent normalization. This is exactly the case, for example, of the so-called Kolmogorov algorithm, examined in subsection 2.8, which leads to a normal form allowing to prove the existence of particular solutions lying on invariant tori. However, we will see also that such a normal form implies that the dynamics, in general, is non-integrable in any open domain around one KAM torus. Thus, the convergence in that case is achieved at the cost of giving up the effort to construct local integrals of motion valid in open domains of the action space.

(iii) The fact that we can have an exponentially small remainder implies that in a Hamiltonian of the form $H = H_0 + \epsilon H_1$ the effect of the perturbation term ϵH_1 should not be considered as entirely destroying the regular character of motions due to H_0. In fact, the largest part of the perturbation only causes *deformation effects*, i.e. it deforms the orbits (and the invariant tori) with respect to the orbits (or tori) in the system with Hamiltonian H_0, and only an

---

[13]This is not the whole story. The way by which we restrict domains as we move forward at successive normalization steps depends itself crucially on the way by which we have chosen to construct the normal form, and, in particular, on the choice of 'Hori kernel' (see footnote 7). In fact, if the function Z_0 playing the role of Hori kernel depends on the action variables by terms of order higher than linear, the way by which we restrict domains is dictated by the requirement to avoid *resonances* in the action space. As an example, if, in the Hamiltonian (3) we chose to include $I_\psi^2/2$ in Z_0, i.e. to follow the book-keeping of Eq.(25), then, it is easy to check that the resulting homological equations appearing in the construction of the Birkhoff normal form are equations of the form

$$\{\omega_* I_\psi + \omega I + \frac{I_\psi^2}{2}, \chi_r\} + h_r^{(r-1)} = 0$$

for some functions $h_r^{(r-1)}$ determined along the normalization process. But then, it is easy to check that χ_r contains divisors of the form $k_1(\omega_* + I_\psi) + k_2\omega$, i.e. linearly depending on the action I_ψ. Assuming, now, that trigonometric terms $\exp(i(k_1\phi_1 + k_2\phi_2))$ of *all* possible wave vectors (k_1, k_2) are generated as we proceed in successive normalization steps (or, as in other models examined in section 3, that they are present already from the start) we can see that a construction of this form cannot be defined in any *open interval* of values of the action I_ψ on the real axis. This is because there is a dense set of values $I_\psi = -\omega_* - (k_2/k_1)\omega$ for which some divisor in the series, at some order, will become *equal to zero exactly*. Thus, with such a normalization scheme, we can only proceed by 1) an algorithm eliminating only a finite number of harmonics at every step, and 2) excluding particular values of the action I_ψ (and some interval around them) from the domain in I_ψ where the series can be valid.

exponentially small part is responsible for long term effects on the stability of orbits. This fact becomes particularly relevant when we examine the speed of diffusion in multidimensional systems, i.e. the phenomenon of Arnold diffusion (section 4).

The asymptotic properties of the normalization process can be seen in our perturbed pendulum example in the following way: After r normalization steps, we find that the remainder function has the following structure:

$$R^{(r)} = H^{(r)}_{r+1} + H^{(r)}_{r+2} + \ldots \tag{46}$$

$$= \sum_{s=r+1}^{\infty} \left(\sum_{\substack{|k_1|+|k_2| \leq 2s-1 \\ k_1,k_2}} \left(b^{(s)}_{k_1,k_2,0} + b^{(s)}_{k_1,k_2,1} I_\psi + \ldots + b^{(s)}_{k_1,k_2,r} (I_\psi)^r \right) e^{i(k_1\psi+k_2\phi)} \right)$$

with real coefficients $b^{(s)}_{k_1,k_2,n}$. We recall again that I_ψ in this expression means the *transformed* action variable $I^{(r)}_\psi$, after r consecutive Lie transformations. Also, the fact that with increasing s we also have an increasing maximum Fourier order $|k_1| + |k_2| \leq 2s - 1$ has been explained in the previous subsection, i.e. the higher order harmonics are generated by the action of repeated Poisson brackets via the Lie operation defining $H^{(r)}$.

If, now, we consider a domain in action space centered around the value $p = p_*$ (which is the origin of our construction), we can estimate the size of the remainder in the interval $-\Delta I_\psi \leq I_\psi \leq \Delta I_\psi$ by taking, e.g., a so-called *majorant series*, i.e. a series in which we take the absolute sum of all terms in Eq.(46), namely:

$$||R^{(r)}||_{\Delta I_\psi} = \tag{47}$$

$$\sum_{s=r+1}^{\infty} \left(\sum_{\substack{|k_1|+|k_2| \leq 2s-1 \\ k_1,k_2}} \left(|b^{(s)}_{k_1,k_2,0}| + |b^{(s)}_{k_1,k_2,1}| \Delta I_\psi + \ldots + |b^{(s)}_{k_1,k_2,r}| (\Delta I_\psi)^r \right) \right) \ .$$

In the computer we cannot store infinitely many remainder terms, thus we have to rely on a finite truncation of the sum (47) at a maximum order s_{max}:

$$||R^{(r)}||_{\Delta I_\psi, \leq s_{max}} = \tag{48}$$

$$\sum_{s=r+1}^{s_{max}} \left(\sum_{\substack{|k_1|+|k_2| \leq 2s-1 \\ k_1,k_2}} \left(|b^{(s)}_{k_1,k_2,0}| + |b^{(s)}_{k_1,k_2,1}| \Delta I_\psi + \ldots + |b^{(s)}_{k_1,k_2,r}| (\Delta I_\psi)^r \right) \right)$$

checking numerically that s_{max} is sufficiently large for the quantity $||R^{(r)}||_{\Delta I_\psi, \leq s_{max}}$ to have practically reached the remainder's limiting value (which corresponds to $s_{max} \to \infty$).

The computation of the quantity $||R^{(r)}||_{\Delta I_\psi, \leq s_{max}}$ allows us to test the convergence properties of the adopted Birkhoff normalization. Namely, by computing the value of $||R^{(r)}||_{\Delta I_\psi, \leq s_{max}}$ as a function of the normalization order r, we can check up to what order $||R^{(r)}||_{\Delta I_\psi, \leq s_{max}}$ keeps decreasing with r. In a
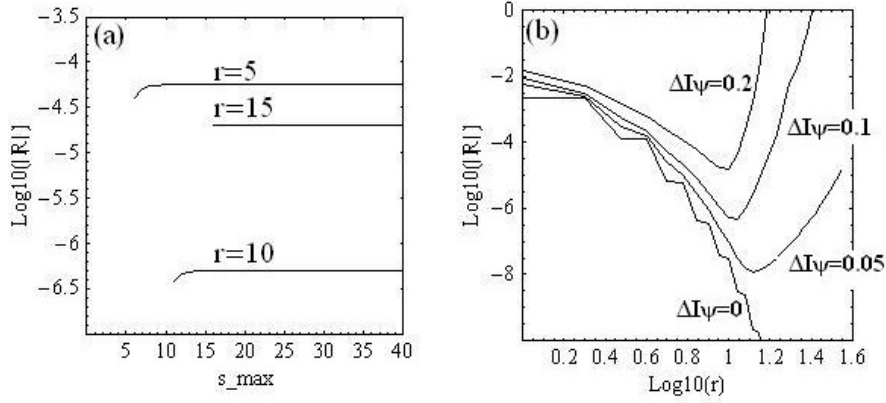
Figure 3.   (a) The numerically computed value of the remainder $||R^{(r)}||_{\Delta I_\psi, \leq s_{max}}$ (Eq.(48)), as a function of the truncation order s_{max}, for the Birkhoff normal form computation of subsections 2.4 and 2.5, for $\Delta I_\psi = 0.1$, and for three different normalization orders, namely $r = 5$, $r = 10$, $r = 15$. The fact that in all three cases $||R^{(r)}||_{\Delta I_\psi, \leq s_{max}}$ stabilizes to a final value as s_{max} increases indicates that the remainder series is absolutely convergent in the domain considered. (b) The quantity $||R^{(r)}||_{\Delta I_\psi, \leq 40}$ for four different values of ΔI_ψ, namely $\Delta I_\psi = 0$, $\Delta I_\psi = 0.05$, $\Delta I_\psi = 0.1$, and $\Delta I_\psi = 0.2$ (curves from bottom to top respectively).

convergent process, we should have $||R^{(r)}||_{\Delta I_\psi, \leq s_{max}} \to 0$ as $r \to \infty$. In reality, however, we find that $||R^{(r)}||_{\Delta I_\psi, \leq s_{max}}$ reaches a minimum at a finite order r.

Figure 3 shows such a numerical test, which illustrates the asymptotic behavior of the remainder for the Birkhoff normal form calculation of our example of subsections 2.4 and 2.5. In producing this figure, we have used a computer program in fortran[14] which computes normal forms up to a high order. We set the maximum truncation order as $s_{max} = 40$, and proceed up to the maximum normalization order $r_{max} = 35$. Figure 3a shows a numerical probe of the convergence of the remainder function in a domain indicated in the figure caption. We set $\Delta I_\psi = 0.1$, and, for different normalization orders r, we compute the truncated remainder at various truncation orders s_{max} in the range $r < s_{max} \leq 40$. The figure shows the value of $||R^{(r)}||_{\Delta I_\psi, \leq s_{max}}$, in the examples of three different normalization orders, namely $r = 5$, $r = 10$, and $r = 15$, as a function of the truncation order s_{max}. Clearly, we see that in all three cases the value of $||R^{(r)}||_{\Delta I_\psi, \leq s_{max}}$ stabilizes as s_{max} increases, indicating the absolute convergence of the remainder function in the domain considered. However, we also observe that the value of $||R^{(r)}||_{\Delta I_\psi, \leq s_{max}}$ for both $r = 5$ and $r = 15$ is *higher* than its value for $r = 10$. This fact implies that the value of the remainder becomes minimum for some normalization order between $r = 5$ and $r = 15$. This is clearly shown in Fig.3b, showing the evolution of the quantity $||R^{(r)}||_{\Delta I_\psi, \leq 40}$, as a function of r, in four different cases, namely $\Delta I_\psi = 0$, $\Delta I_\psi = 0.05$, $\Delta I_\psi = 0.1$, and

---

[14] All the programs used by the author for this tutorial are freely available upon request.

$\Delta I_\psi = 0.2$. Except for the first case, where we have not reached a minimum of $||R^{(r)}||_{\Delta I_\psi, \leq 40}$ up to $r = 35$, in all other cases we find that the minimum occurs at an order within the range considered in Fig.3b. In the case $\Delta I_\psi = 0.1$ (referring to Fig.3a), the optimal normalization order, as indicated by the corresponding minimum in Fig.3b, is $r = 11$ ($\log_{10}(r) = 1.04$). Furthermore, we observe that the optimal normalization order decreases as ΔI_ψ increases, while the optimal remainder value $||R^{(r_{opt})}||_{\Delta I_\psi, \leq 40}$ increases with increasing ΔI_ψ. Despite this increase, however, for $\Delta I_\psi = 0.2$ we still have a rather small optimal remainder value, of order $\sim 10^{-5}$, which indicates that the so-computed normal form still approximates reasonably well the true dynamics.

We now focus on the following basic question: what is the cause of the asymptotic behavior of the remainder shown in figure 3? We can see that, despite the fact that in every normalization step we eliminate from the Hamiltonian terms of higher and higher book-keeping order, the process becomes divergent due to the presence in the series of *small divisors*, and, in particular, due to the fact that these divisors exhibit a bad *accumulation* in the denominators of the series terms.

We recall that small divisors appear in the series due to the solution of the homological equation (36) for all generating functions χ_1, χ_2, etc. More precisely, if the terms to be eliminated at the r-th normalization step (denoted by $h_r^{(r-1)}$) are written in the form of a sum of Fourier terms

$$h_r^{(r-1)} = \sum_{k_1, k_2 \notin \mathcal{M}} b_{k_1, k_2}^{(r-1)}(I_\psi, I) e^{i(k_1 \psi + k_2 \phi)} \tag{49}$$

the homological equation (36) (written for the order r rather than $r + 1$) is satisfied by setting χ_r equal to:

$$\chi_r = \lambda^r \sum_{k_1, k_2 \notin \mathcal{M}} \frac{b_{k_1, k_2}^{(r-1)}(I_\psi, I)}{i(k_1 \omega_* + k_2 \omega)} e^{i(k_1 \psi + k_2 \phi)} \quad . \tag{50}$$

Equation (50) contains divisors of the form $k_1 \omega_* + k_2 \omega$. Notice that in Eqs.(49) and (50), we have changed the way by which we denote those wavenumbers (k_1, k_2) which are excluded from the sum in the r.h.s. Namely, instead of writing $|k_1| + |k_2| \neq 0$ (as in Eq.(33)), we use a new symbol, $\mathcal{M}$, to denote the set of all excluded wavenumbers. This set is called the *resonant module*. In most forms of perturbation theory, the resonant module is determined by the requirement that no extremely small divisors, or divisors equal to zero exactly, appear in the solution (50). In fact, if ω_* and ω are incommensurable, the only possibility for $k_1 \omega_* + k_2 \omega = 0$ is provided by both k_1 and k_2 being equal to zero. Thus, if we choose the resonant module by the requirement to exclude the appearance of only zero divisors, the resonant module is:

$$\mathcal{M} = \{k_1, k_2 : |k_1| + |k_2| = 0\} \quad .$$

However, we will see below (subsection 2.9) that in the theory of *resonant normal forms* the so-arising resonant module necessarily contains also non-zero wavevectors (k_1, k_2).
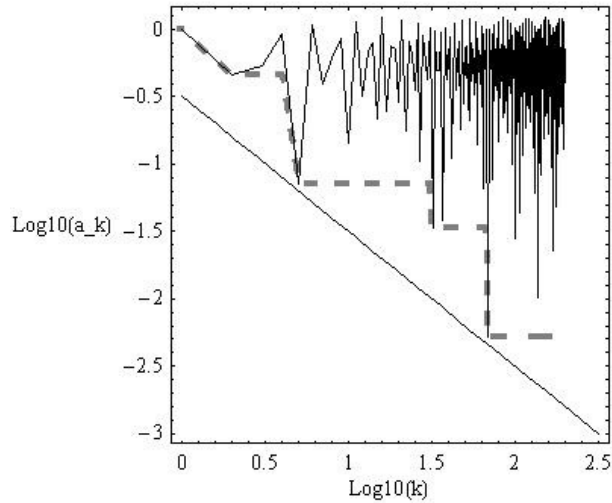
Figure 4.   The minimum divisor a_k (Eq.(51)) as a function of the order k of
the divisor in our numerical example (solid curve). The gray dashed curve
shows the minimum divisor that appears up to the order k. The straight line
represents a low 'diophantine' bound for small divisors explained in subsection
2.8 (see Eq.(62)).

It is possible to see now that, even after the choice of a resonant module,
we are always left in a series with non-zero, but quite small divisors. These
divisors exist as a consequence of the fact that even incommensurable frequen-
cies form ratios which are close to rational numbers. Let us exemplify this
with the frequencies used in our numerical example, namely $\omega_* = 2(\sqrt{3} - 1) =$
1.4641016151377546..., $\omega = 1$. We define $k = |k_1| + |k_2|$ as the *order* of the
divisor $k_1\omega_* + k_2\omega$. Furthermore, we define the minimum of all divisors at a
given order by:

$$a_k = \min\left\{|k_1\omega_* + k_2\omega| : |k_1| + |k_2| = k\right) \quad . \tag{51}$$

Figure 4 shows a_k as a function of k for the numerical frequencies of our ex-
ample. We observe that at most orders k the smallest possible divisor a_k has
relatively large size (above 10^{-1}, and most of them close to 1). However, there
are particular orders at which quite small divisors appear. For example:

$$\text{order 5:} \quad |2\omega_* - 3\omega| = 7.179677 \times 10^{-2}$$

$$\text{order 32:} \quad |13\omega_* - 19\omega| = 3.332099 \times 10^{-2}$$

$$\text{order 69:} \quad |28\omega_* - 41\omega| = 5.154776 \times 10^{-3} \quad .$$

These divisors form sharp reversed spikes in the curve a_k of Fig.4. What are
the important effects they introduce? Let us consider, for example, the case
of the divisor $a_{32} = 3.332099 \times 10^{-2}$. In our example, we have seen that at

the r-th normalization step the harmonics appearing in the Hamiltonian can be up to order $k = 2r - 1$. Thus, a term of the form $\exp(i(k_1\psi + k_2\phi))$ with $|k_1| + |k_2| = 32$ can only appear after the normalization order $r = 17$. However, once it appears, it generates a divisor a_{32} which could be as small as 3.3×10^{-2} in the series, via the solution of Eq.(50) for χ_{17}. The key point now is that this divisor *repeats appearing* at all subsequent normalization steps $r = 18$, $r = 19$, ..., thus generating a *sequence* of terms containing powers of this divisor, i.e. a_{32}, $(a_{32})^2$, ... [15]

By carefully studying our examples so far, it is easy to see that this repetition is a consequence of our chosen normalization scheme. To show this, let us consider a sequence of this type, which starts being formed after the normalization order $r = 17$. Due to Eq.(50), some term in χ_{17} acquires a divisor a_{32}. We introduce a heuristic (although rather unusual) notation to indicate this:

$$\chi_{17} \to \lambda^{17}\frac{e^{i32\Phi}}{a_{32}} \quad .$$

The above notation means: in χ_{17}, which has a book-keeping parameter λ^{17} in front, there are terms of Fourier order 32 containing small divisors bounded from below by a_{32}.

Let us now check how this divisor propagates at subsequent orders. The Hamiltonian normalization at $r = 17$ yields

$$H^{(17)} = \exp(L_{\chi_{17}})H^{(16)} = H^{(16)} + \{H^{(16)}, \chi_{17}\} + ...$$

The Hamiltonian $H^{(16)}$ is in normal form up to order $O(\lambda^{16})$. Thus, at its lowest two orders we have

$$H^{(16)} = Z_0 + \lambda Z_1 + ...$$

Isolate now, in $H^{(17)}$, the following term, arising from the Poisson bracket $\{H^{(16)}, \chi_{17}\}$:

$$\{H^{(16)}, \chi_{17}\} \to \{\lambda Z_1, \chi_{17}\} = \left\{\lambda\frac{I_\psi^2}{2}, \chi_{17}\right\} \to \left\{\lambda\frac{I_\psi^2}{2}, \lambda^{17}\frac{e^{i32\Phi}}{a_{32}}\right\} \quad .$$

The last Poisson bracket produces a term whose size is rather easy to estimate, just thinking in the way by which the Poisson bracket derivatives act on the various parts of it. We have

$$\{\frac{I_\psi^2}{2}, e^{i(k_1\psi+k_2\phi)}\} = -ik_1 I_\psi e^{i(k_1\psi+k_2\phi)} \quad .$$

---

[15] It is important to recall that, depending on the initial Hamiltonian model, a Fourier term $\exp(i(k_1\psi + k_2\phi))$ with the particular combination of wavenumbers (k_1, k_2) producing the smallest possible divisor at order 17 (i.e. in our example $\exp(\pm i(13\psi - 19\phi))$) may or may not have been generated in the normalized Hamiltonian. Thus, a correct reading of all estimates using divisors discussed hereafter is that these are *lower bound* estimates on the size of divisors, or upper bound estimates on the size of the various terms appearing in the series.

The wavenumber k_1 satisfies $|k_1| \leq 2r - 1$, and typically we have the estimate that $|k_1|$ grows linearly with r, or $|k_1| = O(r)$. Thus we can write

$$\left\{ \lambda \frac{I_\psi^2}{2}, \lambda^{17} \frac{e^{i32\Phi}}{a_{32}} \right\} \to \lambda^{18} I_\psi \frac{O(17)e^{i32\Phi}}{a_{32}} \quad .$$

The key remark is that after all the operations, the new Hamiltonian contains the *same* Fourier term as before, but with a different coefficient. That is, by accomplishing one normalization step, besides the generation of new terms of higher and higher harmonics, we also have repetitions of terms of the same harmonics.

However, consider now the next normalization step. The term $\lambda^{18} I_\psi O(17)$ $e^{i32\Phi}/a_{32}$ is part of $h_{18}^{(17)}$, i.e. part of the Hamiltonian to be now normalized. The normalization will be done by the generating function χ_{18}. The crucial remark is, that due again to Eq.(50), the generating function χ_{18} will acquire a second divisor a_{32}, namely:

$$\chi_{18} \to \lambda^{18} \frac{I_\psi O(17)e^{i32\Phi}}{(a_{32})^2} \quad .$$

The same effect appears, now, at all subsequent normalization steps. As a result, after n steps the generating function χ_{17+n} contains a term with n divisors equal to a_{32}, i.e.:

$$\chi_{17+n} \to \lambda^{17+n} \frac{(I_\psi)^n 17^n e^{i32\Phi}}{(a_{32})^{n+1}}$$

Let us generalize: if a new small divisor appears at the normalization order $2r_0 - 1$, the associated term in the generating function produces a sequence of terms at subsequent steps, with coefficients growing *geometrically*:

$$\lambda^{r_0} \frac{1}{a_{2r_0-1}} \to \lambda^{r_0+1} \frac{I_\psi r_0}{a_{2r_0-1}^2} \to \lambda^{r_0+2} \frac{I_\psi^2 r_0^2}{a_{2r_0-1}^3} \ldots \tag{52}$$

The geometric ratio is just $I_\psi r_0/a_{2r_0-1}$, thus, the domain in action space where we expect the geometric progress to be convergent is given by requiring that $|I_\psi r_0/a_{2r_0-1}| < 1$ or:

$$|I_\psi| < a_{2r_0-1}/r_0 \quad . \tag{53}$$

But according to Fig.(4), as we move on to higher and higher order normalization steps, there are particular orders $r_0, r_0', r_0'', \ldots$ where new, smaller and smaller divisors, appear, and we have

$$a_{2r_0-1}/r_0 > a_{2r_0'-1}/r_0' > a_{2r_0''-1}/r_0'' > \ldots \tag{54}$$

Thus, we conclude that with the appearance of every new divisor, the normalization generates a new geometric sequence of terms *of smaller and smaller domain of convergence*. This ensures that the domain of convergence shrinks to zero as $r \to \infty$.

We have not yet answered why at the optimal normalization order we have an exponentially small estimate (Eq.(45)) for the size of the remainder. However, a heuristic derivation of such estimates can be given in more general models than the one considered up to now. Such a derivation will be made in section 4 (see subsection 4.2 and the Appendix).

2.8.   Kolmogorov normal form and the existence of invariant tori

Let us summarize progress so far: we were able to employ a normalization algorithm in order to represent the rotational motions in the perturbed pendulum model (3) via normal forms. This renders possible a number of practical applications stemming from the analytic representation of the regular orbits in terms of series. However, we have seen also that the above normalization process is divergent, a fact implying that we can only reduce the size of the remainder to a finite (albeit, possibly, quite small) optimal lower bound. In many applications, this is sufficient, since we are interested just in approximating the true dynamics by some version of normal form dynamics. However, this type of approach leaves unanswered a question of central interest: can we *prove* that quasi-periodic motions exist? In other words, can we devise a *convergent* normalization algorithm, by which to determine unambiguously the existence of quasi-periodic trajectories moving on invariant tori?

A positive answer to this question is provided by the celebrated Kolmogorov (1954) - Arnold (1963) - Moser (1962) theorem (KAM), which proves the existence of a large set of invariant tori in nearly-integrable Hamiltonian systems satisfying some so-called *analyticity* and *non-degeneracy* conditions.

In the present section we present the normalization algorithm due to Kolmogorov, following a Lie series approach. Also, as in Giorgilli and Locatelli (1997), in our demonstration example we employ a *linear* scheme in which the normalization progresses, as usually, in ascending powers of λ. However, later in this subsection we explain also the so-called *quadratic* scheme, in which the normalization proceeds in groups of terms of book keeping order λ, then λ^2 and λ^3 in one step, then λ^4 to λ^7 in one step, etc. We also discuss why the two schemes are essentially equivalent as far as the accumulation of small divisors in the series terms is concerned. In fact, such accumulation leads to a so-called *quadratic convergence* of the Kolmogorov normal form. Finally, we discuss some practical advantages of Kolmogorov's algorithm with respect to Birkhoff's algorithm regarding the accuracy of computation of particular quasi-periodic solutions in the system (3).

We start with the basic idea of Kolmogorov's scheme, which is simple and drastic: so far, our normalization strategy has been to try to eliminate all terms depending on the angles from the Hamiltonian (3), or its transformed forms $H^{(1)}$, $H^{(2)}$ etc., hoping to give the Hamiltonian a form as close as possible to integrable. Clearly, however, the Hamiltonian (3) is not integrable. Thus, our effort resulted in a bad accumulation of divisors, which eventually causes the series to diverge. In Kolmogorov's scheme, instead, we abandon from the start the process of eliminating from the Hamiltonian *all* the terms depending on the angles. Instead, we only eliminate *a small subset* of terms depending on the angles, which, as shown below, are selected to be the terms preventing to establish the existence, under the Hamiltonian flow, of *a particular equilibrium solution* corresponding to motion on a torus with frequencies selected in advance. Thus, the Kolmogorov normal form continues to represent a non-integrable Hamiltonian, like the original one, which, however possesses a particular equilibrium solution corresponding to a torus.

To fix ideas, let us return to our usual example of the Hamiltonian (3). The first step in Kolmogorov's algorithm is the same as what we have done so far, i.e., to consider a fixed action value p_* and expand the Hamiltonian around it. This leads to the Hamiltonian (23). For reasons that will soon become clear, we now choose to change our book-keeping according to

$$
\begin{aligned}
H &= H_0 + \lambda H_1 = 1.46410 I_\psi + I + \frac{I_\psi^2}{2} \\
&- \lambda \left(0.08 \cos \psi + (0.0078851 + 0.0032 I_\psi) \cos \psi \cos \phi \right) \quad .
\end{aligned}
\tag{55}
$$

We observe that the term H_0 has no dependence on the angles, while the term H_1 has such a dependence, and this is in terms which are either constant or linear in the actions. We now give the following definition:

Let $\mathbf{\Omega}$ be a $n-$dimensional frequency vector. A Hamiltonian function of the form

$$
K = \mathbf{\Omega} \cdot \mathbf{I} + Z(\mathbf{I}) + H_1(\mathbf{I}, \phi)
\tag{56}
$$

where $(\mathbf{I}, \phi)$ are $n-$dimensional action-angle variables, is said to be 'in Kolmogorov normal form' if the functions $Z(\mathbf{I})$, $H_1(\mathbf{I}, \phi)$ are *at least quadratic* in the actions $\mathbf{I}$.

According to this definition, the Hamiltonian (55) is not in Kolmogorov normal form, since H_1 contains terms independent of the actions and linear in the action I_ψ.

Two points must be stressed:

i) A Hamiltonian like (56) is in general non-integrable, since it exhibits a non-linear coupling of the action - angle variables.

ii) However, the fact that Z and H_1 have quadratic (or higher order) dependence on the actions implies that the Hamiltonian K has a particular fixed point solution, i.e. the fixed point $\mathbf{I} = 0$. Indeed, taking into account the quadratic dependence we find immediately that Hamilton's equations for K read:

$$
\begin{aligned}
\dot{\phi} &= \frac{\partial K}{\partial \mathbf{I}} = \mathbf{\Omega} + O(\mathbf{I}) \\
\dot{\mathbf{I}} &= -\frac{\partial K}{\partial \phi} = O(\mathbf{I^2}) \quad .
\end{aligned}
\tag{57}
$$

Thus, if we set $\mathbf{I} = 0$, we find $\dot{\mathbf{I}} = 0$, hence $\mathbf{I}(t) = 0$ at all times t. Furthermore, $\dot{\phi} = \mathbf{\Omega}$ yielding $\phi = \mathbf{\Omega} t + \phi_0$. Thus, in a Hamiltonian of the form (56) we have an invariant torus solution, with fixed frequencies $\mathbf{\Omega}$.

In summary, the Kolmogorov normal form is, in general, a non-integrable Hamiltonian which, however, is guaranteed to possess *one* solution lying on an invariant torus with fixed frequencies.

A Kolmogorov normalization algorithm is an algorithm intending to bring a particular Hamiltonian function in Kolmogorov normal form. For this to be possible, the original Hamiltonian must fulfil particular *analyticity* and *non-degeneracy* conditions. These conditions will be explained in section 3 below, where we give the general algorithm of computation of the Kolmogorov normal

form. Here, instead, we implement Kolmogorov's algorithm in the specific example of the Hamiltonian (55), showing the steps in some detail so as to facilitate further study of the general algorithm.

Starting from the Hamiltonian (55), in the first normalization step the aim is to bring the Hamiltonian in Kolmogorov normal form up to terms of degree 1 in the book-keeping parameter λ. One step is subdivided in the following three substeps:

Substep 1: Eliminate from $H^{(0)}$ all the $O(\lambda)$ terms depending on the angles and independent of the actions, using a generating function $\chi_{1,0}$. Compute the transformed Hamiltonian $H^{(1,0)} = \exp(L_{\chi 1,0})H^{(0)}$.

Substep 2: Eliminate from $H^{(1,0)}$ all the $O(\lambda)$ terms linear in the actions and independent of the angles, using a generating function $\chi_{1,c}$. Compute the transformed Hamiltonian $H^{(1,c)} = \exp(L_{\chi 1,c})H^{(1,0)}$.

Substep 3: Eliminate from $H^{(1,c)}$ all the $O(\lambda)$ terms linear in the actions and depending of the angles, using a generating function $\chi_{1,1}$. Compute the transformed Hamiltonian $H^{(1)} = \exp(L_{\chi 1,1})H^{(1,c)}$.

Let us implement these steps one by one:

i) Substep 1: elimination of $O(\lambda)$ terms independent of the actions and depending on the angles. These are:

$$h_{1,0}^{(0)} = -0.04\left(e^{i\psi} + e^{-i\psi}\right) - 0.0019712\left(e^{i(\psi+\phi)} + e^{-i(\psi+\phi)} + e^{i(\psi-\phi)} + e^{-i(\psi-\phi)}\right) \ .$$

It is easy to see that the generating function $\chi_{1,0}$ must be determined by the same homological equation as in Eq.(31). In fact, if we assume $\chi_{1,0}$ to be a $O(\lambda)$ quantity, then, in the transformed Hamiltonian $H^{(1,0)} = \exp(L_{\chi_{1,0}})H^{(0)}$, the terms of order $O(\lambda)$ are[16]:

$$H_1^{(1,0)} = H_1^{(0)} + \left\{\omega_* I_\psi + \omega I + \frac{I_\psi^2}{2}, \chi_{1,0}\right\} \ .$$

We observe that, due to the different book-keeping that we adopted in Eq.(55), compared to Eq.(24), we have now the presence of the term I_ψ^2 in the Poisson bracket contributing to $O(\lambda)$ terms. However, determining $\chi_{1,0}$ by the usual homological equation:

$$\{\omega_* I_\psi + \omega I, \chi_{1,0}\} + \lambda h_{1,0}^{(0)} = 0 \tag{58}$$

has no consequences, since the extra Poisson bracket $\{I_\psi^2, \chi_{1,0}\}$ only generates a term of order $O(\lambda)$ which is *linear* in the actions, and such terms are to be eliminated at substep 3.

---

[16]Sufficient familiarity with the notation is assumed by now, so that the meaning of various subscripts or superscripts in all expressions below should be straightforward.

In conclusion:

$$\chi_{1,0} = \lambda i \left[0.027321 \left(e^{i\psi} - e^{-i\psi} \right) \right.$$

$$\left. +0.0008 \left(e^{i(\psi+\phi)} - e^{-i(\psi+\phi)} \right) + 0.0042475 \left(e^{i(\psi-\phi)} - e^{-i(\psi-\phi)} \right) \right] \ .$$

Up to terms $O(\lambda^2)$, the Hamiltonian $H^{(1,0)} = \exp(L_{\chi_{1,0}})H^{(0)}$ is given by:

$$H^{(1,0)} = 1.46410 I_\psi + I + \frac{I_\psi^2}{2}$$

$$+\lambda I_\psi \left[0.027321 \left(e^{i\psi} + e^{-i\psi} \right) + 0.0034475 \left(e^{i(\psi-\phi)} + e^{-i(\psi-\phi)} \right) \right]$$

$$+\lambda^2 \left[7.57016 \times 10^{-4} + 9.41880 \times 10^{-5} \left(e^{i\phi} + e^{-i\phi} \right) \right.$$

$$+3.72565 \times 10^{-4} \left(e^{2i\psi} + e^{-2i\psi} \right) - 6.4 \times 10^{-7} \left(e^{2i\phi} + e^{-2i\phi} \right)$$

$$+9.41880 \times 10^{-5} \left(e^{i(2\psi-\phi)} + e^{-i(2\psi-\phi)} \right)$$

$$-3.2 \times 10^{-7} \left(e^{i(2\psi+2\phi)} + e^{-i(2\psi+2\phi)} \right) + 5.6227 \times 10^{-6} \left(e^{i(2\psi-2\phi)} + e^{-i(2\psi-2\phi)} \right) \ .$$

Substep 2: we observe that in $H^{(1,0)}$ there are no $O(\lambda)$ terms linear in the actions and independent of the angles. We thus have to postpone until second order our discussion of how such terms are eliminated.

Substep 3: elimination of the $O(\lambda)$ terms in $H^{(1,0)}$ linear in the actions and depending on the angles. These are:

$$h_{1,1}^{(1,0)} = I_\psi \left[0.027321 \left(e^{i\psi} + e^{-i\psi} \right) + 0.0034475 \left(e^{i(\psi-\phi)} + e^{-i(\psi-\phi)} \right) \right] \ .$$

Again, we use the same type of homological equation to determine the generating function $\chi_{1,1}$:

$$\{ \omega_* I_\psi + \omega I, \chi_{1,1} \} + \lambda h_{1,1}^{(1,0)} = 0 \ . \tag{59}$$

A similar question as before arises: in computing $H^{(1)} = \exp(L_{\chi_{1,1}})H^{(1,0)}$, how do we deal with the non-elimination of $O(\lambda)$ terms generated by the Poisson bracket $\{I_\psi^2, \chi_{1,1}\}$? However, since $\chi_{1,1}$ is linear in the actions, it follows that the terms under question are *quadratic* in the actions. According to the Kolmogorov scheme, such terms are not to be normalized, thus their presence in the Hamiltonian after normalization is consistent with the algorithm. [17]

---

[17]This last remark sounds like a detail, but in the author's opinion, it shows one of the most beautiful aspects of canonical perturbation theory, namely the fact that the physical principles

In summary:

$$\chi_{1,1} = \lambda I_\psi i \left[-0.018660 \left(e^{i\psi} - e^{-i\psi} \right) - 0.0074284 \left(e^{i(\psi-\phi)} - e^{-i(\psi-\phi)} \right) \right]$$

and the final Hamiltonian $H^{(1)} = \exp(L_{\chi_{1,1}})H^{(1,0)}$ up to terms of second degree is given by:

$$H^{(1)} = 1.46410 I_\psi + I + \frac{I_\psi^2}{2}$$

$$-\lambda I_\psi^2 \left[0.018660 \left(e^{i\psi} + e^{-i\psi} \right) + 0.0074284 \left(e^{i(\psi-\phi)} + e^{-i(\psi-\phi)} \right) \right]$$

$$+\lambda^2 \left[7.57016 \times 10^{-4} - 0.0010708 I_\psi + 0.0012102 I_\psi^2 \right.$$

$$+ \left(3.72565 + 1.74103 I_\psi^2 \right) \times 10^{-4} \left(e^{2i\psi} + e^{-2i\psi} \right) - 3.2 \times 10^{-7} \left(e^{i(2\psi+2\phi)} + e^{-i(2\psi+2\phi)} \right)$$

$$+ \left(9.41880 \times 10^{-5} - 2.67279 \times 10^{-4} I_\psi + 4.15846 \times 10^{-4} I_\psi^2 \right) \left(e^{i\phi} + e^{-i\phi} \right)$$

$$+ \left(5.6227 \times 10^{-6} + 2.75904 \times 10^{-5} I_\psi^2 \right) \left(e^{i(2\psi-2\phi)} + e^{-i(2\psi-2\phi)} \right)$$

$$+ \left(9.41880 \times 10^{-5} + 1.38615 \times 10^{-4} I_\psi^2 \right) \left(e^{i(2\psi-\phi)} + e^{-i(2\psi-\phi)} \right)$$

$$\left. -6.4 \times 10^{-7} \left(e^{2i\phi} + e^{-2i\phi} \right) \right] \quad .$$

We observe that the Hamiltonian $H^{(1)}$ is in Kolmogorov normal form up to order $O(\lambda)$, since all terms at this order depend quadratically on the action I_ψ. However, the $O(\lambda^2)$ part contains now terms to be normalized. In particular, we see a term $-0.0010708 I_\psi$ which is independent of the angles. As shown below, such terms are eliminated by a rather unusual form of generating function. Such elimination guarantees that the torus solution found after the Kolmogorov normalization corresponds to the fixed frequencies selected in the start, i.e. in our case $\omega_* = 2(\sqrt{3} - 1)$ and $\omega = 1$.

We give now in detail the three sub-steps for the normalization of $H^{(1)}$:

'guide' our choice of normalization scheme and determine for the most even the formal aspects of normalization algorithms. One example was already pointed out in subsection 2.4, regarding the choice of 'Hori kernel' (see footnote 7). Here, the main advance due to Kolmogorov, is to understand, precisely, that the presence of terms depending on the angles, but quadratic in the actions, does not influence the presence of a torus solution, despite the fact that it is a resignation from our requirement to produce an integrable approximation to the original Hamiltonian.

Substep 1: elimination of the $O(\lambda^2)$ terms independent of the actions and depending on the angles. These are:

$$h_{2,0}^{(1)} = 3.72565 \times 10^{-4} \left(e^{2i\psi} + e^{-2i\psi} \right)$$

$$-3.2 \times 10^{-7} \left(e^{i(2\psi+2\phi)} + e^{-i(2\psi+2\phi)} \right) + 9.41880 \times 10^{-5} \left(e^{i\phi} + e^{-i\phi} \right)$$

$$+5.6227 \times 10^{-6} \left(e^{i(2\psi-2\phi)} + e^{-i(2\psi-2\phi)} \right) + 9.41880 \times 10^{-5} \left(e^{i(2\psi-\phi)} + e^{-i(2\psi-\phi)} \right)$$

$$-6.4 \times 10^{-7} \left(e^{2i\phi} - e^{-2i\phi} \right) \quad .$$

Hence

$$\chi_{2,0} = \lambda^2 i \left[-1.27233 \times 10^{-4} \left(e^{2i\psi} - e^{-2i\psi} \right) \right.$$

$$+6.49324 \times 10^{-8} \left(e^{i(2\psi+2\phi)} - e^{-i(2\psi+2\phi)} \right) - 9.41880 \times 10^{-5} \left(e^{i\phi} - e^{-i\phi} \right)$$

$$-6.05762 \times 10^{-6} \left(e^{i(2\psi-2\phi)} - e^{-i(2\psi-2\phi)} \right) - 4.88476 \times 10^{-5} \left(e^{i(2\psi-\phi)} - e^{-i(2\psi-\phi)} \right)$$

$$\left. +3.2 \times 10^{-7} \left(e^{2i\phi} - e^{-2i\phi} \right) \right] \quad .$$

The Hamiltonian $H^{(2,0)} = \exp(L_{\chi_{2,0}})H^{(1)}$ is in normal form up to terms of order $O(\lambda)$, while at order $O(\lambda^2)$ we have:

$$H_2^{(2,0)} = \lambda^2 \left[7.57016 \times 10^{-4} - 0.0010708 I_\psi + 0.0012102 I_\psi^2 \right.$$

$$+ \left(-2.54467 I_\psi + 1.74103 I_\psi^2 \right) \times 10^{-4} \left(e^{2i\psi} + e^{-2i\psi} \right)$$

$$+ 1.29865 \times 10^{-7} I_\psi \left(e^{i(2\psi+2\phi)} + e^{-i(2\psi+2\phi)} \right)$$

$$+ \left(-2.67279 I_\psi + 4.15846 I_\psi^2 \right) \times 10^{-4} \left(e^{i\phi} + e^{-i\phi} \right)$$

$$+ \left(-1.21152 I_\psi + 2.75904 I_\psi^2 \right) \times 10^{-5} \left(e^{i(2\psi-2\phi)} + e^{-i(2\psi-2\phi)} \right)$$

$$\left. + \left(-9.76951 \times 10^{-5} I_\psi + 1.38615 \times 10^{-4} I_\psi^2 \right) \left(e^{i(2\psi-\phi)} + e^{-i(2\psi-\phi)} \right) \right] \quad .$$

Substep 2: elimination of the $O(\lambda^2)$ terms linear in the actions and independent of the angles.

There is just one such term, $-0.0010708 I_\psi$. This term would be impossible to

eliminate by a generating function $\chi_{2,c}$ defined by the usual form of homological equation, i.e. an equation of the form

$$\{\omega_* I_\psi + \omega I, \chi_{2,c}\} - \lambda^2 0.0010708 I_\psi = 0$$

because looking for a solution of the above equation in the form of Eq.(50) would require division by a *zero* divisor (since $k_1 = k_2 = 0$). However, precisely at this point *we now exploit the fact that Z_0 contains a non-vanishing quadratic part* $I_\psi^2/2$. We set

$$\chi_{2,c} = X_{2,c}\psi \tag{60}$$

where $X_{2,c}$ is a *constant*, and observe that, through the Lie operation $H^{(2,c)} = \exp(L_{\chi_{2,c}})H^{(2,0)}$, $\chi_{2,c}$ generates a term linear in the actions and independent of the angles via the Poisson bracket

$$\exp(L_{\chi_{2,c}})H^{(2,0)} \rightarrow L_{\chi_{2,c}}H^{(2,0)} \rightarrow \{Z_0, \chi_{2,c}\}$$

$$\rightarrow \left\{\frac{I_\psi^2}{2}, X_{2,c}\psi\right\} = -X_{2,c}I_\psi \quad .$$

In order to eliminate the term $-\lambda^2 0.0010708 I_\psi$ from $H^{(2,0)}$, we then simply set $X_{2,c} = -\lambda^2 0.0010708$. [18]

In summary

$$\chi_{c,2} = -\lambda^2 0.0010708\psi \quad .$$

The transformed Hamiltonian is

$$H^{(c,2)} = \exp(L_{\chi_{2,c}})H^{(2,0)} \quad .$$

We can check that $H^{(2,c)}$ does not differ from $H^{(2,0)}$ up to terms $O(\lambda^2)$, apart from the elimination of the term $-\lambda^2 0.0010708 I_\psi$. On the other hand, differences between $H^{(2,c)}$ and $H^{(2,0)}$ exist at all subsequent orders. Finally, it is

---

[18]We see now the reason for keeping the term $I_\psi^2/2$ in the Z_0 part of the Hamiltonian in the choice of book-keeping. This term is used in sub-step 2 of Kolmogorov's algorithm. In fact, the requirement that the original Hamiltonian *posess* a non-vanishing quadratic part proves here to be crucial for the algorithm to proceed. Recalling that the quadratic term was produced after substituting in the original Hamiltonian (3) the expression $p = p_* + I_\psi$ and expanding, we can restate this condition as the requirement that $\partial^2 H_0(p)/\partial p^2 \neq 0$, where , in the original Hamiltonian, we have just $H_0 = p^2/2 + I$. However, under more general Hamiltonian functions of more than one degrees of freedom, the condition, as shown in section 3, can be formulated as the requirement that the determinant of the Hessian matrix of the unperturbed Hamiltonian with respect to the action variables be different from zero, namely $det(\partial^2 H_0/\partial \mathbf{p}^2) \neq 0$. We have, here, one more example of how a formal aspect of perturbation theory is guided by a physical principle. From the physical point of view, a condition of the form $det(\partial^2 H_0/\partial \mathbf{p}^2) \neq 0$ is called a 'twist condition', i.e. it says that the frequencies on different tori should be different, changing with the values of the actions which label the tori. In the Kolmogorov normal form, however, this condition is used in order that it becomes possible to define the generating functions $\chi_c^{(r)}$ by which we eliminate terms linear in the angles and independent of the actions. A posteriori, it can be seen that by this elimination we ensure, precisely, that in our construction of a torus solution, we do not change frequencies with respect to our initial choice, in which the frequencies depended only on the choice of some label values $\mathbf{p}_*$.

straightforward to see that the linear dependence of $\chi_{2,c}$ in the angles produces no terms in $H^{(2,c)}$ which are not trigonometric in the angles.

Substep 3: elimination of the $O(\lambda^2)$ terms linear in the actions and depending on the angles.

These are

$$h_{2,1}^{(2,c)} = I_\psi \Bigg[-2.54467 \times 10^{-4} \left(e^{2i\psi} + e^{-2i\psi} \right)$$

$$+1.29865 \times 10^{-7} \left(e^{i(2\psi+2\phi)} + e^{-i(2\psi+2\phi)} \right) - 2.67279 \times 10^{-4} \left(e^{i\phi} + e^{-i\phi} \right)$$

$$-1.21152 \times 10^{-5} \left(e^{i(2\psi-2\phi)} + e^{-i(2\psi-2\phi)} \right) - 9.76951 \times 10^{-5} \left(e^{i(2\psi-\phi)} + e^{-i(2\psi-\phi)} \right) \Bigg] \ .$$

These terms can be eliminated by a generating function $\chi_{2,1}$, defined by the usual homological equation. Hence

$$\chi_{2,1} = \lambda^2 I_\psi i \Bigg[8.6902 \times 10^{-5} \left(e^{2i\psi} - e^{-2i\psi} \right)$$

$$-2.63513 \times 10^{-8} \left(e^{i(2\psi+2\phi)} - e^{-i(2\psi+2\phi)} \right) + 2.67279 \times 10^{-4} \left(e^{i\phi} - e^{-i\phi} \right)$$

$$+1.30524 \times 10^{-5} \left(e^{i(2\psi-2\phi)} - e^{-i(2\psi-2\phi)} \right) + 5.06664 \times 10^{-5} \left(e^{i(2\psi-\phi)} - e^{-i(2\psi-\phi)} \right) \Bigg] \ .$$

Finally, computing $H^{(2)} = \exp(L_{\chi_{2,1}})H^{(2,c)}$, the Hamiltonian is brought to Kolmogorov normal form up to order $O(\lambda^2)$, i.e.:

$$H^{(2)} = 1.46410 I_\psi + I + \frac{I_\psi^2}{2} + \lambda Z_1 + \lambda^2 Z_2 + \lambda^3 H_3^{(2)} + \lambda^4 H_4^{(2)} + \ldots \qquad (61)$$

where (apart from a constant)

$$Z_1 = -\lambda I_\psi^2 \Bigg[0.018660 \left(e^{i\psi} + e^{-i\psi} \right) + 0.0074284 \left(e^{i(\psi-\phi)} + e^{-i(\psi-\phi)} \right) \Bigg]$$

$$Z_2 = \lambda^2 I_\psi^2 \Bigg[0.0012102 + 3.47907 \times 10^{-4} \left(e^{2i\psi} + e^{-2i\psi} \right)$$

$$+4.15846 \times 10^{-4} \left(e^{i\phi} + e^{-i\phi} \right) + 2.39948 \times 10^{-4} \left(e^{i(2\psi-\phi)} + e^{-i(2\psi-\phi)} \right)$$

$$+5.36951 \times 10^{-5} \left(e^{i(2\psi-2\phi)} + e^{-i(2\psi-2\phi)} \right)$$

$$-5.27027 \times 10^{-8} \left(e^{i(2\psi+2\phi)} + e^{-i(2\psi+2\phi)} \right) \Bigg] \ .$$

This accomplishes the normalization at order $r = 2$ following Kolmogorov's algorithm for the Hamiltonian (55).

Accumulation of divisors and convergence. After having examined the formal aspects of Kolmogorov's algorithm, we can now ask the same question as in subsection 2.6, namely what are the benefits from computing a Kolmogorov normal form as above. In particular, in what aspects should the above presented algorithm be considered as preferential over the algorithm presented in subsections 2.4 and 2.5?

A basic answer to this question is the following: one can prove that the Kolmogorov normalization is *convergent* in an open domain around the solution of interest, i.e. around $I_\psi = 0$. This implies that in the limit of the normalization process, the final Hamiltonian $H^{(\infty)}$ is entirely in Kolmogorov normal form, while the remainder shrinks to zero. Thus, by this method we can *prove the theorem of Kolmogorov* about the existence of invariant tori in a Hamiltonian system like (3).

Are there not small divisors in this case? In fact, the solution to the homological equations defining the generating functions $\chi_{r,0}$ and $\chi_{r,1}$ for $r = 1, 2, \ldots$ introduce precisely the same divisors as in the normalization scheme of subsections 2.4 and 2.5. However, we can demonstrate that by eliminating only some terms depending on the angles, in the Kolmogorov construction we avoid the generation of any divergent sequence due to a bad *accumulation* of divisors. For example, let us consider the sequence of Eq.(52), which leads to the asymptotic behavior of the series considered in subsections 2.4. and 2.5. In this sequence we have a *growing power* of the variable I_ψ, namely a sequence introduces terms $O(I_\psi) \to O(I_\psi^2) \to O(I_\psi^3) \to \ldots$, due to repeated Poisson brackets of some generating function terms with the Hamiltonian term $I_\psi^2/2$. However, in the Kolmogorov normal form *we stop normalizing such terms, precisely, at the point when a term acquires a factor $O(I_\psi^2)$*. In this way, we do not allow such dangerous sequences to propagate in the series.

We are still left with having to demonstrate that there are no other types of dangerous sequences appearing in the Kolmogorov series. The simplest demonstration relies on the employment of the so-called *quadratic scheme*, which leads to the same accumulation of divisors as in the linear scheme developed above.

In order to understand the quadratic scheme, referring to our usual example, consider again the three sub-steps eliminating unwanted terms at order $O(\lambda^2)$. At sub-step 1, we can now remark the following: instead of considering a generating function $\chi_{2,0}$, eliminating the $O(\lambda^2)$ terms depending on the angles and independent of the actions, we could have considered a generating function of the form $\chi_{2+3,0}$, eliminating *simultaneously* the terms $O(\lambda^2)$ and $O(\lambda^3)$ depending on the angles and independent of the actions. Using the "heuristic" notation of subsection 2.7, the corresponding generating function would be of the form:

$$\chi_{2+3,0} \to C \left(\lambda^2 \frac{e^{iK_2\Phi}}{a_2} + \lambda^3 \frac{e^{iK_3\Phi}}{a_3} \right)$$

where C is a factor coming from the normalization procedure up to that point, and the relevant values for the integers K_2, K_3 and divisors a_2, a_3 will be dis-

cussed below. We can see now that the operation $\exp(L_{\chi_{2+3,0}})H^{(1)})$ will not generate new terms at orders $O(\lambda^2)$ and $O(\lambda^3)$ which are depending on the angles and independent of the actions. To see this, we only have in fact to consider Poisson brackets of $\chi_{2+3,0}$ with terms up to order λ in $H^{(1)}$. We have:

$$H^{(1)} = Z_0 + \lambda Z_1 \rightarrow \omega_* I_\psi + \omega I + \frac{I_\psi^2}{2} + \lambda I_\psi^2 e^{iK_1\Phi} + \dots$$

The Poisson bracket $\{\omega_* I_\psi + \omega I, \chi_{2+3,0}\}$ eliminates the unwanted terms of $H^{(1)}$ depending on the angles and independent of the actions. On the other hand

$$\left\{\frac{I_\psi^2}{2} + \lambda I_\psi^2 e^{iK_1\Phi}, \chi_{2+3,0}\right\} \rightarrow$$

$$\left\{\frac{I_\psi^2}{2} + \lambda I_\psi^2 e^{iK_1\Phi}, \lambda^2 \frac{e^{iK_2\Phi}}{a_2} + \lambda^3 \frac{e^{iK_3\Phi}}{a_3}\right\} \rightarrow$$

$$C\left[-i\lambda^2 K_2 I_\psi \frac{e^{iK_2\Phi}}{a_2} - i\lambda^3 I_\psi \left(\frac{K_3 e^{iK_3\Phi}}{a_3} + \frac{2K_2 e^{i(K_1+K_2)\Phi}}{a_2}\right)\right.$$

$$\left. -2i\lambda^4 K_3 I_\psi \frac{e^{i(K_1+K_3)\Phi}}{a_3}\right] \quad .$$

In conclusion, after computing $H^{(2+3,0)} = \exp(L_{\chi_{2+3,0}})H^{(1)}$ we are left with only terms linear in the actions in $H^{(2+3,0)}$.

In precisely the same way, we can see that the operation $H^{(2+3,c)} = \exp(L_{\chi_{2+3,c}}) H^{(2+3,0)}$ leaves no terms of order $O(\lambda^2)$ and $O(\lambda^3)$ in $H^{(2+3,c)}$ which are linear in the actions and independent of the angles, and, finally, that the operation $H^{(2+3)} = \exp(L_{\chi_{2+3,1}}) H^{(2+3,c)}$ leaves no terms of order $O(\lambda^2)$ and $O(\lambda^3)$ in $H^{(2+3)}$ which are linear in the actions and depending on the angles.

In summary, we conclude that, when implementing Kolmogorov's algorithm, the unwanted terms of orders $O(\lambda^2)$ and $O(\lambda^3)$ can be normalized as a group in one step, divided in the usual three sub-steps referred to as above.

In exactly the same way, we can show that the terms $O(\lambda^4)$, $O(\lambda^5)$, $O(\lambda^6)$, $O(\lambda^7)$ can be normalized as a group in one step (divided in the usual three sub-steps), then $O(\lambda^8), \dots, O(\lambda^{15})$ in one step, etc. This particular way of grouping terms which are to be normalized at successive steps is called a quadratic method.

Let us now examine, with the help of the quadratic scheme, why, despite the accumulation of divisors, the Kolmogorov normalization leads to a convergent series. To this end, we will assume that the divisors appearing in the series satisfy a so-called *diophantine condition*, namely, that there are positive constants γ, τ such that for all wavevectors $\mathbf{k}$ one has:

$$|\mathbf{k} \cdot \omega| > \frac{\gamma}{|k|^\tau} \quad . \tag{62}$$

A condition like (62) implies that at any order $|k|$ there is a lower bound on the smallness of all divisors of the order $|k|$, so that very small divisors can only

appear at very high orders. A diophantine condition of the form (62) is satisfied for a subset of large measure in the set of incommensurable frequencies (ω_*, ω), and it is satisfied also in our numerical example, as shown by the straight line plot in Fig.(4) which corresponds to a lower bound to the divisor values given by a power law of the form (62) for $\gamma = 0.223$, $\tau = 1.02$.

In order, now, to establish a bound for the size of the sequence of terms yielding the worst possible accumulation of divisors in the quadratic scheme, recall (from above) that the Lie operation of $\chi_{2+3,0}$ on $H^{(1)}$ generated, at orders $O(\lambda^2)$ and $O(\lambda^3)$, various terms linear in the actions and depending on the angles. Since higher order divisors appear at higher book-keeping orders, we have $K_3 > K_2$, $a_3 < a_2$. Thus, the terms of largest size linear in the actions are $C\lambda^3 K_3 I_\psi e^{iK_3\Phi}/a_3$. Thus

$$\chi_{2+3,1} \to C\lambda^3 K_3 I_\psi e^{iK_3\Phi}/(a_3)^2 \quad .$$

We can now see that, after the operation $H^{(2+3)} = \exp(L_{\chi_{2+3,1}})H^{(2+3,c)}$, the above term generates in $H^{(2+3)}$ a term of order $O(\lambda^6)$, linear in the actions and depending on the angles

$$\exp(L_{\chi_{2+3,1}})H^{(2+3,c)} \to \frac{1}{2}\left\{\{\omega_* I_\psi + \omega I + ..., \chi_{2+3,1}\}, \chi_{2+3,1}\right\}$$

$$\to \lambda^6 C^2 (K_3)^2 I_\psi e^{iK_6\Phi}/(a_3)^2 \quad .$$

Similar terms are generated by the repeated Poisson brackets of $\chi_{2+3,0}$. For example:

$$\exp(L_{\chi_{2+3,0}})H^{(1)} \to \frac{1}{2}\left\{\{\lambda Z_1, \chi_{2+3,0}\}, \chi_{2+3,0}\right\}$$

$$\to \frac{1}{2}\left\{\{\lambda(I_\psi)^2 e^{ik_1\Phi}, \chi_{2+3,0}\}, \chi_{2+3,0}\right\} \to \lambda^7 C^2 (K_3)^2 e^{iK_7\Phi}/(a_3)^2 \quad .$$

But, now, this last term must be eliminated in the normalization of $H^{(2+3)}$ using the generating function $\chi_{4+5+6+7,0}$, which thus acquires a divisor a_7, i.e.:

$$\chi_{4+5+6+7,0} \to \lambda^7 \frac{C^2 (K_3)^2 e^{iK_7\Phi}}{(a_3)^2 a_7} \quad .$$

Similarly, the term $\lambda^6 (K_3)^2 I_\psi e^{iK_6\Phi}/(a_3)^2$ in the previous example must be normalized by the generating function $\chi_{4+5+6+7,1}$. Thus:

$$\chi_{4+5+6+7,1} \to \lambda^6 \frac{C^2 (K_3)^2 I_\psi e^{iK_6\Phi}}{(a_3)^2 a_6} \quad .$$

Carefully examining the above formulae, the main effect of the quadratic scheme can now be stated: [19] *in every normalization step, the worst possible accumulation of divisors is given by the square of the worst possible divisor product of the*

---

[19]In fact, we have exactly the same effect in the linear scheme as well (see Giorgilli and Locatelli (1997,1999), at the end of the steps 1,2,4,8, etc. of that scheme.

previous step times a new divisor equal to the smallest possible in the current group. Namely we have (taking, as an example, the accumulation of divisors in the generating functions $\chi_{1,0}$, $\chi_{2+3,0}$, etc.):

$$\chi_{1,0} \quad \rightarrow \text{ worst possible factor } \frac{\lambda\mu}{a_1}$$

$$\chi_{2+3,0} \quad \rightarrow \text{ worst possible factor } \frac{\lambda^3\mu^3(K_1)^2}{(a_1)^2 a_3}$$

$$\chi_{4+5+6+7,0} \rightarrow \text{ worst possible factor } \frac{\lambda^7\mu^7(K_1)^4(K_3)^2}{(a_1)^4(a_3)^2 a_7}$$

$$\chi_{8+9+\ldots+15,0} \rightarrow \text{ worst possible factor } \frac{\lambda^{15}\mu^{15}(K_1)^8(K_3)^4(K_7)^2}{(a_1)^8(a_3)^4(a_7)^2 a_{15}}$$

$$\ldots$$

where μ is the largest possible numerical coefficient of all Fourier terms in $h_{1,0}^{(0)}$ (in our numerical example, $\mu = 0.04$).

It is possible to see now, that a sequence of worst possible factors as the above *is bounded by a geometric progress.* To see this, we recall that i) $a_{r_1} > a_{r_2}$ if $r_1 < r_2$, ii) the maximum Fourier order K_r of a term at book-keeping order r is $K_r = K'r - 1$, where, in our example, we have $K' = 2$, iii) $K_r > 1$ for all r, and iv) $\mu < 1$. Taking into account the latter relations we deduce that the fastest growing sequence of terms at successive normalization steps in the above example is bounded by the sequence

$$\frac{2K'\mu}{a_2} \rightarrow \frac{((2K')^2 \cdot (4K'))\mu^2}{(a_2)^2 a_4} \rightarrow \frac{((2K')^4 \cdot (4K')^2 \cdot (8K'))\mu^4}{(a_2)^4(a_4)^2 a_8} \rightarrow \ldots$$

Taking finally into account the diophantine inequality (62), after q normalization steps (by the quadratic scheme), the q-th term of the above sequence reads

$$B_q = \mu^{r_q}\left(\frac{K'^{(1+\tau)}}{\gamma}\right)^{L_q} \cdot 2^{P_q} \tag{63}$$

where

$$r_q = 2^{q-1}, \qquad L_q = \sum_{j=1}^{q} 2^{j-1}, \qquad P_q = \sum_{j=1}^{q} j(2^{q-j}) \ .$$

It is now a simple exercise to show that [20]

$$B_q < \left(\frac{\gamma}{4K'^{(\tau+1)}}\right)\left(\frac{16K'^{(2\tau+2)}\mu}{\gamma^2}\right)^{r_q} \tag{64}$$

---

[20] Prove and use the equalities $L_q = 2^q - 1$ and $P_q = 2^{q+1} - 2$.

i.e. *the fastest growing sequence of terms in Kolmogorov's algorithm is bounded by a geometric progress.* But for μ sufficiently small, the geometric ratio becomes smaller than unity. We thus demonstrate that for sufficiently small μ, the Kolmogorov normalization is convergent, and this ensures the existence of a torus solution with the given diophantine frequencies. [21]

It should be emphasized that the quadratic scheme used above is not indispensable in proving that the worst accumulation of divisors in the Kolmogorov scheme is bounded by a geometric progress. If we work, instead, with a 'classical', i.e. order by order, elimination of terms, book-keeping scheme, we can still recover essentially the same accumulation of divisors as in Eq.(63) (see Giorgilli and Locatelli (1997, 1999)). In fact, in numerical computations it turns out that the classical scheme performs somewhat better in establishing limits of applicability of the Kolmogorov construction in concrete physical systems (see, for example, Giorgilli et al. (2009), or Sansottera et al. (2011) and references there in).
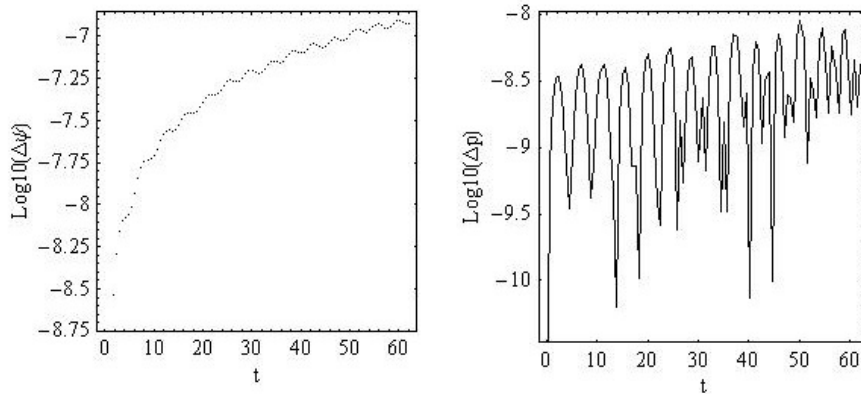
Figure 5.   Time evolution of the absolute error between a theoretical calculation of the time evolution of the variables $\psi(t)$ (left) and $p(t)$ (right) using the Kolmogorov normal form up to normalization order $r = 10$, and the corresponding numerical orbit with initial conditions $\psi = 0$, $p = 1.529162868490481$, up to $t = 60$.

We note finally, that the fact that Kolmogorov's scheme is convergent, has the practical benefit that we can use it in order to obtain very accurate ana-

---

[21]How small should μ be? If we make the calculation in our example, setting $K' = 2$, $\gamma = 0.223$, $\tau = 1.02$, we find $\mu < 1.9 \times 10^{-4}$, while numerically we found that there are invariant tori for μ more than two orders of magnitude larger, i.e. $\mu = 0.04$. However, the source of the discrepancy is the use of unrealistic upper bound limits in the above estimates, while a 'computer-assisted' calculation of the Kolmogorov series up to a high order is likely to yield a much better estimate of the maximum value of μ for which the series converge. This subject has a long history. In one of the first applications of KAM theory by purely analytical means, M. Hénon (1966) showed that in the restricted three body problem we have KAM stable motions if $\mu < 10^{-300}$. However, in recent years there have been quite realistic computer-assisted implementations (or even proofs, using interval arithmetics, of the KAM stability in dynamical systems (see Celletti et al. (2000)), for perturbation parameter values up to 92% of the value for which the existence of a KAM torus is established by numerical means.

lytical expressions for quasi-periodic trajectories using the Lie-transformations from the new to the old canonical variables of the problem under study. The transformations are given by equations quite similar to Eqs.(41). For example, we have:

$$\psi = \exp(L_{\chi_{r,1}}) \exp(L_{\chi_{r,c}}) \exp(L_{\chi_{r,0}}) \ldots \exp(L_{\chi_{1,1}}) \exp(L_{\chi_{1,c}}) \exp(L_{\chi_{1,0}}) \psi^{(r)}$$
(65)

and similarly for all other variables. Furthermore, for the torus solution, we simply substitute in the end $I_\psi^{(r)}(t) = 0$, and $\psi^{(r)}(t) = \omega_* t$ or $\phi^{(r)}(t) = \omega t$. We can then compute theoretically the time evolution of the old variables by the above expressions. Figure (5) shows an example of the error, i.e., the difference between theoretical and numerical calculation for a torus solution in the example treated in the present subsection, and a computation of the Kolmogorov normal form up to order 10 in the book-keeping parameter. The numerical trajectory is found by integrating the original equations of motions with initial conditions obtained by setting $t = 0$ in the above expressions, when the normalization was implemented up to the order $r = 10$. We observe that the error in the computation of the action variable is of the order of 10^{-8}, while in the angles there is a cumulative error of order 10^{-7} at the time $t = 60$, implying a frequency error of the order, again, 10^{-8}. But the error can be reduced to an arbitrarily small amount by normalizing to higher orders r.

2.9. Resonant normal form

So far, we focused on constructing normal forms in the Hamiltonian model (3) around some action value p_* ($p_* = 2(\sqrt{3} - 1)$ in our example) such that the resulting frequency is incommensurable with the perturber's frequency ω. It is immediately clear that the fact that the two frequencies satisfy no resonance condition of the form

$$k_1 \omega_* + k_2 \omega = 0$$
(66)

was essential so far in our success in constructing a normal form, since it ensures that we never encounter in the series (e.g. in the solution (50) of the homological equation) a divisor equal to zero *exactly*. However, such divisors could appear for other choices of p_*. For example, if we choose $p_* = 1$, we have $\omega_* = 1$, thus $\omega_* - \omega = 0$, which is a resonance condition of the form (66), for a particular vector of wavenumbers, denoted hereafter as $\mathbf{k^{(1)}} \equiv (k_1^{(1)}, k_2^{(2)}) = (1, -1)$, or multiples of it.[22] If we choose $p_* = 1$ in the Birkhoff normal form construction

---

[22] In the case of wavenumbers, the use of the superscript (1) has a completely different meaning than in the notation introduced so far for the series. Namely, (1) means 'the first resonance condition'. We will see (section 4) that when we have more than two frequencies, there are cases where we can define more than one linearly independent resonance conditions, the maximum number being equal to the number of frequencies considered minus one. For example, if we have three frequencies $(\omega_1, \omega_2, \omega_3)$, we can have cases with i) no resonance condition (e.g.: $\omega_1 = 1$, $\omega_2 = \sqrt{2}$, $\omega_3 = \sqrt{3}$), ii) one resonance condition (='simply resonant', e.g.: $\omega_1 = 1$, $\omega_2 = \sqrt{2}$, $\omega_3 = 2 + 2\sqrt{2}$, whence $2\omega_1 + 2\omega_2 - \omega_3 = 0$, implying that we have a unique resonant vector $(k_1^{(1)}, k_2^{(1)}, k_3^{(1)}) = (2, 2, -1)$), iii) two independent resonance conditions (='doubly resonant', e.g.: $\omega_1 = 0.4, \omega_2 = 0.2, \omega_3 = 1$, whence $\omega_1 - 2\omega_2 = 0$ but also $2\omega_1 + \omega_2 - \omega_3 = 0$, implying that we have two independent resonant vectors $(k_1^{(1)}, k_2^{(1)}, k_3^{(1)}) = (1, -2, 0)$ and $(k_1^{(2)}, k_2^{(2)}, k_3^{(2)}) =$

of subsections 2.4 and 2.5, we can immediately see that the construction cannot proceed even at order $O(\lambda)$. This, because we can note, for example, that in the expression for the unwanted terms $h_1^{(0)}$, there is a term $e^{i(\psi-\phi)}$ whose elimination, according to Eq.(33), would require a divisor equal to zero exactly. Furthermore, as we have seen, even if the original Hamiltonian contains a finite number of harmonics, the implementation of the normal form algorithm leads to the generation of new ones as we proceed. Thus, it is always possible that we will encounter such a zero divisor at some subsequent normalization step.

What is the physical effect of the so-called 'resonant terms' $e^{i(\psi-\phi)}$, or $e^{-i(\psi-\phi)}$ (or simply $\cos(\psi-\phi)$) in the original Hamiltonian? Returning to the phase portraits of Fig.1, we see that, around $p = p_* = 1$, a conspicuous *island of stability* is formed. In fact, at the center of the island lies a stable periodic orbit, whose fundamental frequency is equal to $\omega = \omega_* = 1$. As shown in Fig.1b, the stability of this orbit is preserved when ϵ is as high as $\epsilon = 1$, and this fact implies the presence of an island of stability around it even when most other invariant curves have been destroyed and replaced by chaotic orbits.

Another relevant remark is that, within resonant domains, the phase portrait takes locally the form of a perturbed pendulum phase portrait, while the limits of the resonant domain (the so-called 'separatrix width', in analogy with the pendulum case), grow with ϵ, at least for small ϵ, up to a point when the separatrix-like chaotic layer formed in the domain around these limits prevails the dynamics. Figure 6 shows these effects. The left column shows a detail of the surface of section computed as in Fig.1, i.e. by numerical orbits, for (a) $\epsilon = 0.04$ and (c) $\epsilon = 0.16$. In fact, a rough estimate shows that by quadrupling the perturbation, the island size grows by a factor ~ 2, i.e. the *square root* of the growth factor of the perturbation.

All these features can be explained by a particular form of normal form theory, called a *resonant normal form*.[23] We will give below a detailed numerical example of how to construct a resonant normal form. However, we can see immediately the outcome of this analysis, shown in the right column of Fig.6, which shows a theoretical phase portrait obtained by a resonant normal form calculation up to order 4, as explained below. We see that the resonant normal form construction: i) explains the fact that the phase portrait in the neighborhood of a resonance has the pendulum form, ii) can predict the size of the resonant domain, and iii) gives a separatrix-like representation for the *shape* of the thin chaotic layer, which is formed in reality by the orbits as computed in

---

$(2, 1, -1)$. In the latter case we can actually form an infinity of different couples of resonant vectors, by taking linear combinations of $\mathbf{k^{(1)}}$ and $\mathbf{k^{(2)}}$ defined as above).

[23]Historically, the implementation of a resonant normal form in the computer preceded that of a non-resonant one. The work of Gustavson (1966) refers to the Hénon - Heiles Hamiltonian, which is a model of the 1:1 resonance. However, the work of Gustavson was itself preceded by the development of a 'direct' method of computation of resonant integrals by Contopoulos (1963) and Contopoulos and Moutsoulas (1965). In the direct method, we perform no canonical transformations, but simply a computation of a formal series defined so as to have vanishing Poisson bracket with the Hamiltonian. The non-resonant case (Whittaker (1916), Cherry (1924), Contopoulos (1960)) is known in galactic dynamics as the 'third integral', i.e. an approximate integral beyond the two exact ones for axisymmetric galaxies, i.e. the energy and the component of the angular momentum along the axis of symmetry.
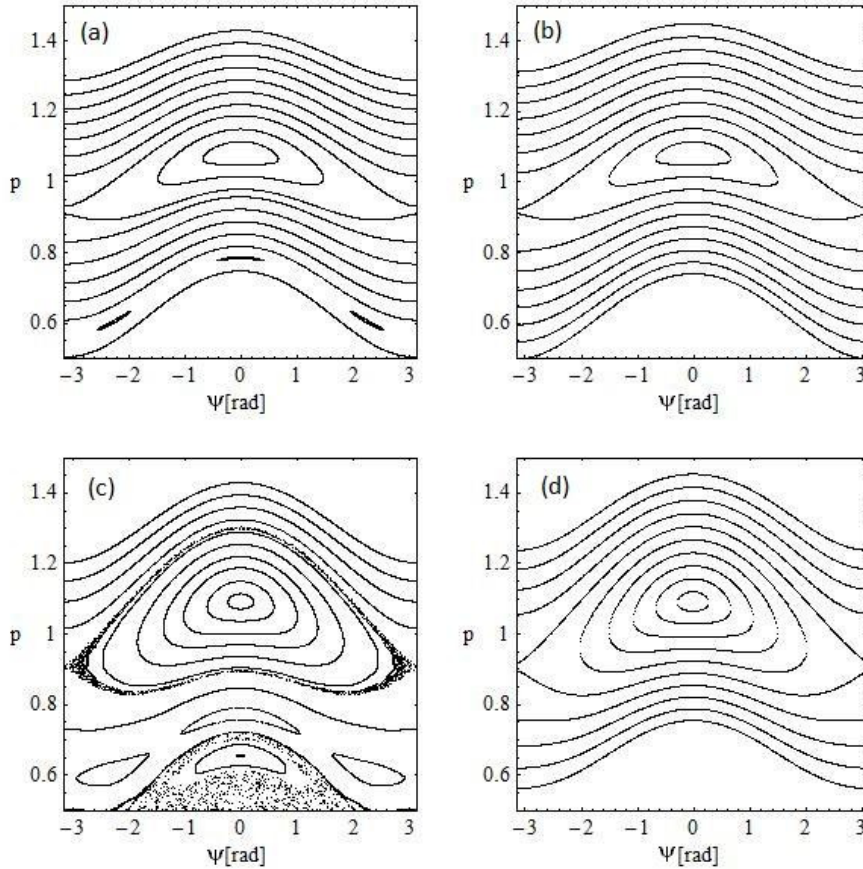
Figure 6. (a) The phase portrait of the model (3) in the 1:1 resonance, for $\epsilon = 0.04$ as computed by numerical integration of orbits, and (b) its theoretical representation by the invariant curves computed via a resonant normal form (see text). (c) and (d) Same as in (a) and (b) respectively, but for $\epsilon = 0.16$. We note that i) the size of the resonant domain has nearly doubled by changing the perturbation by a factor 4, and ii) the theoretical computation yields invariant curves even in a domain where the true dynamics yields a thin chaotic separatrix-like layer.

the full Hamiltonian (3).

The present calculation concerns the model (3), which is of two degrees of freedom. In section 3 we will see that resonant normal form theory can be used in order to describe a new phenomenon taking place within the chaotic layers of resonances in the weakly chaotic regime, namely *Arnold diffusion*. The latter originates from the topological possibility (in three or more degrees of freedom) of having chaotic motions in phase space directions normal to the planes defined by the separatrix-like resonant chaotic layers (while such motions are not possible in systems of two degrees of freedom). In that case, we will see that the *remainder* of the resonant normal form construction accounts for the speed

by which diffusion takes place. A quantitative theory of diffusion in systems of three or more degrees of freedom was developed by Chirikov (1979). This is further reviewed in Cincotta (2002) with a focus on applications in dynamical astronomy. The relation between Chirikov's theory and the outcome of resonant normal form theory is briefly discussed in subsection 4.3.

We now pass to examining the process by which we construct a resonant normal form in the example as above. Essentially, nothing changes with respect to the general normalization algorithm presented in subsection 2.5, apart from recalling that the resonant module (subsection 2.7) contains now wave vectors different from $(k_1, k_2) = 0$. In fact, if we simply define the resonant module to be the set of all wave vectors with associated divisors equal to zero exactly, we can have a consistent normal form algorithm covering all cases, i.e. non-resonant, or resonant of any multiplicity. We thus define:

$$\mathcal{M} = \{\mathbf{k} \equiv (k_1, k_2) : k_1\omega_* + k_2\omega = 0\} \tag{67}$$

and have the following general normalization algorithm:

Recursive normalization algorithm covering all types of resonant conditions

Assuming r algorithm steps have been completed, the Hamiltonian has the form:

$$H^{(r)} = Z_0 + \lambda Z_1 + \ldots + \lambda^r Z_r + \lambda^{r+1} H_{r+1}^{(r)} + \lambda^{r+2} H_{r+2}^{(r)} + \ldots$$

where all Z terms are in normal form, while the remaining terms define the remainder $R^{(r)}$. Then:

1) Isolate from the Hamiltonian $H^{(r)}$ the terms in $H_{r+1}^{(r)}$ to be eliminated at the present step, denoted by $h_{r+1}^{(r)}$. To define these terms, write $H_{r+1}^{(r)}$ as:

$$H_{r+1}^{(r)} = \sum_{k_1, k_2} b_{r+1, k_1, k_2}^{(r)}(\mathbf{I}) e^{i(k_1\psi + k_2\phi)}$$

and choose

$$h_{r+1}^{(r)} = \sum_{k_1, k_2 \notin \mathcal{M}} b_{r+1, (k_1, k_2)}^{(r)}(\mathbf{I}) e^{i(k_1\psi + k_2\phi)}$$

where $\mathcal{M}$ is defined as in Eq.(67).

2) Solve the homological equation

$$\{Z_0, \chi_{r+1}\} + \lambda^{r+1} h_{r+1}^{(r)} = 0$$

and define χ_{r+1}.

3) Compute $H^{(r+1)} = \exp(L_{\chi_{r+1}}) H^{(r)}$. This contains a normal form part:

$$Z^{(r+1)} = Z_0 + \lambda Z_1 + \ldots + \lambda^r Z_r + \lambda^{r+1} Z_{r+1}$$

and a remainder part

$$R^{(r+1)} = \lambda^{r+2} H_{r+2}^{(r+1)} + \lambda^{r+3} H_{r+3}^{(r+1)} + \ldots$$

This completes one full step of the general normalization algorithm.

Let us implement this algorithm in the 1:1 resonant case of the model (3). First, we prepare our Hamiltonian to be in the right form for implementing the algorithm, i.e., as in subsection 2.4, we first set $p = p_* + I_\psi$ in (3) with $p_* = 1$, expand, introduce a book-keeping, and turn trigonometric functions to exponentials. Then (apart from a constant):

$$H^{(0)} = I_\psi + I$$

$$+\lambda \left[\frac{I_\psi^2}{2} - 0.04 \left(e^{i\psi} + e^{-i\psi} \right) \right.$$

$$\left. -(0.0016 + 0.0008 I_\psi) \left(e^{i(\psi+\phi)} + e^{-i(\psi+\phi)} + e^{i(\psi-\phi)} + e^{-i(\psi-\phi)} \right) \right] .$$

Now, we isolate the terms to be normalized at order 1. We observe that the terms $e^{\pm i(\psi-\phi)}$ yield $(k_1, k_2) = (\pm 1, \mp 1)$, which belong to the resonant module $\mathcal{M}$. Thus, they are not to be normalized. In fact, the other term with the same property is $I_\psi^2/2$, whose 'wave vector' is just $(k_1, k_2) = (0, 0)$. In summary:

$$h_1^{(0)} = -0.04 \left(e^{i\psi} + e^{-i\psi} \right) - (0.0016 + 0.0008 I_\psi) \left(e^{i(\psi+\phi)} + e^{-i(\psi+\phi)} \right)$$

leading to

$$\chi_1 = \lambda i \left[0.04 \left(e^{i\psi} - e^{-i\psi} \right) + (0.0008 + 0.0004 I_\psi) \left(e^{i(\psi+\phi)} - e^{-i(\psi+\phi)} \right) \right] .$$

After performing $H^{(1)} = \exp(L_{\chi_1}) H^{(0)}$ we find, up to second order:

$$H^{(1)} = I_\psi + I + \lambda \frac{I_\psi^2}{2} - \lambda(0.0016 + 0.0008 I_\psi) \left(e^{i(\psi-\phi)} + e^{-i(\psi-\phi)} \right)$$

$$+\lambda^2 \left[-1.28 \times 10^{-6} - 6.4 \times 10^{-7} I_\psi + 0.04 I_\psi \left(e^{i\psi} + e^{-i\psi} \right) - 5.6 \times 10^{-5} \left(e^{i\phi} + e^{-i\phi} \right) \right.$$

$$- \left(1.28 \times 10^{-6} + 6.4 \times 10^{-7} I_\psi \right) \left(e^{2i\phi} + e^{-2i\phi} \right)$$

$$+ \left(0.0008 I_\psi + 0.0004 I_\psi^2 \right) \left(e^{i(\psi+\phi)} + e^{-i(\psi+\phi)} \right)$$

$$\left. -8 \times 10^{-6} \left(e^{i(2\psi+\phi)} + e^{-i(2\psi+\phi)} \right) - 3.2 \times 10^{-5} \left(e^{i(2\psi-\phi)} + e^{-i(2\psi-\phi)} \right) \right] .$$

This accomplishes the first step of the resonant normalization process. Clearly, in the second step, the only term to survive in the Hamiltonian (apart from a constant) is $-6.4 \times 10^{-7} I_\psi$. Thus, omitting the details on the form of χ_2, the

Hamiltonian $H^{(2)} = \exp(L_{\chi_2})H^{(1)}$, which is in normal form up to terms $O(\lambda^2)$, is given by

$$H^{(2)} = I_\psi + I + \lambda \frac{I_\psi^2}{2}$$

$$-\lambda(0.0016 + 0.0008I_\psi)\left(e^{i(\psi-\phi)} + e^{-i(\psi-\phi)}\right) - \lambda^2(6.4 \times 10^{-7}I_\psi) + O(\lambda^3) + ...$$

Considering the normal form terms only, setting $\lambda = 1$, ignoring the small linear correction in I_ψ, and passing back to trigonometric functions, we find:

$$Z^{(2)} = I_\psi + I + \frac{I_\psi^2}{2} - (0.0032 + 0.0016I_\psi)\cos(\psi - \phi) \quad . \tag{68}$$

It is now straightforward to see that a Hamiltonian like (68) induces the pendulum dynamics, leading to phase portraits like in Fig.6. The clearest way to see this is by transforming the Hamiltonian (68) to one in *resonant canonical variables* (called sometimes 'fast' and 'slow' variables). For a general resonant wave vector $\mathbf{k}^{(1)}$, the procedure to find the resonant variables is as follows:

i) Define an integer vector $\mathbf{m}$ by the condition

$$\mathbf{m} \cdot \mathbf{k}^{(1)} = 0 \quad . \tag{69}$$

ii) Write:

$$I_\psi = k_1^{(1)}I_R + m_1 I_F \qquad \phi_R = k_1^{(1)}\psi + k_2^{(1)}\phi \tag{70}$$
$$I = k_2^{(1)}I_R + m_2 I_F \qquad \phi_F = m_1\psi + m_2\phi$$

and solve to get I_ψ, I in terms of I_R, I_F. We note that Eq.(70) defines a canonical transformation, since it arises from the generating function $S(\psi, \phi, I_R, I_F) = (k_1^{(1)}\psi + k_2^{(1)}\phi)I_R + (m_1\psi + m_2\phi)I_F$.

iii) Substitute Eq.(70) in the resonant normal form (Eq.(68) in our example). Then, ϕ_F becomes an ignorable variable, implying I_F is an integral of motion of the normal form dynamics. Then, the remaining pair of variables yields the pendulum dynamics.

Let us implement these steps, concerning the resonant normal form $Z^{(2)}$ given in (68). We have:

$$(k_1^{(1)}, k_2^{(1)}) = (1, -1) \qquad \text{implying} \qquad (m_1, m_2) = (1, 1) \quad .$$

Then

$$I_\psi = I_R + I_F, \quad I = -I_R + I_F, \quad \phi_R = \psi - \phi \quad \phi_F = \psi + \phi \quad .$$

Thus, substituting the above expressions into $Z^{(2)}$ we find the form of the resonant Hamiltonian in the resonant canonical variables:

$$Z_{res} = 2I_F + \frac{I_F^2}{2} + I_F I_R + \frac{I_R^2}{2} - (0.0032 + 0.0016(I_R + I_F))\cos(\phi_R) \quad . \tag{71}$$

It is now already clear that, since ϕ_F is ignorable, the variable I_F is an integral of the Hamiltonian flow of Z_{res}. Thus, it can be viewed as a parameter of a one-dimensional pendulum-like Hamiltonian system. In fact, in systems like (3), since I is a dummy action, we can exploit the freedom to assign arbitrary values to I in order to render the value of I_F equal to any convenient value. Since $I_F = (I_\psi + I)/2$, setting, for any initial condition for I_ψ, a corresponding value $I = -I_\psi$, we find $I_F = 0$, and this value is preserved along the flow of Z_{res}. Thus, without loss of generality we can recast Z_{res} as:

$$Z_{res} = \frac{I_R^2}{2} - (0.0032 + 0.0016 I_R) \cos(\phi_R) \quad . \tag{72}$$

We observe that, up to second order, and except for a small correction, this resonant normal form is what we would have found if we only kept the resonant terms in the expression of the original Hamiltonian, but as a function of the new, instead of old, resonant canonical variables. Leaving ϵ as a parameter, we find the following expression for the resonant normal form in this approximation:

$$Z_{res} \approx \frac{I_R^2}{2} - \epsilon \left(0.08 + 0.04 I_R\right) \cos(\phi_R) \quad . \tag{73}$$

The so-called *separatrix width*, which determines the size of the resonance, can be estimated as follows. The unstable equilibrium point of Z_{res} is at $I_R = 0$, $\phi_R = \pm\pi$. Substituting these values in (73) we find the energy value at the unstable equilibrium, which is equal to the (constant) energy value along the whole separatrix, i.e. $E_{sep} = 0.08\epsilon$. The separatrix half-width, now, corresponds to the value of $I_{R,sep}$ at $\phi_R = 0$ along the separatrix. We have

$$E_{sep} = \frac{I_{R,sep}^2}{2} - \epsilon \left(0.08 + 0.04 I_{R,sep}\right) \quad .$$

Solving this equation, we find the values $I_{R,sep}^+$ and $I_{R,sep}^-$, in the upper and lower separatrix branch, where the separatrix intersects the axis $\phi_R = 0$. The difference between these two values defines the separatrix width:

$$\Delta I_{R,sep} = \sqrt{(0.08\epsilon)^2 + 1.28\epsilon} \simeq 0.8\sqrt{2\epsilon} \quad . \tag{74}$$

We obtain a basic result, namely that the resonance width scales as the square root of the perturbation parameter ϵ. This is exemplified by a comparison of the top and bottom row of figure 6, where we see that, by increasing the perturbation by a factor 4, the island of the 1:1 resonance has increased in size by a factor ~ 2.

Staying in the same figure, in order to obtain the theoretical invariant curves shown in panels (b) and (d), we work as follows: Starting from a resonant normal form like (71), we fix different values of the *normal form energy E*, and compute many pairs of values (ϕ_R, I_R) along the curves of constant energy $Z_{res}(\phi_R, I_R) = E$. We also know that $I_F = 0$, and for this value we have $I_R = I_\psi$. Also, $\phi_R = \psi - \phi$, thus, on the surface of section $\phi = 0$ we have $\phi_R = \psi$. This allows to obtain a set of pairs of values (ψ, I_ψ) representing the constant energy level curves of Z_{res}. Now, it must be recalled that the variables (ψ, I_ψ) here

referred to are the *new* canonical variables, after the composition of Lie series with all generating functions χ_1, χ_2, etc. However, using Eqs.(41), from each pair of values of the new canonical variables we can compute a corresponding pair of values of the old canonical variables. Obtaining many such pairs gives the set of invariant curves shown in Figs.6b,d. In particular, we observe that the theoretical separatrix shows a deformation in its top part towards higher values of p. This is an effect due to the canonical transformations (41), while in the new variables the whole portrait at the resonance looks very symmetric around the axis $I_\psi^{(r)} = 0$, for any $r \geq 2$.

Resonant construction for librational motions

An interesting remark, at this point, is that the resonant construction described so far can be applied in the same way in the case of the main separatrix domain of the pendulum, which corresponds to librational motions. In fact, this can be considered as a case of resonance around the action value $p_* = 0$, or $\omega_* = 0$. This corresponds to the resonant wave vector $(k^{(1)}, k^{(2)}) = (1, 0)$, yielding a normal vector $\mathbf{m} = (0, 1)$. We then readily find that the resonant module is simply $\mathcal{M} = \{\mathbf{k} : k_2 = 0\}$. This means to keep in the normal form all trigonometric terms of the form $e^{ik_1\psi}$, i.e. we eliminate all terms containing the angle ϕ. Also, using Eq.(70), in this case we find $I_R = I_\psi$, and $\phi_R = \psi$. The reader can easily fill in now the details for the computation of the corresponding resonant normal form associated with librational motions. The final result, up to fourth order, apart from a constant, and after setting $\lambda = 1$ is:

$$Z_{lib}^{(4)} = I + 5.12 \times 10^{-6} I_\psi + (1 + 7.68 \times 10^{-6}) \frac{I_\psi^2}{2}$$

$$-0.0799999 \cos\psi - (1.28 + 2.56 I_\psi + 1.28 I_\psi^2) \times 10^{-6} \cos(2\psi) + 1.024 \times 10^{-7} \cos(3\psi) .$$

A comparison of the theoretical (with the above formula) versus numerical phase portraits in the libration domain is left as an exercise.

2.10. Hyperbolic normal form and the computation of invariant manifolds

So far, we have explored the possibilities offered by canonical perturbation theory in order to find useful representations for regular orbits in various cases of interest related either to non-resonant or to resonant dynamics. In subsection 2.2, however, it was mentioned that normal form theory can be useful even in representing some features of *chaotic dynamics*, and, in particular, in the computation of the invariant manifolds emanating from unstable periodic orbits in a chaotic domain like the extended chaotic domain of Fig.1b. We now turn our attention to this subject, which, as discussed in the introduction, has lead to a number of quite interesting modern applications.

We recall first the notion of an unstable periodic orbit, and of its asymptotic invariant manifolds. Let $(\psi(t), \phi(t), I_\psi(t), I(t))$ be an orbit of the system (3). This is called *periodic* if there is a time $t = T$ (the period) at which the orbit

closes in itself, i.e.

$$\Big(\psi(t+T), \phi(t+T), I_\psi(t+T), I(t+T)\Big) = \Big(\psi(t)(mod2\pi), \phi(t)(mod2\pi), I_\psi(t), I(t)\Big) \ .$$

Since in our example the angle ϕ evolves linearly with frequency equal to $\omega = 1$, it follows that the only possible periods for periodic orbits in the system (3) are $T = 2\pi$ or multiples of it. We now focus on a particular periodic orbit, denoted P, whose limit, for $\epsilon = 0$, is the unstable equilibrium point at $\psi_P = \pi$ (or $-\pi$), and $p_P = 0$. This equilibrium point continues as a periodic orbit, for values of ϵ different from zero.

Fixing a value of ϵ (say $\epsilon = 1$, as in Fig.1b), how can we locate the initial condition for this periodic orbit on the surface of section of Fig.1b? We recall that such a surface of section can be thought of as providing a mapping of the form (6), where the functions $f(\psi, p)$ and $g(\psi, p)$ are provided just by the numerical computation of the image (ψ', p'), on the surface of section, of some initial point (ψ, p). Now, assume that the point (ψ, p) is the initial condition of a periodic orbit. Then, at return on the section after one period we have $\psi' = \psi$, $p' = p$, or

$$F(\psi, p) = f(\psi, p) - \psi = 0, \qquad G(\psi, p) = g(\psi, p) - p = 0 \ . \tag{75}$$

We now see that the initial condition of the periodic orbit is the solution to a 2×2 set of algebraic equations, namely $F(\psi, p) = 0$, $G(\psi, p) = 0$. In the absence of a trivial solution, we can rely on a numerical method for the solution of this set of equations, as, for example, Newton's method.[24]

---

[24]Newton's method is an iterative method for finding the roots of a set of algebraic equations. In this method, we start with an initial guess (in our case (ψ_0, p_0) on the surface of section), and compute iteratively better approximations (ψ_1, p_1), (ψ_2, p_2), ... to a root of the system of equations $F(\psi, p) = 0$, $G(\psi, p) = 0$. Let (ψ_n, p_n) be the n-th iterate of the method. We compute the next iterate by:

$$\begin{pmatrix} \psi_{n+1} \\ p_{n+1} \end{pmatrix} = \begin{pmatrix} \psi_n \\ p_n \end{pmatrix} + \begin{pmatrix} \frac{\partial F}{\partial \psi_n} & \frac{\partial F}{\partial p_n} \\ \frac{\partial G}{\partial \psi_n} & \frac{\partial G}{\partial p_n} \end{pmatrix}^{-1} \cdot \begin{pmatrix} F(\psi_n, p_n) \\ G(\psi_n, p_n) \end{pmatrix} \ . \tag{76}$$

In practice, we need a numerical procedure for finding the values of both $F(\psi_n, p_n)$, $G(\psi_n, p_n)$, and all their partial derivatives. This can be as follows: given a point (ψ_n, p_n) on the surface of section, we compute its image (ψ'_n, p'_n) under the Poincaré surface of section mapping. This yields numerical values for the functions $f(\psi_n, p_n) = \psi'_n$ and $g(\psi_n, p_n) = p'_n$, and hence, $F(\psi_n, p_n) = f(\psi_n, p_n) - \psi_n$, $G(\psi_n, p_n) = g(\psi_n, p_n) - p_n$. In the same way, we compute the surface of section images of four neighboring initial conditions, i.e. $(\psi_n + h, p_n)$, $(\psi_n - h, p_n)$, $(\psi_n, p_n + h)$, $(\psi_n, p_n - h)$ for some small number h. This yields numerical values for the functions f and g with all four combinations of arguments. Then we can approximate partial derivatives e.g. as:

$$\frac{\partial f}{\partial \psi_n} = \frac{f(\psi_n + h, p_n) - f(\psi_n - h, p_n)}{2h} + O(h^2) \ ,$$

and similarly for $\partial f/\partial p_n$, $\partial g/\partial \psi_n$, $\partial g/\partial p_n$. Steps of order $h = 10^{-5}$ to $h = 10^{-7}$ give optimal results. The error in the computation can always be controlled by computing the determinant $(\partial f/\partial \psi_n)(\partial g/\partial p_n) - (\partial f/\partial p_n)(\partial g/\partial \psi_n)$ which is theoretically equal to 1. Finally, we form the

Implementing Newton's method in our system (3), for $\epsilon = 1$, we find that the periodic orbit is located at $\psi_P = \pi$, $p_P = -0.0073246492486565$, with an uncertainty in the last digit.

In order to check the *stability* of this orbit, we compute the so-called monodromy (or Floquet) matrix, which is the matrix of the linearized mapping equations computed at the point (ψ_P, p_P), namely

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial \psi} & \frac{\partial f}{\partial p} \\ \frac{\partial g}{\partial \psi} & \frac{\partial g}{\partial p} \end{pmatrix}\Bigg|_{\psi=\psi_P, p=p_P} . \tag{77}$$

The stability is characterized by the eigenvalues of the monodromy matrix, given by the solution of the characteristic polynomial $\Lambda^2 - (a + d)\Lambda + (ad - bc) = 0$. By the symplectic condition, we have $ad - bc = 1$, thus

$$\Lambda_{1,2} = \frac{(a + d) \pm \sqrt{(a + d)^2 - 4}}{2} . \tag{78}$$

In the case of unstable periodic orbits, we have two real and reciprocal eigenvalues $\Lambda_1 \Lambda_2 = 1$. The condition for instability, thus, is given by $|a + d| > 2$. Without loss of generality, assume $|\Lambda_1| > 1$ and $|\Lambda_2| < 1$. The eigenvector of M corresponding to the eigenvalue L_1 defines an *unstable eigen-direction* of the linearized mapping around (ψ_P, p_P), while the eigenvector corresponding to Λ_2 defines a *stable* eigen-direction respectively. These directions are easily found by taking the definition of an eigenvector, namely a vector which, acted upon by M, yields a multiple of itself by a factor equal to the eigenvalue. Implementing this definition, if $(\Delta\psi, \Delta p)$ denotes, say, the unstable eigenvector, we have:

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \Delta\psi \\ \Delta p \end{pmatrix} = \Lambda_1 \begin{pmatrix} \Delta\psi \\ \Delta p \end{pmatrix} .$$

The above condition yields two linearly dependent equations, thus we may chose any of them to define the direction of the unstable eigenvector, i.e.

$$\frac{\Delta p}{\Delta\psi} = \frac{\Lambda_1 - a}{b} = \frac{c}{\Lambda_1 - d} . \tag{79}$$

Consider now a small segment of length ΔS on the surface of section along, say, the unstable eigen-direction, starting from the periodic orbit P (Figure 8), and compute the successive images of this segment under the surface of section mapping. By this process we obtain numerically the intersection (with the surface of section) of the *unstable manifold* of the periodic orbit P. In Figs.8a,b, the

---

matrix

$$\begin{pmatrix} \frac{\partial F}{\partial \psi_n} & \frac{\partial F}{\partial p_n} \\ \frac{\partial G}{\partial \psi_n} & \frac{\partial G}{\partial p_n} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial \psi_n} & \frac{\partial f}{\partial p_n} \\ \frac{\partial g}{\partial \psi_n} & \frac{\partial g}{\partial p_n} \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} .$$

The method is successful if the sequence of successive iterates $(\psi_n, p_n) \to (\psi_{n+1}, p_{n+1})$ converges. Numerically, we stop when differences between successive iterates become smaller than a prescribed precision level (which can be very small, say 10^{-14}).

unstable manifold (denoted by W_u) is shown as a thin curve starting from the left point P (which is the same as the right point, modulo 2π). This curve has the form of a straight line close to P, but exhibits a number of oscillations as it approaches the right point P. In the same way, we can compute the stable manifold W_s emanating from the right point P, also shown by a thin curve (in our example the two curves are symmetric with respect to the axis $\psi = 0$, due to a corresponding symmetry in the Hamiltonian (3)). In the case of the stable manifold, we integrate backwards in time all initial conditions on the initial segment ΔS, until finding the pre-image of this segment on the surface of section. We observe that the stable manifold also develops oscillations as it approaches the left point P.

The following are now some more precise definitions. Let

$$\mathcal{D}_P = \left\{ \Big(\psi_P(t), \phi_P(t), I_{\psi,P}(t), I_P(t) \Big), 0 \le t \le T \right\}$$

be the set of all points of the periodic orbit P parametrized by the time t, and $\mathbf{q} = (\psi, \phi, I_\psi, I)$ a randomly chosen point in phase space. We define the distance of an arbitrary point $\mathbf{q}$ of the phase space from the periodic orbit as:

$$d(\mathbf{q}, P) = \min \left\{ dist(\mathbf{q}, \mathbf{q}_P) \text{ for all } \mathbf{q}_P \in \mathcal{D}_P \right\}$$

where $dist()$ means the Euclidean distance. Let now $\mathbf{q}_0$ be a particular initial condition at $t = 0$, and let $\mathbf{q}(t; \mathbf{q}_0)$ denote the orbit resulting from that initial condition. The unstable manifold of P is defined as:

$$W_u^P = \left\{ \mathbf{q}_0 : \lim_{t \to -\infty} d(\mathbf{q}(t; \mathbf{q}_0), P) = 0 \right\} \quad . \tag{80}$$

In words, the unstable manifold W_u^P is the set of all initial conditions in the phase space leading to orbits which tend asymptotically to the periodic orbit P when integrated in the backward sense of time. This, in fact, implies that these orbits recede (on average) from the periodic orbit in the forward sense of time.

Similarly, we define the stable manifold of P by:

$$W_s^P = \left\{ \mathbf{q}_0 : \lim_{t \to \infty} d(\mathbf{q}(t; \mathbf{q}_0), P) = 0 \right\} \quad . \tag{81}$$

In words, it is the set of all initial conditions whose resulting orbits tend asymptotically to the periodic orbit P in the forward sense of time.

The following are some basic properties of W_u and W_s:

i) Both sets are *invariant* under the Hamiltonian flow, i.e. any initial condition on W_u^P leads to an orbit lying always on W_u^P (and similarly for W_s^P).

ii) The unstable (or stable) invariant manifold cannot intersect itself at any of its points. In fact, a simple argument shows that such intersections are inconsistent with causality.

iii) In contrast to (ii), the unstable and stable manifolds may intersect each other. Such intersections are called *homoclinic points*, and their resulting orbits are called homoclinic orbits. Such is the example of the homoclinic point H in figure 8 (see below).

iv) As a consequence of the so-called Grobman (1959) and Hartman (1960) theorem, the unstable manifold W_u^P is tangent to the unstable manifold of the linearized flow around P. In fact, it is this property which allows us to approximate numerically the computation of the invariant manifolds, by taking the successive images of small straight segments along one of the eigendirections close to P.

The study of the asymptotic manifolds of unstable periodic orbits in a chaotic domain is a key element for understanding the structure and mechanisms of chaos in dynamical systems. This key role was already emphasized by Poincaré, who, on the basis of theoretical arguments, predicted the complexity induced in the behavior of chaotic orbits in the phase space due to the manifold dynamics. Various laws characterizing, in particular, the *lobes* and *recurrences* of the invariant manifolds are given in Contopoulos and Polymilis (1993).

We now show that, despite their complexity, the invariant manifolds are objects possible to compute, up to some extent, via an appropriate form of canonical perturbation theory. The resulting normal form, called a *hyperbolic normal form*, was first studied by Moser (1958), while its implementation via canonical transformations was first discussed by Giorgilli (2001).

We give below a specific example of computation of a hyperbolic normal form, and discuss the kind of phenomena whose study becomes possible by using such a normal form.

The idea of a hyperbolic normal form is simple: close to any unstable periodic orbit, we wish to pass from old to new canonical variables $(\psi, \phi, p, I) \rightarrow (\xi, \phi', \eta, I')$, so that the Hamiltonian in the new variables takes locally the form:

$$Z_h = \omega I' + \nu \xi \eta + Z(I', \xi \eta) \qquad (82)$$

where ν is a real constant. In a Hamiltonian like (82), the point $\xi = \eta = 0$ corresponds to a periodic orbit, since, from Hamilton's equation we find $\dot{\xi} = \dot{\eta} = 0 = \dot{I}' = 0$, while $\phi' = \phi'_0 + (\omega + \partial Z(I', 0) \partial I')t$. This implies a periodic orbit, with frequency $\omega' = (\omega + \partial Z(I', 0) \partial I')$. In a system like (3), where the action I is dummy, we find that I' does not appear in the hyperbolic normal form, thus, the periodic solution $\xi = \eta = 0$ has a frequency equal to ω always.

By linearizing Hamilton's equations of motion near this solution, we find that it is always unstable. In fact, we can easily show that the linearized equations of motion for small variations $\delta\xi, \delta\eta$ around $\xi = 0, \eta = 0$ are

$$\dot{\delta\xi} = (\nu + \nu_1(I'))\delta\xi, \quad \dot{\delta\eta} = -(\nu + \nu_1(I'))\delta\eta$$

where $\nu_1(I') = \partial Z(I', \xi\eta = 0)/\partial(\xi\eta)$. The solutions are $\delta\xi(t) = \delta\xi_0 e^{(\nu+\nu_1)t}$, $\delta\eta(t) = \delta\eta_0 e^{-(\nu+\nu_1)t}$. After one period $T = 2\pi/\omega$ we have $\delta\xi(T) = \Lambda_1 \delta\xi_0$, $\delta\eta(T) = \Lambda_2 \delta\xi_0$, where $\Lambda_{1,2} = e^{\pm 2\pi(\nu+\nu_1)/\omega}$. Thus, the two eigendirections of the linearized flow correspond to setting $\delta\xi_0 = 0$, or $\delta\eta_0 = 0$, i.e. they coincide with the axes $\xi = 0$, or $\eta = 0$. These axes are invariant under the flow of (82) and, therefore, *they constitute the unstable and stable manifold of the associated periodic orbit* P.

In summary: if we succeed in finding a canonical transformation of the form

$$
\begin{aligned}
\psi &= \Phi_\psi(\xi, \phi', \eta, I') \\
\phi &= \Phi_\phi(\xi, \phi', \eta, I') \\
p &= \Phi_p(\xi, \phi', \eta, I') \\
I &= \Phi_I(\xi, \phi', \eta, I')
\end{aligned}
\tag{83}
$$

in which the Hamiltonian in the neighborhood of an unstable periodic orbit takes the form (82), then, the expressions i) $\xi = \eta = 0$, and ii) $\xi = 0$ or $\eta = 0$ yield the position of the periodic orbit, and ii) the form of its invariant manifolds in the variables (ξ, ϕ', η, I'). Furthermore, iii) the expressions $\Lambda_{1,2} = e^{\pm 2\pi(\nu+\nu_1)/\omega}$ yield the eigenvalues of the unstable periodic orbit.
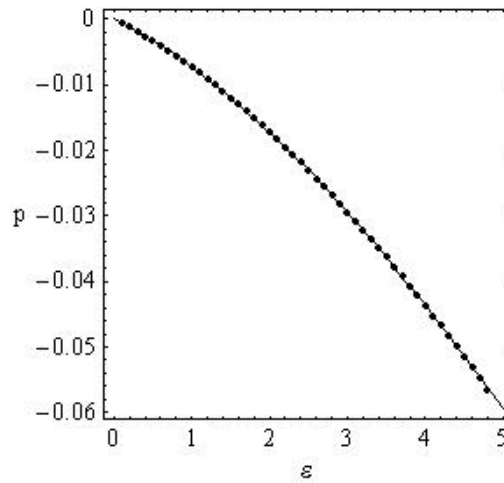
Figure 7. The characteristic curve (value of the fixed point variable p_P on the surface of section) for the main unstable periodic orbit as a function of ϵ. The dots correspond to a purely numerical calculation using Newton's method. The solid curve shows the theoretical calculation using a hyperbolic normal form (similar to formula (96) in text, but for a normalization up to the fifteenth order).

A key point is, now, the following: after we explicitly compute the canonical transformations (83), we can use Eqs.(83) in order to *compute analytically the periodic orbit and its asymptotic invariant manifolds in the original canonical variables as well.* Such a computation, whose details will be presented below, is shown in figures 7 and 8. Figure 7 shows the so-called *characteristic curve* of the main unstable periodic orbit of the system (3), i.e. the orbit which is the continuation, for $\epsilon \neq 0$, of the unstable equilibrium point $(\psi = \pm\pi, p = 0)$ which exists for $\epsilon = 0$. The characteristic curve is a curve yielding the value of the initial conditions on a surface of section, as a function of ϵ, for which the resulting orbit is the periodic one. In our case, we always have $\psi_P = 0$, while p_P varies with ϵ. The dotted curve shows $p_P(\epsilon)$ as computed by a purely numerical process, i.e., implementing Newton's root-finding method, while the solid curve yields $p_P(\epsilon)$ as computed by a hyperbolic normal form at the normalization order

$r = 15$ (see below). The agreement is excellent, and we always recover 8-9 digits of the numerical calculation in this approximation, even for values of ϵ much larger than unity.

How is this computation possible? Returning to Eqs.(83), we simply set $\xi = \eta = 0$, while I', which is an integral of the Hamiltonian flow of (82), can be replaced by a constant label value $I' = c$. [25] Finally, knowing the frequency ω' by which ϕ' evolves, we can set $\phi' = \omega' t + \phi'_0$. Substituting these expressions in the transformation equations (83), we are lead to:

$$
\begin{array}{rcl}
\psi_P(t) & = & \Phi_\psi(0, \omega' t + \phi'_0, 0, c) \\
\phi_P(t) & = & \Phi_\phi(0, \omega' t + \phi'_0, 0, c) \\
p_P(t) & = & \Phi_p(0, \omega' t + \phi'_0, 0, c) \\
I_P(t) & = & \Phi_I(0, \omega' t + \phi'_0, 0, c) \quad .
\end{array}
\tag{84}
$$

The set of Eqs.(84) yields now an *analytic representation of the periodic orbit* P in the whole time interval $0 \le t \le 2\pi/\omega'$. In fact, by the form of the hyperbolic normal form and its normalizing transformations, we find that Eqs.(84) provide a formula for the periodic orbit in terms of a Fourier series, which allows us to find not only its initial conditions on a surface of section, but also the whole time evolution of the set of canonical variables along P.

Similar principles apply to the computation of the invariant manifolds of P. In this case, we first fix a surface of section by setting, say, $\phi' = 0$. Then, assuming, without loss of generality, that the unstable manifold corresponds to setting $\eta = 0$, we find a parametric form for all canonical variables as a function of ξ along the asymptotic curve of the unstable manifold on the surface of section:

$$
\psi_{P,u}(\xi) = \Phi_\psi(\xi, 0, 0, c), \quad p_{P,u}(\xi) = \Phi_p(\xi, 0, 0, c) \quad .
\tag{85}
$$

Due to Eq.(85), ξ can be considered as a *length parameter* along the asymptotic curve of the unstable manifold W_u. Numerically, this allows to compute the asymptotic curve W_u on the surface of section by giving different values to ξ. Such a computation is shown with a thick curve in Fig.8a. We observe that the theoretical curve W_u agrees well with the numerical one up to a certain distance corresponding to $\xi \sim 1$, whereby the theoretical curve starts deviating from the true asymptotic curve W_u. This is because, as we will see, the hyperbolic normal form has a *finite domain of convergence* around P. Thus, by using a finite truncation of the series (83) (representing the normalizing canonical transformations), deviations occur at points beyond the domain of convergence of the hyperbolic normal form.

Similar arguments (and results, as shown in Fig.8) are found for the stable manifold of P. In that case, we substitute $\xi = 0$ in the transformation equations, and employ η as a parameter, namely:

$$
\psi_{P,s}(\eta) = \Phi_\psi(0, 0, \eta, c), \quad p_{P,s}(\eta) = \Phi_p(0, 0, \eta, c) \quad .
\tag{86}
$$

---

[25] In fact, if I is a dummy action, its near identity transformation I' does not appear in the Eqs.(83).
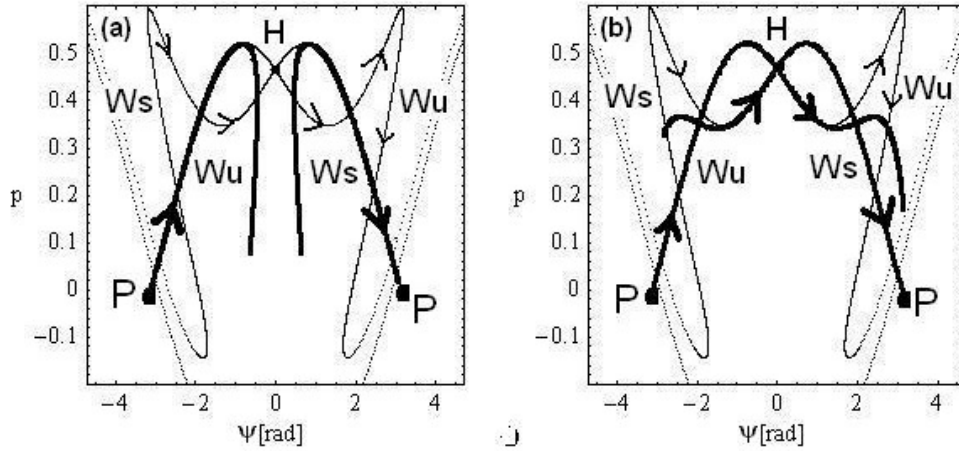
Figure 8.   The thin dotted lines show the unstable (W_u) and stable ((W_s) manifolds emanating from the main unstable periodic orbit (P) in the model (3), for $\epsilon = 1$, after a purely numerical computation (mapping for 8 iterations of 1000 points along an initial segment of length $ds = 10^{-3}$ taken along the unstable and stable eigen-directions respectively. In (a), the thick lines show a theoretical computation of the invariant manifolds using a hyperbolic normal form at the normalization order $r = 15$ (see text). Both theoretical curves W_u and W_s deviate from the true manifolds before reaching the first homoclinic point (H). (b) Same as in (a), but now the theoretical manifolds are computed using the analytic continuation technique suggested in Ozorio de Almeida and Viera (1997). The theoretical curves cross each other at the first homoclinic point, thus, this point can be computed by series expansions.

In figure 8 we see that the domains of convergence of the hyperbolic normal form are small enough so that the two theoretical curves W_u and W_s have no intersection. This implies that we cannot use this computation in order to specify analytically the position of a *homoclinic* point, like H in figure 8. However, in recent years, Ozorio de Almeida and co-workers (Da Silva Ritter et al. (1987), Viera and Ozorio de Almeida (1996), Ozorio de Almeida and Viera (1997)) have considered an extension of the original theory of Moser, which allows to extend considerably the domain of validity of the hyperbolic normal form construction, and to compute homoclinic intersections and even some lobes formed by the asymptotic manifolds in the vicinity of P. Essentially, the technique relies on i) using first the usual construction in order to compute a finite segment of, say, W_u within the domain of convergence of the hyperbolic normal form, and ii) analytically continue this part up to one or more images of it, using *the original Hamiltonian as a Lie generating function*. That is, if **q** is a point computed on the invariant manifold, we compute its image via

$$\mathbf{q}' = \exp(t_n L_H) \exp(t_{n-1} L_H) \dots \exp(t_1 L_H)\mathbf{q} \qquad (87)$$

where $t_n + t_{n-1} + \dots + t_1 = T$, while the times are chosen so as to always lead to a mapping within the analyticity domain of the corresponding Lie series in a complex time domain. Figure 8b shows such a result. Computing a Lie series via (87) with $t_1 = t_2 = t_3 = t_4 = \pi/2$, the thick lines show the theoretical

computation of the images (for the unstable manifold), or pre-images (for the stable manifold), for which we put a minus sign in front of all times t_1 to t_4) of the thick lines shown in panel (a), after (or before) one period. We now see that the resulting series represent the true invariant manifolds over a considerably larger extent, thus, allowing to compute theoretically the position of the homoclinic point H. [26]

We now present in detail the steps leading to the previous results, i.e. a practical example of the calculation of a hyperbolic normal form.

i) *Hamiltonian expansion.* Starting from the Hamiltonian (3), in the neighborhood of P (see phase portraits in Fig.1) it is convenient to expand the Hamiltonian around the value $\psi_0 = \pi$ (or, equivalently, $-\pi$), which corresponds to the position of the unstable equilibrium when $\epsilon = 0$. Setting $\psi = \pi + u$, we then find (up to fourth order): [27]

$$H = \frac{p^2}{2} + I - 0.08 \left(1 + 0.5\epsilon(1+p)(e^{i\phi} + e^{-i\phi}) \right) \left(-1 + \frac{u^2}{2} - \frac{u^4}{24} - ... \right) \quad (88)$$

The hyperbolic character of motion in the neighborhood of the unstable equilibrium is manifested by the combination of terms:

$$H = I + \frac{p^2}{2} - 0.08\frac{u^2}{2} + ... \quad (89)$$

The constant ν appearing in Eq.(82) is related to the constant 0.08 appearing in Eq.(89) via $\nu^2 = 0.08$. In fact, if we write the hyperbolic part of the Hamiltonian as $H_h = p^2/2 - \nu^2 u^2/2$, it is possible to bring H_h in hyperbolic normal form, by introducing a linear canonical transformation:

$$p = \frac{\sqrt{\nu}(\xi + \eta)}{\sqrt{2}}, \quad u = \frac{(\xi - \eta)}{\sqrt{2\nu}} \quad (90)$$

where ξ and η are the new canonical position and momentum respectively. Then H_h acquires the desired form, i.e. $H_h = \nu\xi\eta$.

Substituting the transformation (90) in the Hamiltonian (88) we find

$$H = I + 0.282843\xi\eta - 0.041667\xi\eta^3 + 0.0625\xi^2\eta^2 - 0.041667\xi^3\eta + 0.010417\xi^4$$

$$+\epsilon \left[0.08 + 0.030085\eta - 0.070711\eta^2 - 0.026591\eta^3 + 0.010417\eta^4 \right.$$

---

[26]The reader may wonder why there should be interest in knowing the position of homoclinic points like H. The role of homoclinic points, as well as of their resulting (as initial conditions) homoclinic orbits, in the problem of the so-called *structure of chaos* in Hamiltonian systems cannot be overemphasized. In fact, the homoclinic points form the basis upon which we understand the structure of the *homoclinic tangle*, which is the main local source of chaos close to unstable periodic orbits (see Contopoulos (2002), for a review of the concept and consequences of homoclinic chaos).

[27]In this example we keep ϵ as a parameter, without substituting its numerical value. This is needed in the calculation of the position of the periodic orbit P as a function of ϵ, used in Fig.7.

$$+0.030085\xi + 0.14142\xi\eta + 0.0265915\xi\eta^2 - 0.041667\xi\eta^3$$
$$-0.070711\xi^2 + 0.026591\xi^2\eta + 0.0625\xi^2\eta^2 - 0.026591\xi^3$$
$$-0.041667\xi^3\eta + 0.010417\xi^4 + ...\Bigg] \left(\frac{e^{i\phi} + e^{-i\phi}}{2}\right) \quad .$$

As always, we now have to introduce some book-keeping. In the present case, it is crucial to recognize that the quantities ξ, η themselves can be considered as small quantities describing the neighborhood of a hyperbolic point. However, we want to retain a book-keeping factor λ^0 for the lowest order term $\xi\eta$, since, as we will see, this term appears in the homological equation defining the Lie generating functions in the present case. We thus follow the rule that monomial terms containing a product $\xi^{s_1}\eta^{s_2}$ acquire a book-keeping factor $\lambda^{s_1+s_2-2}$ in front. [28] Finally, we add a book-keeping factor λ to all the terms that are multiplied by ϵ.

In summary, up to $O(\lambda^2)$ the Hamiltonian before any normalization reads:

$$H^{(0)} = I + 0.282843\xi\eta$$

$$+\lambda\epsilon\Bigg[0.04 + 0.0150424(\xi + \eta) - 0.0353553(\xi^2 + \eta^2) + 0.0707107\xi\eta\Bigg](e^{i\phi} + e^{-i\phi})$$

$$+\lambda^2\Bigg[0.0104167(\xi^4 + \eta^4) - 0.0416667(\xi\eta^3 + \xi^3\eta) + 0.0625\xi^2\eta^2$$

$$+0.0132957\epsilon(\xi^2\eta + \xi\eta^2 - \xi^3 - \eta^3)(e^{i\phi} + e^{-i\phi})\Bigg] + \ldots$$

According to the definition of the hyperbolic normal form (Eq.(82)), at first order we want to eliminate i) terms depending on the angle ϕ, or, ii) terms independent of ϕ but depending on a product $\xi^{s_1}\eta^{s_2}$ with $s_1 \neq s_2$. These are

$$h_1^{(0)} = \epsilon\Bigg[0.04 + 0.0150424(\xi + \eta) - 0.0353553(\xi^2 + \eta^2) + 0.0707107\xi\eta\Bigg](e^{i\phi} + e^{-i\phi}) \, .$$

---

[28]This, by the way, is precisely the optimal book-keeping scheme in computing normal forms for polynomial Hamiltonians in general. In the case of elliptic equilibria, in particular, assuming that the lowest order term in the Hamiltonian is the harmonic oscillator model

$$H = \frac{p^2}{2} + \frac{\omega_0^2\psi^2}{2}$$

we introduce complex canonical variables via a linear canonical transformation quite similar to (90), namely

$$p = \frac{\sqrt{\omega_0}(iQ + P)}{\sqrt{2}}, \quad \psi = \frac{(Q + iP)}{\sqrt{2\omega_0}} \quad . \tag{91}$$

The zeroth order Hamiltonian takes the form $H = i\omega_0 QP$. Then, we normalize all monomial terms of the form $Q^{s_1}P^{s_2}$ giving rise to non-zero divisors. This is the way to normalize, for example, the Hamiltonian of the Hénon-Heiles (1964) model. The variables Q, P have a particular importance in quantum mechanics, where Q is called the creation operator, while iP is the annihilation operator.

In order to define the generating function χ_1 performing the first order normalization, we now see that we need a homological equation different from what we have seen so far, namely:

$$\{I + 0.282843\xi\eta, \chi_1\} + \lambda h^{(0)} = 0 \ . \tag{92}$$

The solution of equation (92) is found by noticing that the action of the operator $\{\omega I + \nu\xi\eta, \cdot\}$ on monomials of the form $\xi^{s_1}\eta^{s_2}a(I)e^{ik_2\phi}$ yields

$$\left\{\omega I + \nu\xi\eta, \xi^{s_1}\eta^{s_2}a(I)e^{ik_2\phi}\right\} = -[(s_1 - s_2)\nu + i\omega k_2]\xi^{s_1}\eta^{s_2}a(I)e^{ik_2\phi} \ .$$

Thus, if we write $h_1^{(0)}$ as

$$h_1^{(0)} = \sum_{(s_1,s_2,k_2)\notin\mathcal{M}} b_{s_1,s_2,k_2}(I)\xi^{s_1}\eta^{s_2}e^{ik_2\phi}$$

where the resonant module in this case is defined as:

$$\mathcal{M} = \{(s_1, s_2, k_2) : s_1 = s_2 \text{ and } k_2 = 0\} \tag{93}$$

then, the solution of the homological equation (92) is

$$\chi_1 = \sum_{(s_1,s_2,k_2)\notin\mathcal{M}} \frac{b_{s_1,s_2,k_2}(I)}{(s_1 - s_2)\nu + i\omega k_2}\xi^{s_1}\eta^{s_2}e^{ik_2\phi} \ . \tag{94}$$

In fact, since I is a dummy action, it does not appear in any of the coefficients b_{s_1,s_2,k_2}. But the main observation, regarding Eq.(94), is that *the divisors are complex numbers with a modulus bounded from below by a positive constant*, i.e. we have:

$$\begin{aligned}
|\nu(s_1 - s_2) + ik_2\omega| \ &= \ \sqrt{(s_1 - s_2)^2\nu^2 + k_2^2\omega^2} \geq \min(|\nu|, |\omega|) \\
&\text{for all } (s_1, s_2, k_2) \notin \mathcal{M} \ .
\end{aligned} \tag{95}$$

This last bound constitutes the most relevant fact about the construction of hyperbolic normal forms, because it implies that this construction is convergent. In fact, the sequences of repetitions of divisors encountered in subsection 2.5 appear here also in a quite similar manner, as the reader may readily verify by trying to reconstruct the 'most dangerous' sequence of terms produced by successive Poisson brackets. However, the fact that divisors are bounded from below by a positive constant implies that the radius of convergence remains finite as the normalization order tends to infinity.

Returning to the numerical example, after performing the computations in (94) the generating function χ_1 reads:

$$\chi_1 = \lambda\epsilon i\left[\left(-0.04 + (0.00393948 - 0.0139282i)\xi - (0.00393948 + 0.0139282i)\eta\right.\right.$$

$$-(0.0151515 - 0.0267843i)\xi^2 + (0.0151515 + 0.0267843i)\eta^2 - 0.070711\xi\eta\Bigg)e^{i\phi}$$

$$+\Bigg(0.04 + (0.00393948 + 0.0139282i)\xi - (0.00393948 - 0.0139282i)\eta$$

$$-(0.0151515 + 0.0267843i)\xi^2 + (0.0151515 - 0.0267843i)\eta^2 + 0.070711\xi\eta\Bigg)e^{-i\phi}\Bigg]$$

The normalized Hamiltonian, after computing $H^{(1)} = \exp(L_{\chi_1})H^{(0)}$ is in normal form up to terms of $O(\lambda)$. In fact, we find that there are no new normal form terms at this order, but such terms appear at order λ^2. The reader is invited to make the computation at order 2, which finally yields $H^{(2)} = \exp(L_{\chi_2})H^{(1)}$, in normal form up to order two. We give the final result for verification. Apart from a constant, we have

$$H^{(2)} = I + 0.282843\eta\xi + \lambda^2(0.0625\xi^2\eta^2 - \epsilon^2 0.0042855\xi\eta) + O(\lambda^3) + \dots$$

Let us also give the analytic expressions for the periodic orbit, up to order $O(\lambda^2)$, found by exploiting the normalizing transformations of the hyperbolic normal form. The old canonical variables (ξ, η) are computed in terms of the new canonical variables $(\xi^{(2)}, \eta^{(2)})$ following:

$$\xi = \exp(L_{\chi_2})\exp(L_{\chi_1})\xi^{(2)}$$

$$\eta = \exp(L_{\chi_2})\exp(L_{\chi_1})\eta^{(2)} \quad .$$

This yields functions (up to order $O(\lambda^2)$ $\xi = \Phi_\xi(\xi^{(2)}, \phi^{(2)}, \eta^{(2)})$, and $\eta = \Phi_\eta(\xi^{(2)}, \phi^{(2)}, \eta^{(2)})$. By virtue of the fact that I is a dummy action, we have $\phi^{(2)} = \phi = \omega t = t$, while, for the periodic orbit we set $\xi^{(2)} = \eta^{(2)} = 0$. With these substitutions, we find

$$\xi_P(t) = \Phi_\xi(0, t, 0), \quad \eta_P(t) = \Phi_\eta(0, t, 0) \quad .$$

Finally, we substitute the expressions for $\xi_P(t)$ and $\eta_P(t)$ in the linear canonical transformation (90), in order to find analytic expressions for the periodic orbit in the original variables $p, \psi = \pi + u$. Switching back to trigonometric functions, and setting $\lambda = 1$, we finally find:

$$\begin{align}
\psi_P(t) &= \pi + 0.0740741\epsilon \sin t - 0.000726216\epsilon^2 \sin(2t) \tag{96}\\
p_P(t) &= -0.00592593\epsilon \cos t - 0.00145243\epsilon^2 \cos(2t) \quad .
\end{align}$$

The position of the periodic orbit on the surface of section can be found now by setting $t = 0$ in Eqs.(96). In the actual computation of figures 7 and 8, we compute all expansions up to $O(\lambda^{15})$, after expanding also $\cos\psi$ in the original Hamiltonian up to the same order.

3.   CONSEQUENCES OF ANALYTICITY

3.1.   Book-keeping in models with infinitely many Fourier harmonics

So far, we have examined some basic methods of canonical perturbation theory by presenting a number of numerical examples where such methods are applicable, employing for this purpose the Hamiltonian model (3). However, this model has a number of important limitations. In the present section we focus on one such limitation whose consequences are necessary to discuss, i.e., the fact that the perturbation term in (3) contains only a *finite number of Fourier harmonics.* Namely, in the original Hamiltonian we only have the terms $e^{\pm i\psi}$, $e^{\pm i(\psi+\phi)}$ and $e^{\pm i(\psi-\phi)}$. It has been discussed already that even with few terms initially, a whole spectrum of new Fourier terms are generated in the course of any of the normalization schemes discussed in section 2. However, in the sequel we are interested in a more general case, in which many, or even *all possible Fourier harmonics are present already in the original Hamiltonian.*

It should be made clear that the above situation occurs in almost all problems of practical interest encountered in dynamical astronomy. Consider, for example, the study of a quite common class of problems in galactic dynamics, referring to galaxies with a non-axisymmetric distribution of matter. Such is, for instance, the case of the study of motions in the galactic plane of a spiral galaxy. Using polar coordinates (r, θ) on the plane, the gravitational potential $V(r, \theta)$ can be written in the form of a sum of an axisymmetric term $V_0(r)$, which accounts for the gravitational effects of the disc, and a non-axisymmetric perturbation $\epsilon V_1(r, \theta)$ yielding the gravitational effects of the spiral arms (and/or a bar). The term V_1, in turn, can be expressed in terms of its Fourier decomposition (see subsection 3.5 for details):

$$V_1(r, \theta) = \sum_{k=-\infty}^{\infty} V_k(r)e^{ik\theta} \quad .$$

As discussed in subsection 3.5, the Fourier coefficients $V_k(r)$ can be analyzed in terms of the so-called 'radial' (or epicyclic) action angle variables (I_r, θ_r) as

$$V_k(r) = \sum_{k_r=-\infty}^{\infty} \tilde{V}_{k_r,k}(I_r)e^{ik_r\theta_r} \quad .$$

Then, the gravitational potential of a disc spiral galaxy is a sum of Fourier terms over all possible wavenumbers k_r, k, a fact implying that the original Hamiltonian describing stellar motions in the galaxy contains the whole possible spectrum of Fourier harmonics $\exp(i(k_r\theta_r + k\theta))$.

It is reasonable to expect that the Fourier coefficients of the higher order Fourier terms (i.e. the ones for which $|k_r| + |k|$ is high) are smaller in size. This question, of how the size of the Fourier terms decreases with Fourier order, plays an essential role in the development of almost all perturbative schemes dealing with problems of dynamical astronomy, because it crucially affects the way by which we can choose to implement *book-keeping* in the Hamiltonian normalization. To show this, consider a generic form of Hamiltonian given by

$$H = H_0 + \epsilon(\dots \text{infinitely many harmonics} \dots) = H_0 + \epsilon H_1 \quad . \qquad (97)$$

The whole term ϵH_1 can be considered as a quantity 'of order ϵ'. If we now choose to book-keep according to powers of ϵ, [29], we should write:

$$H = \lambda^0 H_0 + \lambda^1 \epsilon(\ldots \text{infinitely many harmonics}\ldots) \qquad (98)$$

But in this way, Eq.(98) implies that infinitely many terms have to be normalized already at the first order of perturbation theory.

There is nothing fundamentally inconsistent in the above method of doing book-keeping in a Hamiltonian of the form (97). However there are a number of issues rendering the method rather problematic in practice. Let us mention the most important ones:

i) *Memory requirements:* in the computer we have to introduce a maximum order r_{max} in the book-keeping parameter up to which terms can be stored. Following Eq.(98), r_{max} represents also the maximum order in ϵ up to which terms are stored. However, at every order of ϵ we have terms of all Fourier orders. Thus, we must introduce also a maximum Fourier order K_{max}, up to which terms can be stored. The total memory requirements then depend on the size of the integer lattice (r_{max}, K_{max}). We can now see the following: if (r_0, K_0) denotes a so-called *domain of interest*, i.e. a domain in which we are interested in storing all possible terms arising in the *final* normal form series, then, this is only possible if allowance is made to store terms arising in *intermediate* normalization steps within a domain much larger than (r_0, K_0), namely

$$(r_{max}, K_{max}) = (r_0, K_0 r_0) \quad . \qquad (99)$$

A detailed proof is given in Efthymiopoulos (2008). [30] One can see that in most applications the memory requirements due to Eq.(99) become prohibitive. In fact, we have the estimate

$$\frac{\text{number of terms that need to be stored}}{\text{number of terms in domain of interest}} \approx \left(\frac{r_0 + 1}{2}\right)^{2n}$$

where n is the number of degrees of freedom. Since r_0 is usually a number of order ~ 10, the memory requirements at intermediate normalization steps are orders of magnitude bigger than those for storing the final Hamiltonian.

---

[29]This is the most common suggestion in textbooks regarding the application of canonical perturbation theory, namely, to normalize first terms of order ϵ, then ϵ^2, etc.

[30]In summary: the worst memory requirements are caused by the operation $H^{(1)} = \exp(L_{\chi_1})H^{(0)}$. The generating function χ_1 contains terms of the form $\epsilon a(J)\exp(ik\cdot\phi)$, for some of which we have $|k| = K_0$. The Hamiltonian $H^{(0)}$ contains also terms of the form $\epsilon a'(J)\exp(ik\cdot\phi)$, but for some terms we now have $|k| = K_{max}$. Then, the Poisson bracket operation between such terms yields terms of the form $\epsilon^2[a(J)k\cdot\nabla_J a'(J) - a'(J)k'\cdot\nabla_J a(J)]\exp(i(k+k')\cdot\phi)$. Since the components of k and k' are added algebraically, it is possible that $|k + k'| = K_{max} - K_0$. In the same way, the s-th Poisson bracket contained in $L_{\chi_1}^s H^{(0)}$ can produce Fourier terms of order ϵ^{s+1} and Fourier order $K_{max} - sK_0$. The maximum s is then defined by $s + 1 = r_0$, or $s = r_0 - 1$. But then, we must have $K_{max} - (r_0 - 1)K_0 = K_0$, or $K_{max} = r_0 K_0$.

(Non-)validity of some normal forms in open domains of the action space: Let us assume, again a book-keeping of the form (98), and, furthermore, that the zeroth order term H_0 has a nonlinear dependence on the action variables $\mathbf{I}$. A possible way to perform normal form computations is now the following: writing

$$H_1(\mathbf{I}, \phi) = \sum_{all\ \mathbf{k}} H_{1,\mathbf{k}}(\mathbf{I}) \exp(i\mathbf{k} \cdot \phi)$$

and choosing any type of resonant module $\mathcal{M}$, we define a Lie generating function χ_1 by solving a homological equation similar to Eq.(36), namely:

$$\{H_0, \chi_1\} + h_1^{(0)} = 0 \tag{100}$$

where

$$h_1^{(0)}(\mathbf{I}, \phi) = \sum_{\mathbf{k} \notin \mathcal{M}} H_{1,\mathbf{k}}(\mathbf{I}) \exp(i\mathbf{k} \cdot \phi) \quad .$$

However, the solution of (100)

$$\chi_1 = \sum_{\mathbf{k} \notin \mathcal{M}} \frac{H_{1,\mathbf{k}}(\mathbf{I}) \exp(i\mathbf{k} \cdot \phi)}{i\mathbf{k} \cdot \omega(\mathbf{I})} \tag{101}$$

where

$$\omega(\mathbf{I}) = \nabla_I H_0(\mathbf{I}) \tag{102}$$

is the frequency vector of the unperturbed Hamiltonian at the point $\mathbf{I}$ of the action space, has an important difference from the solution (50) of the homological equation (36). Namely, the denominators of Eq.(101) *depend explicitly on the action variables* $\mathbf{I}$. This fact has deep consequences regarding the definition of domains where such a normal form computation is valid [31]. In fact, provided that the functions $\omega(\mathbf{I})$ have at least a linear dependence on $\mathbf{I}$, we can readily prove the following: if we consider any (small whatsoever) open domain $\mathcal{W}$ in the action space, there are infinitely many wavevectors $\mathbf{k}$ for which the corresponding *resonant manifolds*, i.e. the surfaces defined by relations of the form

$$\mathbf{k} \cdot \omega(\mathbf{I}) = 0 \tag{103}$$

pass through $\mathcal{W}$. This is just a consequence of the fact that the set of all resonant manifolds of the form (103) is a dense set in action space [32]. Then, we conclude that if the coefficients $H_{1,\mathbf{k}}(\mathbf{I})$ in Eq.(101) are different from zero, a

---

[31] In fact, there are also many practical consequences rendering the whole approach of this type quite problematic. The most important problem is that, from the point of view of computer-algebraic operations, it is difficult to handle expressions involving rational functions in the action variables, as required by the form of the solutions given in Eq.(101). In conclusion, a method of normalization like in subsections 2.4 and 2.5, i.e. where we expand H_0 around some fixed values of the actions and push everything beyond linear terms to a higher book-keeping order, is preferable in practice over a method retaining H_0 in its original form.

[32] To give a simple example, let us consider the 'convex' case $H_0 = (I_1^2 + I_2^2)/2$. The frequencies are $\omega_1 = I_1$, $\omega_2 = I_2$, thus the resonant manifolds are just straight lines passing through the origin, defined by $k_1 I_1 + k_2 I_2 = 0$. Consider any point (I_{1*}, I_{2*}) of the action space, with

generating function of the form (101) cannot be defined in any open domain $\mathcal{W}$ of the action space, because some terms of χ_1 (and, similarly, in all subsequent generating functions) would have divisors *becoming equal to zero exactly* at some points within $\mathcal{W}$.

A 'remedy' to this problem, used already by Poincaré (1892), is to introduce a 'cut-off' in Fourier space, i.e. a maximum order K beyond which Fourier terms are not normalized. In this case, however, it is required to make a so-called optimal choice of the value of K, ensuring that the un-normalized terms with $|k| > K$ sum to a contribution smaller in size than the remainder terms produced during the normalization process (see Morbidelli and Giorgilli (1997) for a quantitative analysis of this problem). In fact, basic theory (see, for example, Morbidelli and Guzzo (1997)) yields an estimate $K \sim 1/\epsilon^p$, for some exponent p depending on the exact form of the resonant module $\mathcal{M}$. However, such estimates are not very useful in practice, because in the real computation we would need to know precisely the coefficient of proportionality between K and $1/\epsilon^p$.

It is now possible to see that both problems mentioned above are solved in a natural way by choosing a way of book-keeping different from Eq.(98), recognizing[33] the fact that, in a Hamiltonian like (97), the set of all Fourier harmonics can be split into *subsets of different order of smallness*, each subset containing only a *finite number* of harmonics. This immediately suggests a book-keeping of the form

$$H = H_0 + \epsilon[\lambda^1(\text{group of Fourier terms of first order of smallness}) \qquad (104)$$
$$+\lambda^2(\text{group of Fourier terms of second order of smallness}) + \ldots]$$

In subsection 3.2 we will see how to split the Hamiltonian in groups of Fourier terms of different smallness, taking into account a basic property of analytic Hamiltonian functions, namely the *exponential decay* of Fourier coefficients with increasing Fourier order $|k|$. For the moment, however, let us assume that such a splitting has been accomplished. Under the book-keeping (104), we now observe that only a finite number of harmonics are to be eliminated in every normalization step. For example, instead of the generating function of Eq.(101), such a book-keeping would result in a generating function of the form:

$$\chi_1 = \sum_{\mathbf{k} \in \text{ (group 1)}, \mathbf{k} \notin \mathcal{M}} \frac{H_{1,\mathbf{k}}(\mathbf{I}) \exp(i\mathbf{k} \cdot \phi)}{i\mathbf{k} \cdot \omega(\mathbf{I})} \quad . \qquad (105)$$

$I_{1*} > 0$ and $I_{2*} > 0$, and a ball $B_{\delta*}$ of radius δ around I_1*, I_2*. Then, all lines of the form $I_2 = gI_1$, with

$$\frac{I_{2*}/I_{1*} - \delta/\sqrt{I_{1*}^2 + I_{2*}^2 - \delta^2}}{1 + I_{2*}\delta/(I_{1*}\sqrt{I_{1*}^2 + I_{2*}^2 - \delta^2})} \leq g \leq \frac{I_{2*}/I_{1*} + \delta/\sqrt{I_{1*}^2 + I_{2*}^2 - \delta^2}}{1 - I_{2*}\delta/(I_{1*}\sqrt{I_{1*}^2 + I_{2*}^2 - \delta^2})}$$

cross the ball $B_{\delta*}$. Since the set of rational numbers is dense within the set of real numbers, it follows that there are infinitely many rational numbers $g = q/p$ within the above interval. Thus, for arbitrarily small δ, there are infinitely many lines of the form $I_2 = (q/p)I_1$, or $k_1 I_1 + k_2 I_2 = 0$ with $k_1 = q$, $k_2 = -p$, passing through $B_{\delta*}$.

[33]as, already, in the works of Poincaré (1892), and Arnold (1963)

It follows that generating functions of the form (105) can now be defined in *open sets* of the action space. A set of this form results from excluding, from an initial open domain $\mathcal{W}$, a finite number of resonant manifolds corresponding to those wavevectors **k** in Eq.(103) which i) belong to the group 1 and ii) cross $\mathcal{W}$. In practice, we exclude also some neighborhoods of the above manifolds, in order to avoid non-zero, but extremely small divisors in the series. But even after such an exclusion, we are left with an open set where all definitions are possible.

It has been stated that the easiest way to perform the splitting of Fourier terms in groups of different smallness stems from *Fourier theorem*, namely the fact that the Fourier coefficients of an analytic function decay exponentially with increasing Fourier order $|k|$. To this we now turn our attention.

3.2.   Exponential decay of the Fourier coefficients

It is well known that the convergence properties of a series representing a certain function f are determined by the location and structure of the *singularities* of f in a *complex domain* of its arguments. For example, consider the series expansion of the function $f(x) = 1/(1 + x^2)$ around $x_0 = 0$:

$$f(x) = \frac{1}{1 + x^2} = 1 - x^2 + x^4 - x^6 + ..., \quad x \in \mathbf{R} \qquad (106)$$

Applying, e.g., D'Alembert's criterion, it follows that the interval of convergence of the series (106) in the real axis around $x_0 = 0$ is given by $|x| < 1$, despite the fact that f can be defined for all $x \in \mathbf{R}$. This is because the domain of convergence of a series like (106) is defined by the position of singularities of the represented function f in the complex plane $x \in \mathbf{C}$. In our example, f has poles at $x = \pm i$, thus there is a disc of convergence around $x_0 = 0$ with radius $R = |i - x_0| = |i - 0| = 1$. This disc contains the interval $-1 < x < 1$ of the real axis, and this determines the restriction of the domain of convergence of the series (106) for real values of x.

Consider, now, the function

$$F(\phi_1, \phi_2) = \frac{1}{3 + \cos \phi_1 + \cos \phi_2} \quad . \qquad (107)$$

Clearly, F has no poles on the real torus $[0, 2\pi) \times [0, 2\pi)$. However, if ϕ_1 and ϕ_2 are considered to be complex valued quantities, $\phi_j = \phi_{j,R} + i\phi_{j,I}$, $j = 1, 2$, fixing any two values of $\phi_{1,R}$ and $\phi_{2,R}$ we have a singularity at any solution for $(\phi_{1,I}, \phi_{2,I})$ of the set of equations

$$\cos \phi_{1,R} \cosh \phi_{1,I} + \cos \phi_{2,R} \cosh \phi_{2,R} + 3 = 0$$

$$\sin \phi_{1,R} \sinh \phi_{1,I} + \sin \phi_{2,R} \sinh \phi_{2,R} = 0 \quad .$$

The *closest* singularity to the real 2-torus is found if we set $\phi_{1,R} = \phi_{2,R} = \pi$, and $\cosh(\phi_{1,I}) = \cosh(\phi_{2,I}) = 3/2$, or $\phi_{1,I} = \pm\phi_{2,I} = \pm\sigma$, where $\cosh(\sigma) = 3/2$, or

$$\sigma = 0.962424...$$

How will the above singularity be manifested in a series representing $F(\phi_1, \phi_2)$? We first note that since (ϕ_1, ϕ_2) are angular variables, the series representation associated with F is a Fourier series

$$\frac{1}{3 + \cos\phi_1 + \cos\phi_2} = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} h_{k_1,k_2} \exp\left(i(k_1\phi_1 + k_2\phi_2)\right) \qquad (108)$$

where the coefficients h_{k_1,k_2} are given by

$$h_{k_1,k_2} = \frac{1}{4\pi^2} \int_0^{2\pi} \int_0^{2\pi} \frac{\exp\left(-i(k_1\phi_1 + k_2\phi_2)\right)d\phi_1 d\phi_2}{3 + \cos\phi_1 + \cos\phi_2} \qquad (109)$$

and satisfy the relation $h_{k_1,k_2} = h_{-k_1,-k_2}$ by virtue of the even parity of F with respect to both angles ϕ_1 and ϕ_2. Due to the position of the closest singularity, this series cannot be absolutely convergent beyond a so-called complexified torus domain given by

$$\mathbf{T}_\sigma = \left\{ \phi \in \mathbf{C} : Re(\phi) \in [0, 2\pi), Im(\phi) \in (-\sigma, \sigma) \right\} \qquad (110)$$

with $\sigma = 0.962424....$ But now, suppose that we assign ϕ_1 and ϕ_2 with the imaginary values $\phi_1 = \phi_2 = -iu$ for some positive $u < \sigma$. Then, considering, without loss of generality, only the subsequence of all positive wavenumbers k_1, k_2 in the series (108), this subsequence is bounded by the majoring series

$$\sum_{k_1>0, k_2>0}^{\infty} |h_{k_1,k_2}|e^{(k_1+k_2)u} \quad .$$

This series should be absolutely convergent for all $u < \sigma$. However, this is only possible if the coefficients $|h_{k_1,k_2}|$ are bounded by an exponentially decaying function of $|k| = k_1 + k_2$. This is indeed the case, as proven by the so-called *Fourier theorem* on analytic functions (see, for example, Pinsky (2002)), which states that for all $\sigma' < \sigma$ one has:

$$|h_{k_1,k_2}| < A_{\sigma'}e^{-\sigma'|k|}, \quad \text{for } A_{\sigma'} = \sup_{\mathbf{T}'_\sigma} |F(\phi_1, \phi_2)| \quad . \qquad (111)$$

Thus, choosing any value of σ' in the interval $u < \sigma' < \sigma$ provides an exponentially decaying bound on the size of the coefficients $|h_{k_1,k_2}|$.

In practice, we find optimal bounds if we set σ' to values very close to σ. Figure 9 shows the exponential decay of the coefficients h_{k_1,k_2} defined by Eq.(109). The coefficients are computed using the equations

$$h_{k_1,k_2} = \sum_{n=N}^{\infty} \sum_{j=0}^{n-N} \frac{(-1)^{2n}(2n)!}{2^{2n}3^{2n+1}(|k_1|+j)!j!(|k_2|+n-j-N)!(n-j-N)!}$$

$$\text{with } N = \frac{|k_1|+|k_2|}{2} \text{ if } |k_1|+|k_2| \text{ even} \qquad (112)$$

$$h_{k_1,k_2} = \sum_{n=N}^{\infty} \sum_{j=0}^{n-N} \frac{(-1)^{2n+1}(2n+1)!}{2^{2n+1}3^{2n+2}(|k_1|+j)!j!(|k_2|+n-j-N)!(n-j-N)!}$$

$$\text{with } N = \frac{|k_1|+|k_2|-1}{2} \text{ if } |k_1|+|k_2| \text{ odd}$$
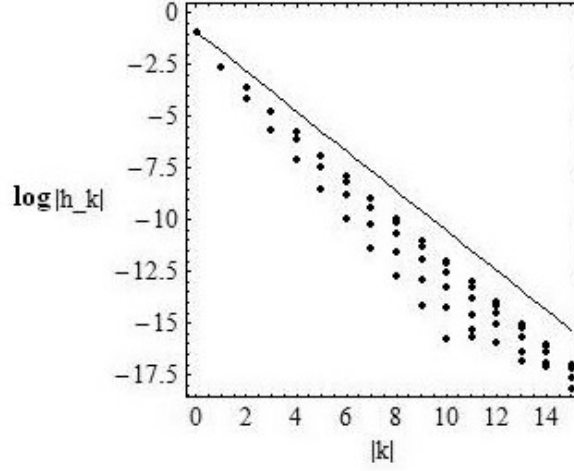
Figure 9.   The logarithm of the absolute value of all coefficients h_{k_1,k_2} given by Eq.(109) as a function of the Fourier order $|k| = |k_1| + |k_2|$. The straight line corresponds to an upper bound law of the form (114), with $A = 0.4$, and $\sigma = 0.96$.

which follow directly after computing the Taylor expansion

$$\frac{1}{3+x} = \frac{1}{3} - \frac{x}{9} + \frac{x^2}{27} - ..., \tag{113}$$

and substituting x by $\cos\phi_1 + \cos\phi_2$. In Fig.9, the value of every coefficient h_{k_1,k_2} is computed approximately, by truncating the corresponding sum in Eq.(112) up to the order $n = |k_1| + |k_2| + 20$. The upper bound for the size of all coefficients is given by a straight line in Fig.9, which corresponds to the exponential law

$$|h_k| < A \exp(-|k|\sigma) \tag{114}$$

where, σ, now, denotes a numerical value set slightly below the value of the singularity, namely $\sigma = 0.96$, while $A = 0.4$.

The property of the exponential decay of Fourier coefficients provides the basis for a splitting of the Fourier harmonics of a Hamiltonian function in groups of different orders of smallness, i.e., for implementing a book-keeping of the form suggested in Eq.(104). In fact, based on Eq.(114), for any positive integer K', we can define the following groups of terms, of different smallness, according to the Fourier order $|k|$ of each term (see Giorgilli 2002, p.90-91, for details):

terms of order $0 \leq |k| < K'$ smallness $= O\left(A(e^{-K'\sigma})^0\right)$

terms of order $K' \leq |k| < 2K'$ smallness $= O\left(A(e^{-K'\sigma})^1\right)$

terms of order $2K' \leq |k| < 3K'$ smallness $= O\left(A(e^{-K'\sigma})^2\right)$

$$\cdots$$

We observe that the various groups are in *ascending powers of the quantity* $e^{-\sigma K'}$, which can thus be regarded as a *natural small quantity* appearing in a Hamiltonian with exponentially decaying Fourier coefficients.

How to choose K'? The simplest (and in many aspects optimal, see Giorgilli 2002) choice is to set $K' \sim 1/\sigma$, implying that the grouping is done in powers of the quantity $1/e$. An alternative choice is to determine K' so that the 'natural' (due to the exponential decay) small parameter $e^{-\sigma K'}$ becomes equal to the other small parameter of the Hamiltonian (97), i.e. ϵ. We thus require that $e^{-\sigma K'} \sim \epsilon$, or:

$$K' = \left[- \frac{\log(\epsilon)}{\sigma} \right] \quad . \tag{115}$$

In practice, due to the logarithmic dependence of K' on ϵ, as well as the fact discussed already, namely that we have a certain flexibility in defining a book-keeping, we find the following rule for practical normal form computations:

Practical rule for choosing K': set K' constant and equal to an average of the values found by Eq.(115) within the range of values of ϵ encountered in the particular problem under study.

This leads to the following

Practical rule for book-keeping in Hamiltonians of the form (97):

$$
\begin{aligned}
\epsilon(\text{terms of order} \quad 0 \leq |k| < K') &\quad \to \lambda^1 \\
\epsilon(\text{terms of order} \quad K' \leq |k| < 2K') &\quad \to \lambda^2 \\
\epsilon(\text{terms of order} \quad 2K' \leq |k| < 3K') &\quad \to \lambda^3
\end{aligned}
\tag{116}
$$
$$\cdots$$

Further implications arise by the fact that the Fourier coefficients of a Hamiltonian expansion usually exhibit explicit dependence on the action variables of the problem under study. However, it is possible to see that this dependence takes place via rules which essentially allow to reproduce a book-keeping of the form (116) with a small modification, examples of which are given below.

3.3.   Application 1: Width of resonances in a simple model

As an application of the book-keeping rules discussed in subsection 3.2, we consider now a simple Hamiltonian system of two degrees of freedom, depending on a small parameter ϵ, in which an appropriate form of resonant normal form theory yields the size of the islands of stability corresponding to various resonances, for small enough values of ϵ.

We consider the Hamiltonian function

$$H \equiv H_0 + \epsilon H_1 = \frac{I_1^2 + I_2^2}{2} + \epsilon \left(\frac{1 + I_1 + 2I_2 + I_1^2 - 3I_2^2 + 2I_1 I_2}{3 + \cos\phi_1 + \cos\phi_2} \right) \quad . \tag{117}$$

The Hamiltonian (117) is a 2D variant of a 3D Hamiltonian model introduced in Froeschlé et al. (2000) and used in subsequent studies of diffusion in the so-called Arnold web in the weakly chaotic regime (see section 4).

The Hamiltonian flow under the unperturbed part of the Hamiltonian, i.e. H_0, is quite simple: we have $\dot{I}_i = 0$, or $I_i = const.$, and $\dot{\phi}_i = I_i = \omega_i$, $i = 1, 2$. Thus, both actions I_i, $i = 1, 2$ are integrals of the flow of H_0, corresponding to motions under constant frequencies $\omega_i = I_i$.
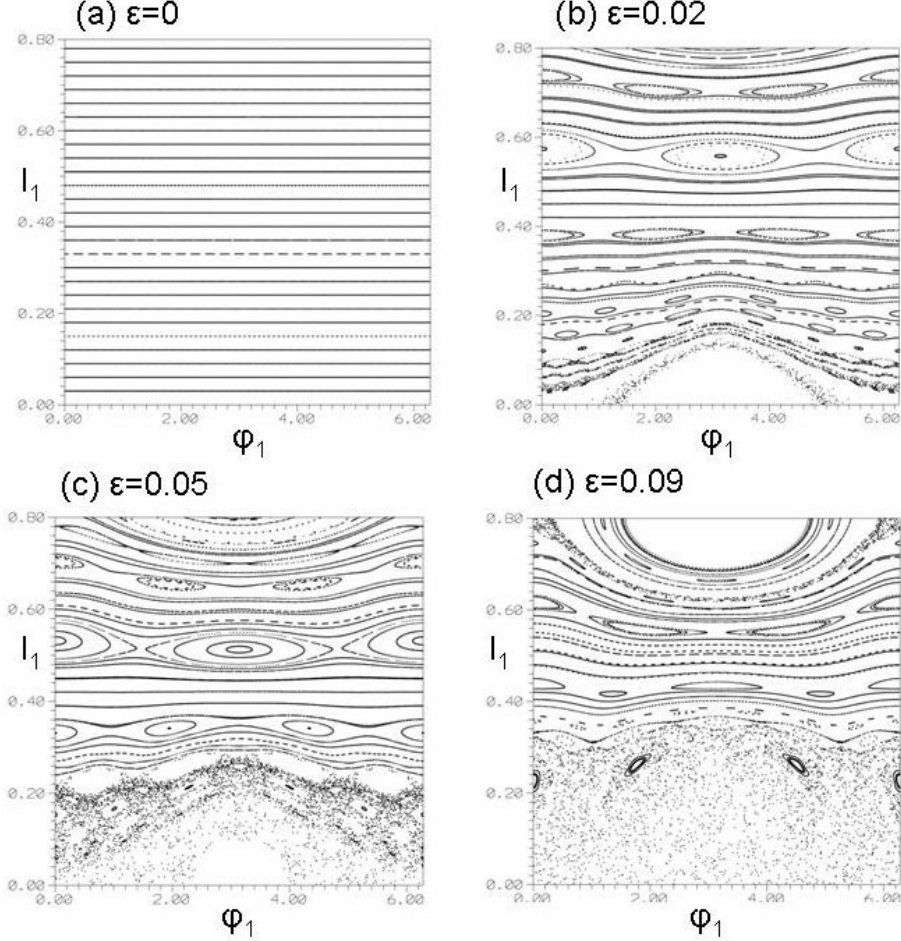
Figure 10.   Phase portraits (surfaces of section (ϕ_1, I_1) for $mod(\phi_2, 2\pi) = 0$, $\dot{\phi}_2 > 0$), in the Hamiltonian model (117), for the energy $E = 1$, and (a) $\epsilon = 0$, (b) $\epsilon = 0.02$, (c) $\epsilon = 0.05$, and (d) $\epsilon = 0.09$.

As in Fig.1, we will use a surface of section, defined by the condition $\phi_2 mod 2\pi = 0$, $\dot{\phi}_2 > 0$, to provide phase portraits of the dynamics of the system (117) for various values of ϵ. Figure 10a corresponds to the motion in the unperturbed case $\epsilon = 0$, for a constant energy $H = E = 1$. Since, for $\epsilon = 0$ the actions remain constant in time for any pair of initial conditions $(I_1(0), I_2(0)) = (I_{1*}, I_{2*})$, the surface of section (ϕ_1, I_1) (Fig.10a) yields a set of straight lines $I_1 = I_{1*}$, representing the intersection, with the surface of section, of a foliation of *invariant tori* of the system under study.

The set of all invariant tori of Fig.10a can be divided in two distinct sets, of so-called *non-resonant*, or *resonant* tori. A non-resonant torus is one for which the frequencies $(\omega_{1*}, \omega_{2*}) = (I_{1*}, I_{2*})$ satisfy no commensurability relation, i.e no relation of the form:

$$k_1\omega_{1*} + k_2\omega_{2*} = 0, \quad (k_1, k_2) \in \mathbf{Z}, \quad |k| = |k_1| + |k_2| \neq 0 \ . \tag{118}$$

Conversely, if (118) is satisfied for some integer vector (k_1, k_2), the torus is called resonant. In the surface of section of Fig.10a, any initial condition on a resonant torus yields a distinct set of points along the corresponding straight line of Fig.10a. For example, in the case of the resonance

$$2\omega_{1*} - \omega_{2*} = 2I_{1*} - I_{2*} = 0$$

we have $\phi_1 - \phi_1(0) = \omega_{1*}t = (1/2)\omega_{2*}t = (1/2)(\phi_2 - \phi_2(0))$. Thus, whenever ϕ_2 completes two periodic circles, ϕ_1 completes one full circle, returning to the initial condition $\phi_1(0)$. Thus, in the surface of section we obtain two distinct points along the straight line $I_1 = I_{1*}$, i.e. one point for each of the two periodic circles of ϕ_2. After completion of the second circle, the motion is repeated periodically, thus yielding always a repetition of the same two consequents in the surface of section. This is called *a periodic orbit* of multiplicity two.

On the other hand, if the actions I_{1*}, I_{2*} are chosen so that the associated frequencies ω_{1*}, ω_{2*} satisfy no commensurability relation of the form (118), the associated non-resonant torus in Fig.10a is filled densely by the consequents of a single orbit produced by taking any initial condition on the torus.

For a fixed energy (e.g. $E = 1$, as in Fig.10), we can find precisely the values of I_{1*} for which the motion is on a resonant torus, by solving simultaneously the equations

$$k_1 I_{1*} + k_2 I_{2*} = 0, \quad \frac{I_{1*}^2 + I_{2*}^2}{2} = E$$

for any pair of integers (k_1, k_2) with $|k_1| + |k_2| \neq 0$. The positive solutions for I_1^* are given by:

$$
\begin{aligned}
I_{1*} &= 0 &&\text{if } k_2 = 0 \\
I_{1*} &= \sqrt{\frac{2E}{1 + \frac{k_1^2}{k_2^2}}} &&\text{if } k_2 \neq 0 \ .
\end{aligned}
\tag{119}
$$

Eq.(119) can be used to find approximately the position of resonances also if $\epsilon \neq 0$, but small. The evolution of the phase portrait with increasing ϵ is shown in Fig.10. These portraits exemplify the well known phenomenon of transition to large scale chaos via the mechanism of *resonance overlap*. Namely, we see that, as ϵ increases, various resonant zones containing islands of stability occupy a larger and larger part of the phase space. For ϵ small, however, nearly every island chain appears delimited by a thin separatrix-like border, which is delimited, in turn, by two rotational tori, one below and one above the resonance. Nevertheless, beyond some critical ϵ (see below) the size of some resonant zones becomes so great that the zones start *overlapping*, thus forming an extended chaotic domain. In Fig.10, the first resonances to overlap as ϵ increases are

those closer to the $I_1 = 0$ axis. In fact, as ϵ increases, the formation of the chaotic domain due to resonance overlapping appears to progress from the lower part of the phase portrait upwards.

We describe now the main steps of construction of a resonant normal form enabling us to explain some of the above features of the phase portraits of Fig.10. We are interested, in particular, in: i) estimating by analytical means the size of the resonant zones, or, the so-called *separatrix width*, for various islands of Fig.10, and ii) making estimates about the critical value of ϵ where we have the onset the resonance overlapping regime leading to the appearance of extended chaos in Fig.10.

Let us start with the example of an important resonance appearing in the phase portraits of Figs.10b,c for values of I_1 roughly in the interval $0.50 \leq I_1 \leq 0.60$. This resonance corresponds to the choice $k_1^{(1)} = 2$, and $k_2^{(1)} = -1$ (otherwise called a '2:1 resonance') [34]. In the phase portraits of Figs.10b,c, we see the formation of two islands of stability within the corresponding resonant zone (one of the islands appears broken due to the modulo 2π evaluation of the angle ϕ_1).

As explained already, for $\epsilon = 0$ we have no islands but simply a resonant torus (straight line in Fig.10a) corresponding to the 2:1 resonance. The associated action value I_{1*} is found via Eq.(119), for $E = 1$, while the value of I_{2*} is found by the equation $k_1^{(1)} I_{1*} + k_2^{(1)} I_{2*} = 0$, for $(k_1^{(1)}, k_2^{(1)}) = (2, -1)$. We find

$$I_{1*} = 0.632456, \quad I_{2*} = 1.264912 \ . \tag{120}$$

For $\epsilon \neq 0$ but relatively small (Fig.10b, $\epsilon = 0.02$), we observe that the whole resonant zone moves downwards, although it remains relatively close to the value of I_{1*} given by Eq.(120). Furthermore, as ϵ increases, the width of the resonant zone increases also. By numerical calculations we find that this tendency is maintained up to a value of $\epsilon \simeq 0.06$. However, for higher values of ϵ the size of the 2:1 islands starts decreasing, and at $\epsilon = 0.09$ (Fig.10d), the 2:1 island chain disappears all together.

The implementation of resonant normal form theory allows to interpret these phenomena, both qualitatively and quantitatively. The construction of a resonant normal form is done in nearly the same way as in the example examined in subsection 2.9, the main difference being in the book-keeping of the Hamiltonian, which in the present case follows the practical rules developed in subsection 3.2. We have the following steps:

Shift of center and expansion of the Hamiltonian (117) in the action variables with respect to the values I_{1*}, I_{2*}. This is analogous to the expansion of the Hamiltonian (3) around a resonant value p_*. In the present case, we introduce

---

[34]Use is made here of the same notation as in subsection 2.9. Namely, the superscript (1) means 'the first resonance condition'. Similarly to subsection 2.9, in the present case as well the use of the superscript (1) to enumerate a resonant vector is rather redundant, since in two degrees of freedom there can be no more than one linearly independent resonance conditions. However, this notation is consistent and actually allows one to more easily follow the analysis of resonant dynamics in the next section, where we pass from examining systems of two degrees of freedom to examining systems of three degrees of freedom

the canonical transformation:

$$I_1 = I_{1*} + I_1', \quad I_2 = I_{2*} + I_2' \quad . \tag{121}$$

Substituting (121) into (117), dropping a constant term, and substituting the numerical values of Eq.(119), the Hamiltonian takes the form:

$$H(\phi_1, \phi_2, I_1', I_2') = 0.632456 I_1' + 1.264912 I_2' + \frac{I_1'^2 + I_2'^2}{2}$$
$$+ \frac{\epsilon \left(1.36228 + 4.79473 I_1' - 4.32456 I_2' + I_1'^2 + 2 I_1' I_2' - 3 I_2'^2\right)}{3 + \cos\phi_1 + \cos\phi_2} \quad . \tag{122}$$

Action rescaling: In subsection 2.9 it was shown that the separatrix width of a resonance scales proportionally to $\epsilon^{1/2}$. As shown below, this is a generic property, maintained also in systems of the form (97). Such a property can be 'a priori' taken into account in the normalization algorithm by introducing a *scaling* transformation $I_1', I_2' \to J_1, J_2$ defined by:

$$J_i = \epsilon^{1/2} I_i', \quad i = 1, 2 \quad . \tag{123}$$

The transformation (123) is not canonical. However, it is straightforward to check that if we substitute (123) in the Hamiltonian (122), the equations of motion take the correct form in the new variables (ϕ, J) under a new Hamiltonian, given by:

$$H'(\phi, J) = \epsilon^{-1/2} H(\phi, J) \quad . \tag{124}$$

In the case of the Hamiltonian (122) we find:

$$H'(\phi_1, \phi_2, J_1, J_2) = 0.632456 J_1 + 1.264912 J_2 + \epsilon^{1/2} \frac{J_1^2 + J_2^2}{2} \tag{125}$$

$$+ \frac{\epsilon^{1/2}\left(1.36228 + \epsilon^{1/2}(4.79473 J_1 - 4.32456 J_2) + \epsilon(J_1^2 + 2 J_1 J_2 - 3 J_2^2)\right)}{3 + \cos\phi_1 + \cos\phi_2} \quad .$$

A careful inspection of the new Hamiltonian (125) shows an essential feature: the fact that the terms quadratic in the actions are of order at least $O(\epsilon^{1/2})$ allows us to use only linear terms in the kernel of the homological equations computed at subsequent steps. This is analogous to the introduction of the book-keeping factor λ in front of the term $I_\psi^2/2$ in the example of subsection 2.3 (Eq.(24)).

Book-keeping: In the Hamiltonian (125) all terms (apart from the first two linear in the actions) are multiplied by some power of the quantity $\epsilon^{1/2}$. Thus, in order to split the terms in the Fourier series of (125) in groups of different orders of smallness compatible with powers of the quantity $\epsilon^{1/2}$, we use Eq.(115), but with $\epsilon^{1/2}$ instead of ϵ, i.e. we set

$$K' = \left[-\frac{\log(\epsilon^{1/2})}{\sigma}\right] \tag{126}$$

with $\sigma = 0.96$. The mean value of K' when, as in Fig.10, ϵ is varied in the interval $0.01 \leq \epsilon \leq 0.1$, is $K' = 2$. Finally, we implement the book-keeping rule of Eq.(116), again with $\epsilon^{1/2}$ instead of ϵ. In fact, since in (125) we have the appearance of terms with factors depending explicitly on a power of $\epsilon^{1/2}$, the book-keeping rule (116) is supplemented by the requirement to introduce an additional factor λ^p in front of any of the terms in Eq.(125) which appears multiplied by some factor $(\epsilon^{1/2})^p$.

In summary:

$$H'(\phi_1, \phi_2, J_1, J_2) = 0.632456 J_1 + 1.264912 J_2 + \lambda \epsilon^{1/2} \frac{J_1^2 + J_2^2}{2}$$

$$+ \ \frac{\lambda \epsilon^{1/2}\left(1.36228 + \lambda \epsilon^{1/2}(4.79473 J_1 - 4.32456 J_2) + \lambda^2 \epsilon(J_1^2 + 2 J_1 J_2 - 3 J_2^2)\right)}{3 + \cos\phi_1 + \cos\phi_2} =$$

$$0.632456 J_1 + 1.264912 J_2 + \lambda \epsilon^{1/2} \frac{J_1^2 + J_2^2}{2}$$

$$+ \ \lambda \epsilon^{1/2}\left(1.36228 + \lambda \epsilon^{1/2}(4.79473 J_1 - 4.32456 J_2) + \lambda^2 \epsilon(J_1^2 + 2 J_1 J_2 - 3 J_2^2)\right)$$

$$\times \ \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} \lambda^{\left[\frac{|k_1| + |k_2|}{K'}\right]} h_{k_1, k_2} \exp\left(i(k_1\phi_1 + k_2\phi_2)\right) \tag{127}$$

with h_{k_1, k_2} given by (109), and $K' = 2$.

Computing the Fourier coefficients as in Eqs.(112), the Hamiltonian up to Fourier order 5 reads:

$$H'(\phi_1, \phi_2, J_1, J_2) = 0.632456 J_1 + 1.264912 J_2 + \lambda \epsilon^{1/2} \frac{J_1^2 + J_2^2}{2}$$

$$+ \ \lambda \epsilon^{1/2}\left(1.36228 + \lambda \epsilon^{1/2}(4.79473 J_1 - 4.32456 J_2) + \lambda^2 \epsilon(J_1^2 + 2 J_1 J_2 - 3 J_2^2)\right)$$

$$\times \Bigg[\ 0.384023 - 0.0760351(e^{i\phi_1} + e^{-i\phi_1} + e^{i\phi_2} + e^{-i\phi_2})$$

$$+ 0.016008\lambda(e^{2i\phi_1} + e^{-2i\phi_1} + e^{2i\phi_2} + e^{-2i\phi_2})$$

$$+ 0.0280895\lambda(e^{i\phi_1+i\phi_2} + e^{-i\phi_1+i\phi_2} + e^{i\phi_1-i\phi_2} + e^{-i\phi_1-i\phi_2})$$

$$- 0.00354632\lambda(e^{3i\phi_1} + e^{-3i\phi_1} + e^{3i\phi_2} + e^{-3i\phi_2})$$

$$- 0.00823336\lambda(e^{i\phi_1+2i\phi_2} + e^{-i\phi_1+2i\phi_2} + e^{i\phi_1-2i\phi_2} + e^{-i\phi_1-2i\phi_2}$$

$$+ e^{2i\phi_1+i\phi_2} + e^{-2i\phi_1+i\phi_2} + e^{2i\phi_1-i\phi_2} + e^{-2i\phi_1-i\phi_2})$$

$$+ 0.000817397\lambda^2(e^{4i\phi_1} + e^{-4i\phi_1} + e^{4i\phi_2} + e^{-4i\phi_2}) \tag{128}$$

$$+0.00222626\lambda^2(e^{i\phi_1+3i\phi_2} + e^{-i\phi_1+3i\phi_2} + e^{i\phi_1-3i\phi_2} + e^{-i\phi_1-3i\phi_2}$$

$$+e^{3i\phi_1+i\phi_2} + e^{-3i\phi_1+i\phi_2} + e^{3i\phi_1-i\phi_2} + e^{-3i\phi_1-i\phi_2})$$

$$+0.00307641\lambda^2(e^{2i\phi_1+2i\phi_2} + e^{-2i\phi_1+2i\phi_2} + e^{2i\phi_1-2i\phi_2} + e^{-2i\phi_1-2i\phi_2})$$

$$-0.000194075\lambda^2(e^{5i\phi_1} + e^{-5i\phi_1} + e^{5i\phi_2} + e^{-5i\phi_2})$$

$$-0.000581994\lambda^2(e^{i\phi_1+4i\phi_2} + e^{-i\phi_1+4i\phi_2} + e^{i\phi_1-4i\phi_2} + e^{-i\phi_1-4i\phi_2}$$

$$+e^{4i\phi_1+i\phi_2} + e^{-4i\phi_1+i\phi_2} + e^{4i\phi_1-i\phi_2} + e^{-4i\phi_1-i\phi_2})$$

$$-0.000995866\lambda^2(e^{2i\phi_1+3i\phi_2} + e^{-2i\phi_1+3i\phi_2} + e^{2i\phi_1-3i\phi_2} + e^{-2i\phi_1-3i\phi_2}$$

$$+e^{3i\phi_1+2i\phi_2} + e^{-3i\phi_1+2i\phi_2} + e^{3i\phi_1-2i\phi_2} + e^{-3i\phi_1-2i\phi_2}) + ... \Bigg]$$

Hamiltonian normalization. We now implement the usual resonant normal form procedure, in precisely the same way as in subsection 2.9. The resonant module is:

$$\mathcal{M} = \{k \text{ such that } k \cdot m = 0\}, \quad \text{where } m = (1, 2) \quad . \tag{129}$$

We give below the form of the generating functions χ_1 and χ_2 arising in the first two normalization steps:

$$\chi_1 = \lambda i \Bigg[0.163776\epsilon^{1/2}(e^{i\phi_1} - e^{-i\phi_1}) + 0.0818879\epsilon^{1/2}(e^{i\phi_2} - e^{-i\phi_2}) \Bigg]$$

$$\chi_2 = \lambda^2 i \Bigg[-0.0172402\epsilon^{1/2}(e^{2i\phi_1} - e^{-2i\phi_1}) - 0.0086201\epsilon^{1/2}(e^{2i\phi_2} - e^{-2i\phi_2})$$

$$+0.0605033\epsilon^{1/2}(e^{i(\phi_1-\phi_2)} - e^{-i(\phi_1-\phi_2)}) - 0.0201678\epsilon^{1/2}(e^{i(\phi_1+\phi_2)} - e^{-i(\phi_1+\phi_2)})$$

$$+0.0025462\epsilon^{1/2}(e^{3i\phi_1} - e^{-3i\phi_1}) + 0.0012731\epsilon^{1/2}(e^{3i\phi_2} - e^{-3i\phi_2})$$

$$+0.0035468\epsilon^{1/2}(e^{i(\phi_1+2\phi_2)} - e^{-i(\phi_1+2\phi_2)}) - 0.0059114\epsilon^{1/2}(e^{i(\phi_1-2\phi_2)} - e^{-i(\phi_1-2\phi_2)})$$

$$+0.0044336\epsilon^{1/2}(e^{i(2\phi_1+\phi_2)} - e^{-i(2\phi_1+\phi_2)}) + (0.317480J_1 - 0.519907J_2)\epsilon(e^{i\phi_1} - e^{-i\phi_1})$$

$$+(0.288216J_1 - 0.324691J_2)\epsilon(e^{i\phi_2} - e^{-i\phi_2}) \Bigg] \quad .$$

After two normalization steps, the Hamiltonian reads (omitting a constant)

$$H^{(2)} = 0.632456J_1 + 1.264912J_2 + \lambda 0.5\epsilon^{1/2}(J_1^2 + J_2^2)$$

$$+\lambda^2 \Bigg[1.84129\epsilon J_1 - 1.66073\epsilon J_2 - 0.011216\epsilon^{1/2}(e^{i(2\phi_1-\phi_2)} + e^{-i(2\phi_1-\phi_2)}) \Bigg] + O(\lambda^3) \quad .$$

In order to give $H^{(2)}$ the usual 'pendulum' form, we pass, again as in subsection 2.9, to resonant variables (ϕ_R, J_R), introducing also a pair of 'fast' canonical variables (ϕ_F, J_F). The fast angle will be ignorable, implying that J_F is a second

integral of the normal form flow. The transformation from the original to resonant variables is exactly the same as in Eq.(70) (with J_1 in the place of I_ψ and J_2 in the place of I), setting $k^{(1)} = (2, -1)$ and $m = (1, 2)$. We find $J_1 = 2J_R + J_F$, $J_2 = -J_R + 2J_F$, and $\phi_R = 2\phi_1 - \phi_2$. Substituting these expressions to $H^{(2)}$, transforming back to trigonometric functions, and setting $\lambda = 1$, we find the following expression of the resonant normal form in the variables (ϕ_R, J_R, J_F) (apart from constant terms):

$$
\begin{aligned}
Z_{res} &= 3.16228 J_F + 2.5\epsilon^{1/2} J_F^2 \\
&+ \epsilon^{1/2}\left(2.5 J_R^2 - 0.022432 \cos\phi_R\right) \\
&+ \epsilon(5.34331 J_R - 1.48017 J_F) \quad .
\end{aligned}
\tag{130}
$$

Eq.(130) clearly shows that, neglecting terms of order ϵ or higher, the resonant normal form takes the form of a pendulum Hamiltonian. This, in turn, can be used to estimate the size of the 2:1 islands by computing the *separatrix half-width* of the associated pendulum. Neglecting terms of order ϵ, from the second line of Eq.(130) we can estimate the separatrix half-width using the same method as in subsection 2.9. We find:

$$
\Delta J_R = \sqrt{\frac{2 \cdot 0.022432}{2.5}} \simeq 0.134 \quad .
\tag{131}
$$

Since $J_1 = 2J_R + J_F$, and taking into account that J_F is an integral of the Hamiltonian (130) and hence exhibits no variations along the resonance separatrix, we have $\Delta J_1 = 2\Delta J_R \simeq 0.268$. Finally, passing back to the original variable I_1 before the re-scaling of Eq.(123), we have

$$
\Delta I_1 = \Delta I_1' \simeq 0.268\epsilon^{1/2} \quad .
\tag{132}
$$

We recover here the well known result that the separatrix half-width scales proportionally to $\epsilon^{1/2}$. It should be noted that all the above estimates refer to the new canonical variables, obtained after the Lie transformations with the generating functions χ_1 and χ_2. However, these estimates remain precise when transforming back to the old variables, since the corrections induced by such transformation can be only of higher order, i.e. $O(\epsilon)$. In fact, for $\epsilon = 0.02$, $\epsilon = 0.05$, or $\epsilon = 0.09$ we find $\Delta I_1 \simeq 0.038$, $\Delta I_1 \simeq 0.06$ and $\Delta I_1 \simeq 0.08$ respectively. The first two of these values compare quite well with the size of the 2:1 island as found in the corresponding phase portraits of Figs.10b and c respectively. However, in Fig.10d it appears that the 2:1 island chain has been destructed. This is due to the resonance overlap mechanism examined below.

Another relevant phenomenon that can be predicted by the above resonant theory concerns the shift, downwards, of the position of the stable periodic orbit at the center of the 2:1 islands of stability, as ϵ increases. Considering, for definiteness, the island intersecting the axis $\phi_1 = 0$, in Fig.10 we see that the vertical position of the central periodic orbit, which, for $\epsilon = 0$ corresponds to the value $I_1 = I_{1*} = 0.632456$ has shifted to $I_1 \simeq 0.57$ for $\epsilon = 0.02$, and $I_1 \simeq 0.53$ for $\epsilon = 0.05$. This downward shift can be predicted theoretically as follows. From Eq.(130), the position of the stable periodic orbit in the surface of section

corresponds to $\phi_1 = 0$, $\phi_2 = 0$, as well as to the value $J_R = J_{R0}$ for which the condition

$$\frac{\partial Z_{res}}{\partial J_R} = 0$$

is satisfied. We find

$$J_{R0} = -1.06866\epsilon^{1/2} + \dots \tag{133}$$

On the other hand, in evaluating the resonant normal form, we have started from the Hamiltonian (122), which originates from the original Hamiltonian (117) after dropping a constant term equal to $(I_{1*}^2 + I_{2*}^2)/2$. However, I_{1*} and I_{2*} themselves are chosen so that $E = (I_{1*}^2 + I_{2*}^2)/2$, where E is the numerical value of the energy. Thus, for the numerical value of Z_{res} we have

$$E_{Z_{res}} = \frac{1}{\epsilon^{1/2}} \left(E - \frac{I_{1*}^2 + I_{2*}^2}{2} \right) = 0 \quad .$$

Setting $Z_{res} = 0$ in (130), and substituting (133), we find an equation for J_F, whose solution $J_F = J_{F0}$ can be given in form of series in powers of $\epsilon^{1/2}$. We find

$$J_{F0} = 0.007\epsilon^{1/2} + \dots \tag{134}$$

We finally have

$$J_{10} = 2J_{R0} + J_{F0} \simeq -2.13\epsilon^{1/2} + \dots \tag{135}$$

Passing back to the non-scaled variable $I_1 = I_{1*} + \epsilon^{1/2}J_1$, we find

$$I_{10} \simeq 0.63 - 2.13\epsilon + \dots \tag{136}$$

It should be noted, again, that these values refer to the *new* canonical variables, after the transformation by the generating functions χ_1 and χ_2. However, only a small change (of order $O(\epsilon^{3/2})$) takes place by passing back from (136) to the corresponding equation for the old variable I_1. In fact, if we set $\epsilon = 0.02$, or $\epsilon = 0.05$ in Eq.(136) we find $I_{10} = 0.59$ and $I_{10} = 0.52$ respectively, in good agreement with the numerical values.

On the other hand, for $\epsilon = 0.09$ we have $I_{10} = 0.44$. However, we will see that at this perturbation level, the resonance overlapping mechanism is effective, thus significantly altering all estimates based on a resonant normal form theory of the above form. To this we now turn our attention.

Resonance overlapping criterion. The resonance overlapping mechanism (Contopoulos (1966), Rosenbluth et al. (1966), Chirikov (1979)) stems from the following remark: we observe (e.g. via Eq.(136)) that the position of a stable periodic orbit changes, as a function of ϵ by a $O(\epsilon)$ quantity, while the size of the associated islands of stability increases as a function of ϵ by a much larger, i.e. $O(\epsilon^{1/2})$, quantity. Thus, if we consider the resonant zones around *two different* periodic orbits, there is a critical ϵ beyond which the zones necessarily overlap. However, the interaction of two or more resonances introduces chaos. Thus, by using the resonance overlapping criterion we can estimate the value of the perturbation ϵ beyond which we have the onset of a substantial degree of chaos in the system.

In the case of the system (117), we can invoke the resonance overlapping criterion in order to estimate the value of ϵ at which the chaotic domain in the lower part of the surfaces of section of Fig.10 extends up to covering the 2:1 resonance. To this end, we note first that the most important island in the lower part of Figs.10b,c,d corresponds to the resonance $\omega_{1*} = I_{1*} = 0$. For that resonance, the resonant wave-vector is $k^{(1)} = (1,0)$, implying $m = (0,1)$. The value of I_{2*} is given by $I_{2*} = \sqrt{2E}$.

In order to estimate the size of the island formed by this resonance, we implement again resonant normal form theory by exactly the same steps as in the case of the 2:1 resonance, using, however, the parameters stated above for the 1:0 resonance. [35] The reader is invited to accomplish him/herself all necessary calculations, which, in the present case, are sufficient to carry on up to order $O(\lambda)$. We give the final result. The generating function χ_1 is given by:

$$\chi_1 = \lambda 0.116754 i\epsilon^{1/2}(e^{i\phi_2} - e^{-i\phi_2})$$

while the Hamiltonian $H^{(1)} = \exp L_{\chi_1} H^{(0)}$ is given by (apart from constants):

$$H^{(1)} = 1.41421 J_2 + \lambda\epsilon^{1/2}\left[0.5(J_1^2 + J_2^2) + 0.16511(e^{i\phi_1} + e^{-i\phi_1})\right] + O(\lambda^2) \ .$$

The variables (ϕ_1, J_1) form a pair of resonant canonical variables. Following the same procedure as in the case of the 2:1 resonance, we find the separatrix half-width, which, in the present case, is given by

$$\Delta I_1 \simeq 1.15\epsilon^{1/2} \ . \tag{137}$$

For $\epsilon = 0.02$ or $\epsilon = 0.05$ we find $\Delta I_1 \simeq 0.16$, and $\Delta I_1 \simeq 0.26$ respectively, which are, again, in good agreement with the corresponding island sizes estimated numerically in Figs.10b,c.

The position of the stable periodic orbit in this approximation is at $I_{10} = 0$.

We can now implement the resonance overlap criterion: we estimate that an extended chaotic domain is formed between the two resonances at values of ϵ beyond the value at which *the sum of the half-widths of the two resonances becomes equal to the separation between the two central periodic orbits.* This corresponds to the value of ϵ at which the theoretical separatrices of the two resonances become tangent one to the other.

In our example, in view of Eq.(136), the separation ΔI_{per} between the two periodic orbits 2:1 and 1:0 is equal to the position of the 2:1 periodic orbit itself, i.e.

$$\Delta I_{per} = 0.63 - 2.13\epsilon$$

Setting this equal to the sum of the two separatrix half-widths (Eqs.(132) and (137) respectively), we arrive at the equation defining the critical ϵ for the overlapping of the two resonances, namely:

$$0.63 - 2.13\epsilon_c \simeq 1.42\epsilon_c^{1/2} \ . \tag{138}$$

---

[35]This is also called an 'adiabatic' resonance, since one of the frequencies is zero, or, in higher order approximation, very close to zero.

The smallest root is $\epsilon_c = 0.09$. In fact, this value is an *overestimate* of the critical value for resonance overlapping, since, for $\epsilon = 0.09$ (Fig.10d), the 2:1 island chain has already disappeared. This overestimation is due to the fact that we neglected the effect of all other resonances between 2:1 and 1:0, i.e. the resonances 3:1, 4:1, etc. More accurate estimates are found by taking these resonances into account. However, even with a simple calculation based on the overlapping of the most conspicuous resonances, we are lead to a relevant and useful theoretical estimate for the critical value ϵ_c.

3.4.   Kolmogorov normal form in analytic hamiltonian functions

In subsection 2.8 we discussed the implementation of Kolmogorov's algorithm used in proofs of the KAM theorem, in the case of a simple Hamiltonian model like (3). In the present subsection, we give the general form of Kolmogorov's algorithm for Hamiltonian systems of the form (97) for which the following conditions hold:

i) The Hamiltonian is analytic in a complexified domain of the action - angle variables, and

ii) The determinant of the Hessian matrix of the 'unperturbed part' H_0 is different from zero.

As an application, we compute the Kolmogorov normal form in the example of the Hamiltonian (117). We emphasize again the practical aspects of the construction of the Kolmogorov normal form, and in particular the way by which we introduce book-keeping in order to properly take into account the size of the various terms appearing in the Fourier expansion of the Hamiltonian.

The **general algorithm of construction of the Kolmogorov normal form** can be stated as follows:

Step 1: torus fixing. In a Hamiltonian of the form:

$$H(\phi, I; \epsilon) = H_0(I) + \epsilon H_1(\phi, I; \epsilon)$$

analytic in all its arguments, where (ϕ, I) are n-dimensional action-angle variables, we fix a frequency vector $\omega_* \equiv (\omega_{1*}, \ldots, \omega_{n*})$ such that the frequencies $\omega_{1*}, \ldots, \omega_{n*}$ satisfy no commensurability relation. Then, we compute the actions I_* for which $\omega_* = \nabla H_0(I_*)$.

Step 2: shift of center. We define new action variables $J \equiv (J_1, J_2, \ldots, J_n)$ by the 'shift' transformation $I = I_* + J$. We substitute in the Hamiltonian, and expand around I_*. We perform the book-keeping rule (116), with K' given by the rule of Eq.(115). Then, the Hamiltonian takes the form (apart from a

constant): [36]

$$H^{(0)} = \omega_* \cdot J + \frac{1}{2} \left(J \cdot M \cdot J^T \right) + H_0^{>2}(J)$$

$$+\sum_{s=1}^{\infty} \lambda^s \sum_{k \in G_s} \left(h_{s,k,0}^{(0)}(\epsilon, I_*)e^{ik\cdot\phi} + (h_{s,k,1}^{(0)}(\epsilon, I_*) \cdot J)e^{ik\cdot\phi} + h_{s,k,\geq 2}^{(0)}(\epsilon, I_*, J)e^{ik\cdot\phi} \right)$$

where:

i) M is the $n \times n$ Hessian matrix of H_0 computed at I_*

$$M_{ij} = \frac{\partial^2 H_0(I_*)}{\partial I_i \partial I_j} \quad .$$

ii) $H_0^{>2}$ is a series in powers of the action variables J, starting with terms of degree 3.

iii) G_s means the set of wavevectors whose associated Fourier terms should be book-kept at the order s according to the chosen book-keeping rules.

iv) $h_{s,k,0}^{(0)}(\epsilon, I_*)$ are constants given as functions of the (also constant) indicated arguments.

v) $h_{s,k,1}^{(0)}(\epsilon, I_*)$ are n-component constant vectors, depending on the indicated arguments. The notation ($\cdot$) means inner product.

vi) $h_{s,k,\geq 2}^{(0)}(\epsilon, I_*, J)$ are functions of the actions (and of the other indicated arguments), given as polynomial series which contain terms of degree 2 or higher in the actions.

Step 3: recursive normalization algorithm. Suppose r normalization steps have been accomplished. The Hamiltonian has the form:

$$H^{(r)} = \omega_* \cdot J + \frac{1}{2} \left(J \cdot M \cdot J^T \right) + H_0^{>2}(J)$$

$$+\lambda Z_1(\phi, J; \epsilon) + +\lambda^2 Z_2(\phi, J; \epsilon) + \ldots + \lambda^r Z_r(\phi, J; \epsilon)$$

$$+ \sum_{s=r+1}^{\infty} \lambda^s \sum_{k \in G_s} \left(h_{s,k,0}^{(r)}(\epsilon, I_*) + (h_{s,k,1}^{(r)}(\epsilon, I_*) \cdot J) + h_{s,k,\geq 2}^{(r)}(\epsilon, I_*, J) \right) e^{ik\cdot\phi}$$

where the functions Z_i, $i = 1, \ldots, r$ are *in Kolmogorov normal form, i.e. their dependence on the actions is in terms of degree 2 or higher.* Then:

Substep 3.1: eliminate the $O(\lambda^{r+1})$ terms depending on the angles and independent of the actions. To this end, define $\chi_{r+1,0}$ by solving the homological equation:

$$\{\omega_* \cdot J, \chi_{r+1,0}\} + \lambda^{r+1} \sum_{k \in G_{r+1}, |k| \neq 0} h_{r+1,k,0}^{(r)}(\epsilon, I_*)e^{ik\cdot\phi} = 0$$

---

[36] Once again, the reader is invited not to be discouraged by the apparent complexity of the formulae below. In in simple words: H_0 must be decomposed in three parts: i) linear in J, ii) quadratic in J, and iii) all the rest. Also, after performing the book-keeping, all Fourier coefficients of H_1 must be decomposed also in three parts, namely: i) terms independent of the actions, ii) terms linear in the actions, and iii) all the rest.

and compute
$$H^{(r+1,0)} = \exp(L_{\chi_{r+1,0}})H^{(r)} \quad .$$

The hamiltonian takes the form

$$H^{(r+1,0)} = \omega_* \cdot J + \frac{1}{2}\left(J \cdot M \cdot J^T\right) + H_0^{>2}(J)$$

$$+\lambda Z_1(\phi, J; \epsilon) + \lambda^2 Z_2(\phi, J; \epsilon) + \ldots + \lambda^r Z_r(\phi, J; \epsilon)$$

$$+\lambda^{r+1} \sum_{k \in G_{r+1}} \left((h_{r+1,k,1}^{(r+1,0)}(\epsilon, I_*) \cdot J) + h_{r+1,k,\geq 2}^{(r+1,0)}(\epsilon, I_*, J) \right) e^{ik \cdot \phi}$$

$$+ \sum_{s=r+2}^{\infty} \lambda^s \sum_{k \in G_s} \left(h_{s,k,0}^{(r+1,0)}(\epsilon, I_*) + (h_{s,k,1}^{(r+1,0)}(\epsilon, I_*) \cdot J) + h_{s,k,\geq 2}^{(r+1,0)}(\epsilon, I_*, J) \right) e^{ik \cdot \phi} \quad .$$

Substep 3.2 (frequency fixing): eliminate the $O(\lambda^{r+1})$ terms linear in the actions and independent of the angles. To this end, define $\chi_{r+1,c}$ by

$$\chi_{r+1,c} = C \cdot \phi, \quad C = (M^{-1})h_{r+1,0,1}^{(r+1,0)}(\epsilon, I_*)$$

and compute
$$H^{(r+1,c)} = \exp(L_{\chi_{r+1,c}})H^{(r+1,0)} \quad .$$

The hamiltonian takes the form

$$H^{(r+1,c)} = \omega_* \cdot J + \frac{1}{2}\left(J \cdot M \cdot J^T\right) + H_0^{>2}(J)$$

$$+\lambda Z_1(\phi, J; \epsilon) + \lambda^2 Z_2(\phi, J; \epsilon) + \ldots + \lambda^r Z_r(\phi, J; \epsilon)$$

$$+\lambda^{r+1} \sum_{k \in G_{r+1},|k|\neq 0} \left((h_{r+1,k,1}^{(r+1,c)}(\epsilon, I_*) \cdot J) + h_{r+1,k,\geq 2}^{(r+1,c)}(\epsilon, I_*, J) \right) e^{ik \cdot \phi}$$

$$+ \sum_{s=r+2}^{\infty} \lambda^s \sum_{k \in G_s} \left(h_{s,k,0}^{(r+1,c)}(\epsilon, I_*) + (h_{s,k,1}^{(r+1,c)}(\epsilon, I_*) \cdot J) + h_{s,k,\geq 2}^{(r+1,c)}(\epsilon, I_*, J) \right) e^{ik \cdot \phi} \quad .$$

At this step, precisely, use is made of the required property that *the Hessian matrix M should have a determinant different from zero*, i.e. it should be an invertible matrix. This is required in order that the vector $C = (M^{-1})h_{r+1,0,1}^{(r+1,0)}(\epsilon, I_*)$ can be computed. In fact, the generating function $\chi_{r+1,c}$ has the same property as encountered in subsection 2.8, namely it is a function linear rather than trigonometric in the angles. However, again as in the example of subsection 2.8, we can see that the linear dependence of $\chi_{r+1,c}$ on the angles introduces no formal issues in the algorithm, i.e. only terms of trigonometric dependence on the angles survive in the transformed Hamiltonian $H^{(r+1,c)}$.

Substep 3.3: eliminate the $O(\lambda^{r+1})$ terms linear in the actions and depending on the angles. To this end, define $\chi_{r+1,1}$ by solving the homological equation:

$$\{\omega_* \cdot J, \chi_{r+1,1}\} + \lambda^{r+1} \sum_{k \in G_{r+1}, |k| \neq 0} (h_{r+1,k,1}^{(r+1,c)}(\epsilon, I_*) \cdot J)e^{ik \cdot \phi} = 0$$

and compute

$$H^{(r+1)} = \exp(L_{\chi_{r+1,1}})H^{(r+1,c)} \quad .$$

The hamiltonian takes the form

$$H^{(r+1)} = \omega_* \cdot J + \frac{1}{2}\left(J \cdot M \cdot J^T\right) + H_0^{>2}(J)$$

$$+\lambda Z_1(\phi, J; \epsilon) + \lambda^2 Z_2(\phi, J; \epsilon) + \ldots + \lambda^{r+1}Z_{r+1}(\phi, J; \epsilon)$$

$$+ \sum_{s=r+2}^{\infty} \lambda^s \sum_{k \in G_s} \left(h_{s,k,0}^{(r+1)}(\epsilon, I_*) + (h_{s,k,1}^{(r+1)}(\epsilon, I_*) \cdot J) + h_{s,k,\geq 2}^{(r+1)}(\epsilon, I_*, J)\right)e^{ik \cdot \phi}$$

i.e. it has been brought in Kolmogorov normal form up to terms of order $O(\lambda^{r+1})$. This completes one full step of the recursive algorithm of computation of the Kolmogorov normal form.

Example: We will construct the Kolmogorov normal form in the case of the Hamiltonian system (117), for a torus located in between the resonances considered in subsection 2.3, namely for the frequency ratio $\omega_{1*}/\omega_{2*} = I_{1*}/I_{2*} = q = 1/\gamma^2$, where $\gamma = (\sqrt{5}+1)/2$. To this end, we implement the steps of the general algorithm exposed above as follows:

Torus fixing: We have $q = 0.381966$, where $q = I_{1*}/I_{2*}$. Setting (for $\epsilon = 0$) $E = H_0$, we find $I_{2*} = \sqrt{2E/(1+q^2)}$. Hence, for $E = 1$, we have $I_{1*} = 0.504623$, and $I_{2*} = 1.32112$.

Shift of center: We set

$$I_1 = I_{1*} + J_1, \quad I_2 = I_{2*} + J_2 \quad . \tag{139}$$

Substituting (139) into (117), and omitting a constant, the Hamiltonian takes the form:

$$H = 0.504623J_1 + 1.32112J_2 + \frac{J_1^2 + J_2^2}{2}$$

$$+\epsilon\left(\frac{0.49877 + 4.65148J_1 - 4.91747J_2 + J_1^2 + 2J_1J_2 - 3J_2^2}{3 + \cos\phi_1 + \cos\phi_2}\right) \quad .$$

Following now the general book-keeping rule of Eq.(116), we compute a constant K' from the average of Eq.(115) in the considered range of values of ϵ. Within the interval $0.01 \leq \epsilon \leq 0.1$, we find $K' = 3$.

In summary:

$$H^{(0)}(\phi_1, \phi_2, J_1, J_2) = 0.504623 J_1 + 1.32112 J_2 + \frac{J_1^2 + J_2^2}{2}$$

$$+ \quad \lambda \epsilon \Big(0.49877 + 4.65148 J_1 - 4.91747 J_2 + J_1^2 + 2 J_1 J_2 - 3 J_2^2 \Big)$$

$$\times \Bigg[\quad 0.384023 - 0.0760351 (e^{i\phi_1} + e^{-i\phi_1} + e^{i\phi_2} + e^{-i\phi_2})$$

$$+ 0.016008 (e^{2i\phi_1} + e^{-2i\phi_1} + e^{2i\phi_2} + e^{-2i\phi_2})$$

$$+ 0.0280895 (e^{i\phi_1 + i\phi_2} + e^{-i\phi_1 + i\phi_2} + e^{i\phi_1 - i\phi_2} + e^{-i\phi_1 - i\phi_2})$$

$$- 0.00354632 \lambda (e^{3i\phi_1} + e^{-3i\phi_1} + e^{3i\phi_2} + e^{-3i\phi_2})$$

$$- 0.00823336 \lambda (e^{i\phi_1 + 2i\phi_2} + e^{-i\phi_1 + 2i\phi_2} + e^{i\phi_1 - 2i\phi_2} + e^{-i\phi_1 - 2i\phi_2}$$

$$+ e^{2i\phi_1 + i\phi_2} + e^{-2i\phi_1 + i\phi_2} + e^{2i\phi_1 - i\phi_2} + e^{-2i\phi_1 - i\phi_2})$$

$$+ 0.000817397 \lambda (e^{4i\phi_1} + e^{-4i\phi_1} + e^{4i\phi_2} + e^{-4i\phi_2}) \qquad (140)$$

$$+ 0.00222626 \lambda (e^{i\phi_1 + 3i\phi_2} + e^{-i\phi_1 + 3i\phi_2} + e^{i\phi_1 - 3i\phi_2} + e^{-i\phi_1 - 3i\phi_2}$$

$$+ e^{3i\phi_1 + i\phi_2} + e^{-3i\phi_1 + i\phi_2} + e^{3i\phi_1 - i\phi_2} + e^{-3i\phi_1 - i\phi_2})$$

$$+ 0.00307641 \lambda (e^{2i\phi_1 + 2i\phi_2} + e^{-2i\phi_1 + 2i\phi_2} + e^{2i\phi_1 - 2i\phi_2} + e^{-2i\phi_1 - 2i\phi_2})$$

$$- 0.000194075 \lambda (e^{5i\phi_1} + e^{-5i\phi_1} + e^{5i\phi_2} + e^{-5i\phi_2})$$

$$- 0.000581994 \lambda (e^{i\phi_1 + 4i\phi_2} + e^{-i\phi_1 + 4i\phi_2} + e^{i\phi_1 - 4i\phi_2} + e^{-i\phi_1 - 4i\phi_2}$$

$$+ e^{4i\phi_1 + i\phi_2} + e^{-4i\phi_1 + i\phi_2} + e^{4i\phi_1 - i\phi_2} + e^{-4i\phi_1 - i\phi_2})$$

$$- 0.000995866 \lambda (e^{2i\phi_1 + 3i\phi_2} + e^{-2i\phi_1 + 3i\phi_2} + e^{2i\phi_1 - 3i\phi_2} + e^{-2i\phi_1 - 3i\phi_2}$$

$$+ e^{3i\phi_1 + 2i\phi_2} + e^{-3i\phi_1 + 2i\phi_2} + e^{3i\phi_1 - 2i\phi_2} + e^{-3i\phi_1 - 2i\phi_2}) + ... \Bigg]$$

Normalization: We are now ready to perform the first recursive step of the normalization algorithm, separated in the three sub-steps. We give the corresponding formulae for the generating functions and for the final Hamiltonian after the first normalization step. We have:

$$\chi_{1,0} = \lambda \epsilon i \Bigg[0.0751533 (e^{i\phi_1} - e^{-i\phi_1}) + 0.028706 (e^{i\phi_2} - e^{-i\phi_2})$$

$$- 0.0079112 (e^{2i\phi_1} - e^{-2i\phi_1}) - 0.0030218 (e^{2i\phi_2} - e^{-2i\phi_2})$$

$$- 0.0076737 (e^{i(\phi_1 + \phi_2)} - e^{-i(\phi_1 + \phi_2)}) + 0.0171589 (e^{i(\phi_1 - \phi_2)} - e^{-i(\phi_1 - \phi_2)}) \Bigg] \quad .$$

$$\chi_{1,c} = \lambda \epsilon (1.78628 \phi_1 - 1.88842 \phi_2)$$

$$\chi_{1,1} = \lambda \epsilon i \Bigg[(0.551942 J_1 - 0.74095 J_2)(e^{i\phi_1} - e^{-i\phi_1})$$

$$+ (0.267709 J_1 - 0.304746 J_2)(e^{i\phi_2} - e^{-i\phi_2})$$

$$+(-0.0581015J_1 + 0.0779978J_2)(e^{2i\phi_1} - e^{-2i\phi_1})$$
$$+(-0.028181J_1 + 0.0320798J_2)(e^{2i\phi_2} - e^{-2i\phi_2})$$
$$+(0.181038J_1 - 0.190188J_2)(e^{i(\phi_1-\phi_2)} - e^{-i(\phi_1-\phi_2)})$$
$$+(-0.0673611J_1 + 0.0798595J_2)(e^{i(\phi_1+\phi_2)} - e^{-i(\phi_1+\phi_2)})\Bigg]\quad.$$

The Hamiltonian, after the first normalization step is given by

$$H^{(1)} = \exp(L_{\chi_{1,1}})\exp(L_{\chi_{1,c}})\exp(L_{\chi_{1,0}})H^{(0)}\quad.$$

We find:

$$H^{(1)} = \lambda\epsilon\Bigg[1.78498 + 0.384023J_1^2 + 0.768047J_1J_2 - 1.15207J_2^2$$

$$(0.475907J_1^2 - 0.89302J_1J_2 + 0.228105J_2^2)(e^{i\phi_1} + e^{-i\phi_1})$$
$$+(-0.0760351J_1^2 + 0.115639J_1J_2 - 0.0766411J_2^2)(e^{i\phi_2} + e^{-i\phi_2})$$
$$+(-0.100195J_1^2 + 0.188012J_1J_2 - 0.048024J_2^2)(e^{2i\phi_1} + e^{-2i\phi_1})$$
$$+(0.016008J_1^2 - 0.024346J_1J_2 + 0.0161356J_2^2)(e^{2i\phi_2} + e^{-2i\phi_2})$$
$$+(0.209127J_1^2 - 0.315047J_1J_2 + 0.10592J_2^2)(e^{i(\phi_1-\phi_2)} + e^{-i(\phi_1-\phi_2)})$$
$$+(-0.0392716J_1^2 + 0.0686773J_1J_2 - 0.00440891J_2^2)(e^{i(\phi_1+\phi_2)} + e^{-i(\phi_1+\phi_2)})\Bigg]\quad.$$

We close this subsection with a note on the convergence of the Kolmogorov normal form. The accumulation of divisors in the present scheme follows nearly the same pattern as exposed in subsection 2.8. In fact, it turns out that the accumulation of divisors is the same whether one progresses by normalizing according to the quadratic scheme (i.e. in groups of terms $O(\lambda)$, then $O(\lambda^2)$, $O(\lambda^3)$, then $O(\lambda^4)$, $O(\lambda^5)$, $O(\lambda^6)$, $O(\lambda^7)$, etc.), or according to a linear scheme (Giorgilli and Locatelli 1997). Using the same heuristic notation as in subsection 2.8, the situation is summarized with the help of a figure (Fig.11). The arrows indicate the most important Poisson brackets formed by the generating functions $\chi_{r,0}$ and $\chi_{r,1}$ in one example, of, say, the order $r = 5$. The terms specified on top of each arrow indicate the Hamiltonian terms with which the most important repetitions of Poisson brackets take place, either by the generating function $\chi_{5,0}$, or by $\chi_{5,1}$. The symbol (X) denotes a path produced by repeated Poisson brackets of the generating functions $\chi_{r,0}$ or $\chi_{r,1}$, leading to terms that *stop being normalized* at subsequent normalization steps. In fact, this can happen for two reasons: i) a repeated Poisson bracket leads eventually to zero new terms (this can happen if both functions in a bracket are independent of the actions), or it leads to terms depending quadratically on the actions. A careful inspection of Fig.11 shows now that the worst possible accumulation of divisors in Kolmogorov's scheme is quadratic. For example, starting from the normalization order $r = 5$ we have an accumulation of the form:

$$a_5 \to a_5^2 a_{10} \to a_5^4 a_{10}^2 a_{20} \to \ldots$$

Propagation of small divisors
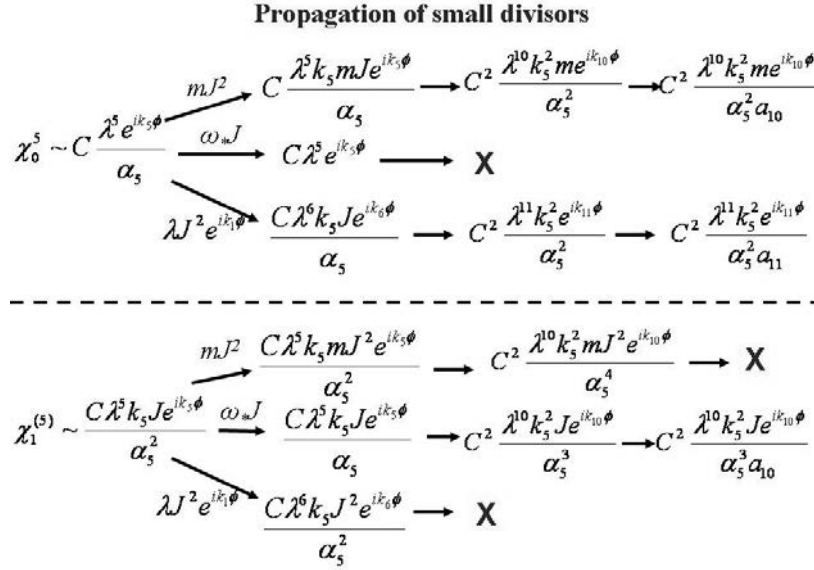
Figure 11. Propagation of divisors in the Kolmogorov normal form construction. Using the heuristic notation introduced in subsections 2.7 and 2.8, the figure shows the most important chains of terms produced in the Kolmogorov series by the repeated Poisson brackets of the generating functions $\chi_{r,0}$ and $\chi_{r,1}$ at the normalization order $r = 5$ (see text for details).

Using the same arguments as in subsection 2.8, we can show that such an accumulation ensures the convergence of the Kolomogorov normal form for sufficiently small values of ϵ. This, in turn, proves the existence of an invariant torus with the given frequencies for sufficiently small ϵ. In fact, a numerical convergence test shows that the convergence in the present example persists up to $\epsilon \simeq 0.06$. This is about 70% of the maximum perturbation value up to which the torus is found to exist by purely numerical means.

4. THREE DEGREES OF FREEDOM: DIFFUSION IN THE ARNOLD WEB

In the present section, we consider nearly-integrable Hamiltonian systems of three degrees of freedom of the form $H = H_0 + \epsilon H_1$, exhibiting a phenomenon of diffusion of weakly chaotic orbits in the web or resonances, otherwise called *Arnold diffusion* (Arnold (1964), Arnold and Avez (1968)).

In nearly integrable systems of two degrees of freedom, the presence of two-dimensional invariant tori, whose existence is a consequence of the Kolmogorov - Arnold - Moser theorem, poses topological restrictions to the possible excursions that a chaotic orbit can undergo within any hypersurface of constant energy $H = E$ embedded in the system's phase space. In fact, in systems of two degrees of freedom the maximal dimension of invariant tori is two, while the hyper-surface $H = E$ has dimension three. Thus, in such a hyper-surface the maximal

tori do not allow communication between chaotic domains in their interior and in their exterior. [37]. However, in systems of three degrees of freedom, the invariant KAM tori can no longer act as absolute barriers to the communication of the various chaotic domains co-existing in the same hyper-surface of constant energy. This is because the latter's dimension (equal to five) exceeds by two the maximal dimension of invariant tori (equal to three). The non-existence of topological restrictions renders *a priori* possible, in turn, that a chaotic orbit started at a certain point of a constant energy hyper-surface undergoes a long excursion, hence visiting (possibly after a quite long time) the neighborhood of any other point in a connected chaotic domain within the same hyper-surface. This topological possibility exists even if the perturbation ϵ is infinitesimally small, and the so-resulting diffusion of chaotic orbits is called Arnold diffusion.

It should be stressed that besides topological arguments, a demonstration that the Arnold diffusion really takes place in a concrete system requires establishing the existence of a *mechanism of transport* of the chaotic orbits within the Arnold web (see Lochak (1999) for a detailed discussion of the distinction between the terms 'Arnold diffusion' and 'Arnold's mechanism'). In fact, Arnold (1964) gave an example of such mechanism, whose existence, however, is guaranteed only in a rather special type of Hamiltonian model of three degrees of freedom. As shown in subsection 4.5, Arnold's model shares some common features with models arising from the implementation of *resonant normal form theory* in systems of three degrees of freedom satisfying some so-called *convexity conditions* (see subsection 4.3 below). However, there is also a crucial difference between the two types of models, discussed in subsection 4.5, which renders difficult (and non-existent, to date) a generalization of the proof of whether Arnold's mechanism exists in generic Hamiltonian systems of three or more degrees of freedom (see Simó and Valls (2001), and Mather (2004)).

On the other hand, following some early works in systems of three degrees of freedom or 4D symplectic mappings (Froeschlé (1970a,b, 1972), Froeschlé and Scheidecker (1973), Tennyson (1982), weakly-chaotic diffusion in the web of resonances has been observed numerically in an extended list of works in recent years referring to model Hamiltonian systems or mappings. Indicative references are: Kaneko and Konishi (1989), Wood et al. (1990), Laskar (1993) Dumas and Laskar (1993), Skokos et al. (1997), Efthymiopoulos et al. (1998), Lega et al. (2003), Giordano and Cincotta (2004), Froeschlé et al. (2005), Guzzo et al. (2005, 2006), Lega et al. (2009, 2010a, 2010b), Cincotta and Giordano (2012), Mestre et al. (2012), Mestre (2012), Efthymiopoulos and Harsoula (2012). Also, various aspects of the problem of diffusion have been studied in concrete astronomical dynamical models, as for example, the motion of asteroids in solar system dynamics (e.g. Holman and Murray (1996), Murray and Holman (1997), Levison et al. (1997), Nesvorny and Morbidelli (1998), Morbidelli and Nesvorny (1999), Marzari et al. (2003), Tsiganis et al. (2005), Cordeiro and Mendes de Souza (2005), Cordeiro (2006), Robutel and Gabern (2006), Lhotka et al. (2008),

---

[37]This property is manifested also when considering Poincaré surfaces of section. For example, in Figs.1a,b any chaotic orbit started in a domain surrounded by a closed invariant curve cannot find itself in another chaotic domain outside the curve. Thus, two chaotic domains, one in the interior and the other in the exterior of the closed KAM curve, do not communicate. Likewise, two domains above and below a rotational KAM curve do not communicate.

Tsiganis (2008, 2010), Cachucho et al. 2010), or the motion along resonances in galactic dynamics (e.g. Papaphilippou and Laskar, (1996, 1998), Wachlin and Ferraz-Mello (1998), Muzzio et al. (2005), Kalapotharakos and Voglis (2005), Cincotta et al. (2006), Aquilano et al. (2007), Valluri et al. (2012)). In fact, there is evidence that such a diffusion takes place in systems satisfying a variety of different convexity (or, more generally *steepness*) conditions, as well as conditions of so-called *a priori stability* (or instability). Thus, not all diffusion phenomena appearing in works as the above should be characterized as Arnold diffusion. Also, an important comment is in order at this point, regarding the *slowness* of Arnold diffusion, as observed in most numerical experiments. In fact, a comparison of the timescale of Arnold diffusion with the timescales of interest in astronomical systems renders often quite questionable whether this type of diffusion is relevant in realistic applications (see also a discussion of this point in the article by Cincotta et al. (2012) in the present volume of proceedings). For example, the orbits of stars in galaxies evolve in a timescale of 10^2 – 10^4 periods, which is quite a small time compared to the timescale in which Arnold diffusion produces observable effects (10^6 – 10^{11} periods for perturbation values as those encountered in Table 1). Thus, it appears that Arnold diffusion plays a small role in galaxies, except perhaps for orbits passing very close to the galactic center. However, the time scale of Arnold diffusion is quite relevant to applications in solar system dynamics, where the characteristic timescales are of order 10^8 – 10^{10} periods.

It should be noted also that in most applications we find that the distinction between the so-called *Nekhoroshev regime* (see below), where the rate of diffusion becomes exponentially small in the inverse of a small parameter, and the *resonance overlap* regime, where the diffusion speed is given by a power-law of ϵ, is not so sharp in practice. However, it is important to emphasize that normal form theory can be applied to some extent in both regimes. In fact, as shown below the most relevant quantity characterizing the speed diffusion turns to be the *size of the remainder R_{opt} at the optimal normalization order*. Thus, one can obtain realistic estimates of the speed of diffusion using R_{opt} in both regimes, i.e. independently on whether R_{opt} depends on $1/\epsilon$ exponentially or algebraically.

In the remaining part of this section we present the main elements of resonant normal form theory as well as some of its applications in describing weakly-chaotic diffusion in the web of resonances in systems of three degrees of freedom. This implementation allows one to derive useful tools for describing, both qualitatively and quantitatively, the *geometric* properties of diffusion. In particular, by the resonant normal form theory we can explain why the diffusion progresses preferentially in some directions of the phase space, and what invariant objects (like periodic orbits or low-dimensional tori) are responsible for such preference. Furthermore, via normal forms we can compute quantitative estimates of the diffusion rate, which are found to compare well with the results of numerical experiments.
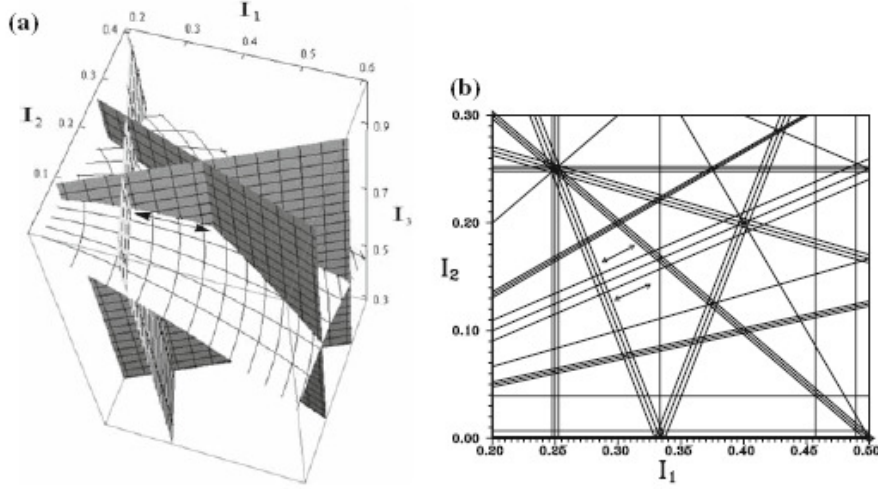
Figure 12.   (a) A part of the paraboloid of constant energy condition $I_1^2 + I_2^2 + 2I_3 = 2E$ in the action space for the model (141) with $\epsilon = 0$, and for the value of the energy $E = 1$. Gray-shaded planes indicate some resonant manifolds in this model. The intersection of all the resonant planes with the paraboloid of constant energy produces a set of parabolic curves which is the 'web of resonances'. (b) The projection of the web of resonances $k_1 I_1 + k_2 I_2 + k_3 = 0$ for $|k| \leq 5$ on the (I_1, I_2) plane. Single or triple lines correspond to resonances whose corresponding coefficient in the Fourier series of H_1 is positive or negative respectively (after Efthymiopoulos (2008)).

4.1.   Diffusion along resonances: a numerical example

As a tool serving to introduce basic concepts, let us consider a simple example of a Hamiltonian model of three degrees of freedom (Froeschlé et al. (2000)):

$$H = H_0 + \epsilon H_1 = \frac{I_1^2 + I_2^2}{2} + I_3 + \frac{\epsilon}{4 + \cos\phi_1 + \cos\phi_2 + \cos\phi_3} \quad . \qquad (141)$$

The integrable part H_0 produces a simple dynamics. All the orbits lie on invariant tori, and we have $\dot{I}_i = 0$, while $\dot{\phi}_1 = \omega_{0,1} = I_1$, $\dot{\phi}_2 = \omega_{0,2} = I_2$, $\dot{\phi}_3 = \omega_{0,3} = 1$.

The surface of constant energy $H_0 = E$ is a paraboloid in the action space given by $I_1^2 + I_2^2 + 2I_3 = 2E$. Part of this paraboloid is shown as a grid-lined inclined surface in Fig.12a.

We define the *resonant manifolds* corresponding to the integrable Hamiltonian H_0 as the set of all two-dimensional manifolds defined by relations of the form:

$$k_1 \omega_{0,1} + k_2 \omega_{0,2} + k_3 \omega_{0,3} = k_1 I_1 + k_2 I_2 + k_3 = 0 \qquad (142)$$

with $k \equiv (k_1, k_2, k_3) \in \mathbf{Z}^3$, $|k| \equiv |k_1| + |k_2| + |k_3| \neq 0$. In fact, we see that for any choice of k, Eq.(142) implies that the invariant manifolds in this example are just planes, which are always normal to the plane (I_1, I_2).

Consider now the intersections of all the resonant planes with a surface of constant energy $E = H_0$. These form a set of parabolic curves, which form a *web* on the surface $E = H_0$ (see Fig.12a). This is called a *web of resonances*.

Viewed from the top of Fig.12a, the projection of the resonance web on the plane (I_1, I_2) is a set of straight lines (Fig. 12b). It will be shown below that, when $\epsilon \neq 0$, a small *resonant domain*, of thickness $O(|h_k\epsilon|^{1/2})$, is formed around each resonant manifold, where h_k is the coefficient of the term $\exp(ik \cdot \phi)$ in the Fourier development of the perturbation H_1. The intersections of the resonant domains with the surface of constant energy form *resonant zones*, also of thickness $O(|h_k\epsilon|^{1/2})$. As a result, when projected on the plane (I_1, I_2), the web of resonances along with the resonant zones looks like in Fig.12b. In fact, in this figure some resonances are represented by a single line, while other by a triple line, i.e. a line corresponding to the central resonance and two nearly parallel lines on either side of it, representing the border of the resonant zone. This difference reflects a real phenomenon encountered when using a numerical method in order to depict the resonance web, that will be explained below.

The main phenomenon, of chaotic diffusion in the Arnold web, is now shown with the example of Fig.13. In both panels, the color (or grayscale) background corresponds to a so-called *FLI map* (Froeschlé et al. (2000)), allowing to clearly see the resonant structure of the system in the action space when (a) $\epsilon = 0.01$, and (b) $\epsilon = 0.015$. The *Fast Lyapunov Indicator* (Froeschlé et al (1997)) is a numerical index of the degree of chaotic behavior of individual orbits, computed via a numerical integration of the linearized equations of motion together with the full equations of motion. The reader is deferred to the original article (Froeschlé et al (1997)), as well as to Maffione et al. (2011) for a review of numerical methods of chaos detection and/or quantification in conservative dynamical systems. For our purpose in the present tutorial, it suffices to note that a dark (blue or green) color in Fig.12 indicates ordered orbits, while a light (yellow) color means chaotic orbits. By the FLI map, we are able to clearly see the network formed by the projection of different resonant manifolds on the plane I_1, I_2, as well as the form of one conspicuous (and other, smaller) *resonance junctions*, i.e. the chaotic domains formed at the loci where the manifolds of different resonances intersect each other.

Using normal form theory, in subsequent subsections it will be shown that the possibility to distinguish the various resonances in Fig.13 is due to the fact that *in the neighborhood of any resonant domain the dynamics can be locally approximated by the dynamics of the perturbed pendulum*. Namely, we will show that the resonant normal form for a Hamiltonian like (141) can be written as:

$$
\begin{aligned}
H_{res} &= Z(I_{R_2}, I_F) + \frac{1}{2}\beta I_R^2 + 2\epsilon g_k \cos\phi_R + \ldots \\
&+ R(\phi_R, \phi_{R_2}, \phi_F, I_R, I_{R_2}, I_F)
\end{aligned}
\tag{143}
$$

where the variables $(\phi_R, \phi_{R_2}, \phi_F, I_R, I_{R_2}, I_F)$ are resonant action-angle variables derived from the original canonical variables by a linear canonical transformation. The derivation of Eq.(143) (see subsection 3.4) follows from a quite similar analysis as in subsection 3.3 for the corresponding 2D problem. We observe that the Hamiltonian (143) is split in three parts: a part depending only on the
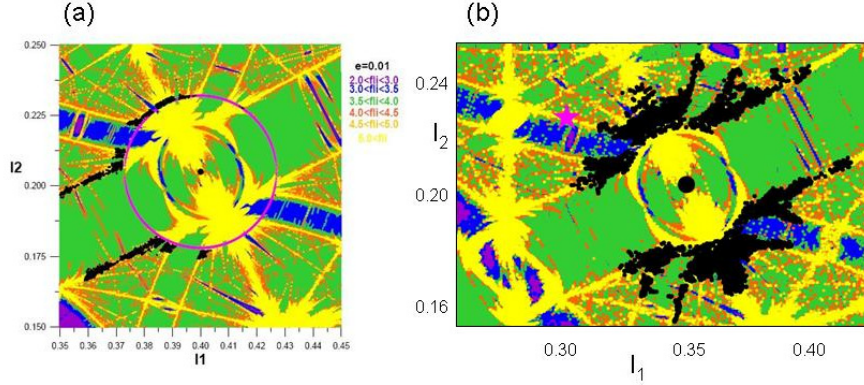
Figure 13. (a) Arnold diffusion in the model (141) for $\epsilon = 0.01$. An orbit with initial conditions in the chaotic separatrix-like layer of the 2:1 resonance slowly diffuses along the resonance, until reaching a doubly-resonant domain. The motion along the simple resonance, as well as the circular motion within the doubly-resonant domain are both interpreted by resonant normal form theory (see subsections 4.3 and 4.4 respectively). (b) Chaotic diffusion for $\epsilon = 0.015$. The same chaotic orbit visits several parts of the Arnold web.

actions I_{R_2}, I_F, a second part given by the pendulum Hamiltonian:

$$H_{pend} = \frac{1}{2}\beta I_R^2 + 2\epsilon g_k \cos \phi_R \tag{144}$$

where β and g_k are constants, and finally, a third part, i.e. the remainder $R(\phi_R, \phi_F, \phi_3, I_R, I_F, I_3)$. The coefficient g_k is nearly equal to the coefficient h_k of the term $\exp(ik\cdot\phi)$ in the Fourier series representing H_1 in the original Hamiltonian, where k is the wave-vector corresponding to the particular resonance considered.

If we neglect the remainder effect, the actions I_{R_2} and I_F are integrals of the Hamiltonian H_{res}. This allows to calculate the separatrix half-width of the resonance in the same way as in subsection 3.3. We find

$$\Delta I_{R,sep} = \sqrt{\frac{8|g_k\epsilon|}{\beta}} \quad . \tag{145}$$

In reality, however, there is no true separatrix at the border of each resonance, since *the effect of the remainder is to introduce chaos in the system.* Thus, instead of a separatrix, we have the formation of a separatrix-like layer at the border of any resonance [38]. Any initial condition started within such a chaotic layer yields a large value of the FLI. Thus, computing many such orbits allows to obtain a clear picture of the structure of all chaotic layers, and hence of the whole

---

[38]This is similar to the effect shown in Fig.6 for the simple perturbed pendulum model (3). Namely, the theoretical invariant curves computed by a normal form in that example yield a separatrix curve, while, in reality, we have instead a thin separatrix-like chaotic layer as revealed by a numerical integration of orbits.

resonance web. In fact, when the resonance web is visualized numerically in the action space, e.g. by the FLI method, one has to make an appropriate choice of Poincaré surface of section in order to fix the angles at which the FLI maps are computed. In Fig.12, we make the choice $\phi_3 = 0$, $|\phi_1| + |\phi_2| \leq 0.05$. This corresponds essentially to setting the resonant angle $\phi_R = k_1\phi_1 + k_2\phi_2 + k_3\phi_3$ to a value very close to zero, i.e., $\phi_R \approx 0$. We then see that this means to plot two sets of points in the plane (I_1, I_2), corresponding to a chaotic orbit passing close to the maxima or minima of the theoretical separatrices, when $g_k < 0$, or one set of points on the plane (I_1, I_2), corresponding to passing close to the X-point of each separatrix, when $g_k > 0$. Consequently, the thin chaotic borders of the resonances appear as a pair of thick lines in the surface of section if $g_k < 0$, or as a single line if $g_k > 0$. This rule is used for plotting the resonances and their borders in Fig.12b.

Returning to Fig.13, the black points show now the excursion traveled in the action space along the resonance web by *one* chaotic orbit, when (a) $\epsilon = 0.01$ and (b) $\epsilon = 0.015$. In both cases, the orbit starts in the left part of the corresponding plots, within the separatrix layer of the resonance $I_1 - 2I_2 = 0$. In Fig.13a, the orbit undergoes a phase of diffusion along a simple resonance. This phase lasts for a quite long time ($t \sim 10^8$), while, afterwards, the orbit enters a so-called doubly-resonant domain formed by the intersection of many resonances. The orbit remains in that domain up to the end of the numerical integration ($t = 2 \times 10^8$). We observe that despite the fact that the central part of the doubly resonant domain appears quite chaotic, the motion of the chaotic orbit in this domain appears confined in a nearly circular arc, while the diffusion rate is quite slow in the direction normal to the arc. In Fig.13b, on the other hand, for somewhat larger ϵ ($\epsilon = 0.015$) we observe that the orbit, within a similar timescale, covers a much more extended part of the resonance web, passing via more than one doubly-resonant domains and undergoing also changes of direction from along one resonance to along another.

Both phenomena, of diffusion along simple resonances, or within doubly resonant domains, as well as the geometric properties of the motion in either case, are explained by an appropriate form of resonant perturbation theory. Furthermore, it is possible to obtain quantitative estimates regarding the *rate of diffusion* (or the value of the diffusion coefficient, assuming, as a first approximation, that the diffusion has a normal character)[39], in terms of the *size of the remainder* of the normal form series computed either in the simply resonant, or doubly resonant case. We will now give the general theory of resonant normal form dynamics referring to both cases. However, we will return to a specific numerical example concerning the Hamiltonian model (141) in subsection 4.4, illustrating Arnold diffusion in a concrete calculation where an appropriate set of resonant canonical variables, produced via a normal form computation, are used.

---

[39]There has been some numerical evidence (see e.g. the review article of Lega et al. (2008)), that the diffusion along simple resonances has a normal character. However, counter-examples can be found (see e.g. Giordano and Cincotta (2004)).

4.2.   Resonant normal form theory: Exponential estimates and the role of convexity.

We hereafter focus on systems of three degrees of freedom of the form

$$H = H_0 + \epsilon H_1$$

which satisfy necessary conditions for the holding of the *theorem of Nekhoroshev* (Nekhoroshev (1977), Benettin et al. (1985), Benettin and Galavotti (1986), Lochak (1992), Pöschel (1993)). The Nekhoroshev theorem ensures that for any initial datum in the action space, the speed of diffusion is bounded by a quantity *exponentially small* in the inverse of the small parameter ϵ, namely one has:

$$|I(t) - I(0)| < \epsilon^a, \quad \text{for all } t < T = O\left(\exp\left(\frac{\epsilon_0}{\epsilon}\right)^b\right) \tag{146}$$

where a, b, ϵ_0 are positive constants.

The basic formulation of the Nekhoroshev theorem relies on a number of assumptions relevant to the phenomenon of Arnold diffusion. In particular, the so-called *geometric part* of Nekhoroshev theorem examines the question of whether there are allowable directions left in the action space along which a chaotic orbit can undergo a fast drift, after all restrictions due to resonant conditions have been properly taken into account in any local domain within the action space. It is then found that additional restrictions to the diffusion are posed if some so-called *convexity* (or, more generally, *steepness*) conditions are fulfilled by the unperturbed part H_0 of the Hamiltonian. We now present a simple form of such conditions.

Analyticity and convexity conditions. We assume that the Hamiltonian function satisfies the following conditions:

Analyticity: We assume that:

i) there is an open domain $\mathcal{I} \subset \mathbf{R}^3$ and a positive number ρ such that for all points $I_* \equiv (I_{1*}, I_{2*}, I_{3*}) \in \mathcal{I}$ and all complex quantities $I'_i \equiv I_i - I_{i*}$ satisfying the inequalities $|I'_i| < \rho$, the function H_0 can be expanded as a convergent Taylor series

$$H_0 = H_{0*} + \omega_* \cdot I' + \sum_{i=1}^{3}\sum_{j=1}^{3} \frac{1}{2} M_{ij*} I'_i I'_j + \dots \tag{147}$$

where $\omega_* = \nabla_I H_0(I_*)$, while M_{ij*} are the entries of the Hessian matrix of H_0 at I_*, denoted by M_*.

ii) For all $I \in \mathcal{I}$, H_1 admits an absolutely convergent Fourier expansion

$$H_1 = \sum_k h_k(I) \exp(ik \cdot \phi) \tag{148}$$

in a domain where all three angles satisfy $0 \leq Re(\phi_i) < 2\pi$, $|Im(\phi_i)| < \sigma$ for some positive constant σ. We recall that, exactly as in subsection 3.2, the latter condition implies that the size of the Fourier coefficients h_k is bounded by a quantity decaying exponentially with $|k|$, i.e. the inequality (114) is satisfied.

iii) We finally assume that all coefficients h_k can be expanded around I_*

$$h_k = h_{k*} + \nabla_{I_*} h_k \cdot I' + \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} h_{k,ij*} I'_i I'_j + \dots \qquad (149)$$

and that the series (149) are convergent in the same union of domains as the series (147) for H_0. [40]

Convexity: We assume that for all $I_* \in \mathcal{I}$ either two of the (real) eigenvalues of the Hessian matrix M_* have the same sign and one is equal to zero, or all three eigenvalues have the same sign. It is noted that the fact that the eigenvalues are real follows from the property that M_* is a symmetric matrix.

Multiplicity of resonances: We now give some definitions allowing to characterize the multiplicity of resonant dynamics. As in subsection 3.3, we define a *resonant manifold* (denoted hereafter by $\mathcal{R}_k$) associated with a non-zero wavevector k with co-prime integer components $k \equiv (k_1, k_2, k_3)$ as the two-dimensional surface in the action space defined by the relation

$$\mathcal{R}_k = \{I \in \mathcal{I} : k_1 \omega_1(I) + k_2 \omega_2(I) + k_3 \omega_3(I) = 0\} \quad , \qquad (150)$$

where $\omega_i(I) = \partial H_0 / \partial I_i$.

Let $I_* \in \mathcal{I}$ be such that all three frequencies $\omega_i(I_*)$, $i = 1, 2, 3$ are different from zero. We distinguish the following three cases:

i) *Non-resonance:* no resonant manifold $\mathcal{R}_k$ contains I_*.

ii) *Simple resonance:* one resonant manifold $\mathcal{R}_k$ contains I_*.

iii) *Double resonance:* more than one resonant manifolds contain I_*. In the latter case, it is possible to choose two linearly independent vectors $k^{(1)}, k^{(2)}$ such that all resonant manifolds $\mathcal{R}_k$ containing I_* are labeled by vectors k which are linear combinations of the chosen vectors $k^{(1)}, k^{(2)}$ with rational coefficients. The intersection of these manifolds forms a one-dimensional *resonant junction*. A doubly-resonant point I_* corresponds to the intersection of a resonant junction with a constant energy surface $H_0(I_*) = E$.

Normal form computation

In computing a resonant normal form for a system with the above properties, we follow a similar method as in subsection 3.3. Namely:

i) We define an *expansion center*, i.e. a point I_* in the action space around which all quantities are expanded. The point I_* must be chosen so that:

a) the orbits whose study we are interested in should reside in a neighborhood of I_*, and

b) This neighborhood should belong to the domain of analyticity of the *optimal* normal form series.

_____

[40] Actually, a more precise statement is that we assume that the intersection of all the domains where the series (147) and (149) are convergent is a non-empty open set $\mathcal{I}$.

Both conditions require some further clarification:

a) Let us consider the left part of Fig.13a. The plotted chaotic orbit undergoes a diffusion within the separatrix chaotic layer along the resonance $I_1 - 2I_2 = 0$. We seek to properly choose a value for the actions I_* in order to compute a resonant normal form in this domain, i.e., well before the orbit approaches the resonant junction at the center of the same figure. For many practical reasons, it proves convenient to choose I_* in such a way that the corresponding frequencies ω_* *always satisfy a doubly-resonant condition*, i.e. there are two linearly independent conditions of the form $k^{(i)} \cdot \omega_* = 0$, $i = 1, 2$. For reasons clarified immediately below, we can see that, provided that $|k^{(2)}| >> |k^{(1)}|$, i.e. that the second resonant wave-vector $k^{(2)}$ corresponds to a resonance of order *much higher* than the order of the first resonant wave-vector $k^{(1)}$, the so-resulting normalization will effectively correspond to a simply resonant normal form, even if, formally, I_* is a doubly-resonant point (see subsection 4.3 below). On the other hand, if $|k^{(2)}|$ and $|k^{(1)}|$ are of similar order, the so resulting normal form is found to represent all features of dynamics in a doubly-resonant domain (subsection 4.4).

Let us give an example: in the case of the Hamiltonian (141) we have $\omega_{1*} = I_{1*}$ and $\omega_{2*} = I_{2*}$, while $\omega_{3*} = 1$. Returning to Fig.13a, if, looking to the leftmost part of the plot, we choose $I_{1*} = 0.355$, $I_{2*} = 0.1775$, we find that the two independent resonance conditions satisfied for this choice, and corresponding to the minimum possible $|k^{(1)}|$ and $|k^{(2)}|$ are: a) $I_{1*} - 2I_{2*} = 0$, corresponding to $k^{(1)} = (1, -2, 0)$, $|k^{(1)}| = 3$, and b) $200I_{1*} - 71 = 0$, corresponding to $k^{(2)} = (200, 0, -71)$, $|k^{(2)}| = 271$. Hence, we have $|k^{(2)}| >> |k^{(1)}|$. In subsection 4.3 we will see that the normal form constructed by such a choice of expansion center effectively describes simply resonant dynamics, despite the fact that the point I_* is formally doubly-resonant, because trigonometric terms of only the resonant angle $k^{(1)} \cdot \phi$ are present in the final normal form formula. On the other hand, as the orbit moves along the resonance in the right direction in Fig.13a, we find that at some later time the orbit passes from the neighborhood of the point $I_{1*} = 4/11 = 0.3636...$, $I_{2*} = 2/11 = 0.1818...$. Now, this is also a doubly-resonant point with $k^{(1)} = (1, -2, 0)$, and $k^{(2)} = (4, 3, -2)$. Thus, $|k^{(2)}| = 9$, i.e. $|k^{(2)}|$ is now much smaller than in the previous example (although still larger than $|k^{(1)}|$). When computing the normal form, we find that for ϵ sufficiently small the so resulting formula contains trigonometric terms of both resonant angles $k^{(1)} \cdot \phi$ and $k^{(2)} \cdot \phi$, thus, it accounts for phenomena due to the double resonance condition. In fact, in this case the effect of the second resonance is visible even numerically, since in Fig.13a we distinguish a thin chaotic layer intersecting transversally the basic (2:1) the main guiding resonance. Finally, as the orbit continues to move rightwards, the orbit approaches, and eventually enters into the conspicuous doubly-resonant domain at the center of Fig.(13a). Then, the orbit exhibits the most prominent effects due to the double resonance. Such effects can be accounted for by a third choice of expansion center, namely $I_{1*} = 0.4$, $I_{2*} = 0.2$, leading to $k^{(2)} = (2, 1, -1)$ and $|k^{(2)}| = 4$ (while $|k^{(1)}| = 3$ always). We observe that both resonant vectors are now of quite low order. Then, it turns out that the resulting normal form contains always trigonometric terms of both resonant angles $k^{(1)} \cdot \phi$ and $k^{(2)} \cdot \phi$ independently of the value of ϵ. In fact, the so resulting normal form can explain the features of the chaotic

motion within the doubly resonant domain seen e.g. in the example of the orbit of Fig.13a.

b) In the above discussion, we saw that different choices of I_* must be made in order to describe the dynamics locally and in different time intervals, i.e. as the orbit gradually drifts from left to right along the resonance $I_1 - 2I_2 = 0$. However, a question arises: in order that the normal form associated with each choice of I_* provides a valid description for the numerical data set of an orbit, we must guarantee that all data points of the orbit considered are contained within the analyticity domain of the normalized Hamiltonian $H^{(r)}$ around I_*, where r is the maximum normalization order considered in the normal form calculation. The question is how to perform this check. A precise description of the techniques by which we specify the analyticity domains of a normal form construction at successive normalization steps is too technical to be presented here. However, we can mention some basic tools used in such specification (the reader may consult Giorgilli (2002) for a more advanced but still quite pedagogical presentation). Briefly, in order to check that a function $f(\phi, J)$ is analytic in a complexified domain of its arguments

$$\mathcal{W}_{\rho,\sigma} = \{(\phi, J) : J \in \mathbf{C}^n : |J_i| < \rho, \phi \in \mathbf{T}_\sigma^n\}$$

where n is the number of degrees of freedom, we compute the so-called *Fourier-weighted norm*

$$||f(\phi, J)||_{\rho,\sigma} = \sum_k \sup |f_k(J)|_\rho e^{|k|\sigma} \tag{151}$$

were f_k are the coefficients of the Fourier development of $f(\phi, J)$, i.e.

$$f_k(J) = \frac{1}{(2\pi)^n} \int_0^{2\pi} \int_0^{2\pi} \ldots \int_0^{2\pi} f(\phi, J) e^{-ik\cdot\phi} d^n\phi$$

and $\sup|\cdot|_\rho$ denotes the supremum of a quantity in the union of the domains $|J_i| < \rho$. The criterion whether the function $f(\phi, J)$ is analytic in a domain $\mathcal{W}_{\rho,\sigma}$ is that Fourier-weighted norm (151) should have a finite upper bound, i.e. $||f(\phi, J)||_{\rho,\sigma} < \infty$.

It is now possible to show that in normal form theory, at every normalization step, the size of the domain of analyticity of both the transformed Hamiltonian and the normalizing transformation series in general *decreases* with respect to the previous step. As a consequence, we have to check that the points of any particular orbit that we intend to study via a normal form continue to remain within the domain of analyticity of these functions at least until a calculation up to the so-called optimal order (see below, and also subsection 2.7), where the normalization can stop. In fact, since at every step we pass from old to new canonical variables, one has to find also the canonical transformations allowing to see how a final domain of analyticity, computed in the new variables, transforms when mapped back to the old variables at which the numerical orbits are computed. Without giving more details, we mention that these tests are necessary ingredients of any study and/or computer-algebraic program in the framework of normal form calculations.

ii) *Resonant module:* After choosing the resonant vectors $k^{(1)}$ and $k^{(2)}$ so that $|k^{(1)}| + |k^{(2)}|$ is minimal, the resonant module can be defined by

$$\mathcal{M} \equiv \left\{ k \in \mathcal{Z}^3 : k \cdot m = 0 \right\} \tag{152}$$

where the vector $m \equiv (m_1, m_2, m_3)$ is defined by the relations

$$m_1 = k_2^{(1)} k_3^{(2)} - k_2^{(2)} k_3^{(1)}, \ \ m_2 = k_3^{(1)} k_1^{(2)} - k_3^{(2)} k_1^{(1)}, \ \ m_3 = k_1^{(1)} k_2^{(2)} - k_1^{(2)} k_2^{(1)} \ . \tag{153}$$

If m_1, m_2, m_3 are not co-prime integers, we re-define m by dividing the m_i by their maximal common divisor. One sees immediately that by the above definition m is a vector parallel to ω_* (m is sometimes called the 'pseudo-frequency' vector).

Action re-scaling and book-keeping: We implement the same action rescaling as in subsection 3.3, namely

$$J_i = \epsilon^{-1/2}(I_i - I_{i*}) = \epsilon^{-1/2} I_i', \quad i = 1, 2, 3 \tag{154}$$

and work with the Hamiltonian function $h(J, \phi) = \epsilon^{-1/2} H(I_* + \epsilon^{1/2} J, \phi)$ producing the correct equations of motion in the re-scaled action variables. Also, in order to separate the Fourier terms in groups of different order of smallness we use again Eq.(115) rewritten here as

$$K' = \left[-\frac{1}{2\sigma} \ln(\epsilon) \right] \tag{155}$$

to define an average value for K' in the domain of interest for the values of ϵ under consideration.

After all previous steps, we finally implement the book-keeping formula

$$\begin{aligned} h &= \omega_* \cdot J + \lambda \epsilon^{1/2} \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{1}{2} M_{ij*} J_i J_j + \ldots + \sum_k \left(\lambda^{1+[|k|/K']} \epsilon^{1/2} h_{k*} \right. \\ &+ \left. \lambda^{2+[|k|/K']} \epsilon \nabla_I h_k \cdot J + \lambda^{3+[|k|/K']} \frac{\epsilon^{3/2}}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} h_{k,ij*} J_i J_j + \ldots \right) \exp(ik \cdot \phi) \ . \end{aligned} \tag{156}$$

Setting $Z_0 = \omega_* \cdot J$, the Hamiltonian (156) takes the form

$$h = H^{(0)}(J, \phi) = Z_0 + \sum_{s=1}^{\infty} \lambda^s H_s^{(0)}(J, \phi; \epsilon^{1/2}) \tag{157}$$

where the superscript (0) denotes, as usually, the original Hamiltonian, and the functions $H_s^{(0)}$ are given by

$$H_s^{(0)} = \sum_{\mu=1}^{s} \epsilon^{\mu/2} \sum_{k=K'(s-\mu)}^{K'(s-\mu+1)-1} H_{\mu,k}^{(0)}(J) \exp(ik \cdot \phi) \tag{158}$$

where $H_{\mu,k}^{(0)}(J)$ are polynomials containing terms of degree $\mu - 1$ or μ in the action variables J. Precisely, we have:

$$H_{\mu,k}^{(0)}(J) = \sum_{\mu_1=0}^{\mu-1} \sum_{\mu_2=0}^{\mu-1-\mu_1} \sum_{\mu_3=0}^{\mu-1-\mu_1-\mu_2} \frac{1}{\mu_1!\mu_2!\mu_3!} \frac{\partial^{\mu-1} h_{1,k}(I_*)}{\partial^{\mu_1} I_1 \partial^{\mu_2} I_2 \partial^{\mu_3} I_3} J_1^{\mu_1} J_2^{\mu_2} J_3^{\mu_3}$$

if $|k| > 0$, or

$$H_{\mu,k}^{(0)}(J) = \sum_{\mu_1=0}^{\mu} \sum_{\mu_2=0}^{\mu-\mu_1} \sum_{\mu_3=0}^{\mu-\mu_1-\mu_2} \frac{1}{\mu_1!\mu_2!\mu_3!} \frac{\partial^{\mu} H_0(I_*)}{\partial^{\mu_1} I_1 \partial^{\mu_2} I_2 \partial^{\mu_3} I_3} J_1^{\mu_1} J_2^{\mu_2} J_3^{\mu_3}$$

$$+ \sum_{\mu_1=0}^{\mu-1} \sum_{\mu_2=0}^{\mu-1-\mu_1} \sum_{\mu_3=0}^{\mu-1-\mu_1-\mu_2} \frac{1}{\mu_1!\mu_2!\mu_3!} \frac{\partial^{\mu-1} h_{1,0}(I_*)}{\partial^{\mu_1} I_1 \partial^{\mu_2} I_2 \partial^{\mu_3} I_3} J_1^{\mu_1} J_2^{\mu_2} J_3^{\mu_3}$$

if $k = 0$.

Hamiltonian normalization: The Hamiltonian normalization is performed by the usual recursive equation:

$$H^{(r)} = \exp(L_{\chi_r}) H^{(r-1)} \qquad (159)$$

where χ_r is the r-th step generating function defined by the homological equation

$$\{\omega_* \cdot J^{(r)}, \chi_r\} + \lambda^r \tilde{H}_r^{(r-1)}(J^{(r)}, \phi^{(r)}) = 0 \qquad (160)$$

and $\tilde{H}_r^{(r-1)}(J^{(r)}, \phi^{(r)})$ denotes all terms of $H^{(r-1)}$ which do not belong to the resonant module $\mathcal{M}$.

Remainder and optimal normalization order: After r normalization steps, the transformed Hamiltonian $H^{(r)}$ has the form

$$H^{(r)}(\phi, J) = Z^{(r)}(\phi, J; \lambda, \epsilon) + R^{(r)}(\phi, J; \lambda, \epsilon) \qquad (161)$$

where $Z^{(r)}(J^{(r)}, \phi^{(r)}; \lambda, \epsilon)$ and $R^{(r)}(J^{(r)}, \phi^{(r)}; \lambda, \epsilon)$ are the normal form and the remainder respectively. The normal form is a finite expression which contains terms up to order r in the book-keeping parameter λ, while the remainder is a convergent series containing terms of order λ^{r+1} and beyond.

In a similar way as in subsection 2.7, where we discussed the asymptotic character of the series for the simple model (3), it is now possible to see that the above normalization process has also an asymptotic character. Namely, i) the domain of convergence of the remainder series $R^{(r)}$ shrinks as the normalization order r increases, and ii) the size $||R^{(r)}||$ of $R^{(r)}$, where $|| \cdot ||$ is a properly defined norm in the space of trigonometric polynomials, initially decreases, as r increases, up to an optimal order r_{opt} beyond which $||R^{(r)}||$ increases with r. In the so-called *Nekhoroshev regime*, one has $||Z^{(r_{opt})}|| >> ||R^{(r_{opt})}||$. Thus, stopping at r_{opt} best unravels the dynamics, which is given essentially by the Hamiltonian flow of $Z^{(r_{opt})}$ slightly perturbed by $R^{(r_{opt})}$.

A basic result, now, of Nekhoroshev theory is that the optimal normalization order r_{opt} depends on ϵ via an inverse power-law namely

$$r_{opt} \sim \epsilon^{-a} \quad , \tag{162}$$

for some positive exponent a. A heuristic derivation of this basic result is given in the Appendix. Taking into account that the leading terms in the remainder are $O(\lambda^{r_{opt}+1})$, while, due to Eq.(155), the book-keeping itself was determined so as to split terms in groups corresponding to powers of the quantity $e^{-\sigma K'} \sim \epsilon^{1/2}$, we conclude that the size of the remainder can be estimated as of order $O(\epsilon^{(r_{opt}+1)/2})$, implying (viz.Eq.(155)):

$$||R^{(r_{opt})}|| \sim \epsilon^{1/2} \exp\left(\frac{-K'\sigma}{\epsilon^a}\right) \tag{163}$$

i.e. *the remainder at the optimal normalization order is exponentially small in $1/\epsilon$.*

The Fourier order

$$K_{opt}(\epsilon) = K' r_{opt}(\epsilon) \tag{164}$$

is called the optimal K–truncation order. All the normal form terms of $H^{(r_{opt})}$ are of Fourier order $0 \le |k| < K_{opt}(\epsilon)$.

The role of convexity in resonant dynamics

We now examine in detail the dynamics induced by the combined effects of the normal form and of the remainder in the above resonant normal form construction.

To this end, let us recall, first, that the normal form at the optimal normalization order r_{opt} is a function of the new canonical variables $\phi^{(r_{opt})}, J^{(r_{opt})}$, which are near-identity transformations of the old canonical variables (ϕ, J). [41] The most important terms in the normalized Hamiltonian are i) the lowest order terms independent of the angles, and ii) the resonant terms, depending on the angles via linear combinations of the resonant arguments $k^{(1)} \cdot \phi^{(r_{opt})}$, or $k^{(2)} \cdot \phi^{(r_{opt})}$. The Hamiltonian takes the form

$$h(J^{(r_{opt})}, \phi^{(r_{opt})}) = \omega_* \cdot J^{(r_{opt})} + \epsilon^{1/2} \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{1}{2} M_{ij*} J_i^{(r_{opt})} J_j^{(r_{opt})} + \ldots \tag{165}$$

$$+ \epsilon^{1/2} \sum_{n_1, n_2 \in Z^2} g_{n_1, n_2}(J^{(r_{opt})}) \exp(i(n_1 k^{(1)} + n_2 k^{(2)}) \cdot \phi^{(r_{opt})}) + \ldots$$

$$+ R_{nonres}(J^{(r_{opt})}, \phi^{(r_{opt})})) \quad .$$

---

[41] The reader should be reminded at this point of the convention used throughout this tutorial, that was mentioned in subsection 2.4, namely that we drop superscripts of the form (r) from all symbols referring to canonical variables just for the purpose of simplifying notation, bearing, however, in mind, that in any Hamiltonian function denoted as $H^{(r)}(\phi, J)$ the arguments are the transformed variables, i.e. $(\phi, J) \equiv (\phi^{(r)}, J^{(r)})$ after r consecutive normalization steps. Here, however, we restore the original notation because both the original and the transformed variables appear in some formulae.

The term $R_{nonres}(J^{(r_{opt})}, \phi^{(r_{opt})}))$, called the *non-resonant remainder*, contains all terms depending on the angles via combinations different from the resonant ones, i.e. $k^{(1)} \cdot \phi^{(r_{opt})}$, or $k^{(2)} \cdot \phi^{(r_{opt})}$.

As in subsection 3.3, let us introduce, now, resonant canonical action - angle variables. We consider the canonical transformation

$$(J_1^{(r_{opt})}, J_2^{(r_{opt})}, J_3^{(r_{opt})}, \phi_1^{(r_{opt})}, \phi_2^{(r_{opt})}, \phi_3^{(r_{opt})}) \rightarrow (J_{R_1}, J_{R_2}, J_F, \phi_{R_1}, \phi_{R_2}, \phi_F)$$

defined by

$$
\begin{aligned}
J_1^{(r_{opt})} &= k_1^{(1)} J_{R_1} + k_1^{(2)} J_{R_2} + m_1 J_F \\
J_2^{(r_{opt})} &= k_2^{(1)} J_{R_1} + k_2^{(2)} J_{R_2} + m_2 J_F \\
J_3^{(r_{opt})} &= k_3^{(1)} J_{R_1} + k_3^{(2)} J_{R_2} + m_3 J_F \\
\phi_{R_1} &= k_1^{(1)} \phi_1^{(r_{opt})} + k_2^{(1)} \phi_2^{(r_{opt})} + k_3^{(1)} \phi_3^{(r_{opt})} \\
\phi_{R_2} &= k_1^{(2)} \phi_1^{(r_{opt})} + k_2^{(2)} \phi_2^{(r_{opt})} + k_3^{(2)} \phi_3^{(r_{opt})} \\
\phi_F &= m_1 \phi_1^{(r_{opt})} + m_2 \phi_2^{(r_{opt})} + m_3 \phi_3^{(r_{opt})}
\end{aligned}
\tag{166}
$$

where $m \equiv (m_1, m_2, m_3)$ has been defined in Eq.(153). The Hamiltonian takes the form

$$
h(J_{R_1}, J_{R_2}, J_F, \phi_{R_1}, \phi_{R_2}) = (\omega_* \cdot m) J_F
$$
$$
+ \epsilon^{1/2} \sum_{i,j=1}^{3} \frac{1}{2} M_{ij*} (k_i^{(1)} J_{R_1} + k_i^{(2)} J_{R_2} + m_i J_F)(k_j^{(1)} J_{R_1} + k_j^{(2)} J_{R_2} + m_j J_F)
$$
$$
+ \epsilon^{1/2} \sum_{n_1, n_2 \in Z^2} g_{n_1, n_2}(J_{R_1}, J_{R_2}, J_F) \exp(i(n_1 \phi_{R_1} + n_2 \phi_{R_2})) + \dots \tag{167}
$$
$$
+ R_{nonres}(J_{R_1}, J_{R_2}, J_F, \phi_{R_1}, \phi_{R_2}) \ .
$$

If we neglect the remainder term R_{nonres}, the constant-valued action J_F is an integral of motion of the Hamiltonian flow of (167). On the other hand, the terms

$$
Z_0(J_{R_1}, J_{R_2}; J_F) = (\omega_* \cdot m) J_F \tag{168}
$$
$$
+ \epsilon^{1/2} \sum_{i,j=1}^{3} \frac{1}{2} M_{ij*} (k_i^{(1)} J_{R_1} + k_i^{(2)} J_{R_2} + m_i J_F)(k_j^{(1)} J_{R_1} + k_j^{(2)} J_{R_2} + m_j J_F)
$$
$$
+ \dots
$$

define an 'integrable part' of the Hamiltonian (167), while the remaining terms depending on the resonant angles, which are of order at least $\epsilon^{1/2}$, can be considered as a perturbation. It is now possible to show the following: *Due to the convexity conditions satisfied by the original Hamiltonian, the constant energy condition $Z_0 = E_Z$, for various values of E_Z, defines a set of invariant ellipses on the plane of the variables J_{R_1}, J_{R_2}. Furthermore, it will be shown that if the effects of the remainder are omitted, any chaotic orbit in the considered resonant*

domain, moving under the normal form Hamiltonian flow alone, stays confined on one ellipse defined for a particular value of E_Z (which remains constant under the normal form dynamics).

This last remark is of crucial importance for understanding the geometric consequences of the convexity condition on the way by which the diffusion of the chaotic orbits proceeds in the action space. In fact, the existence of an invariant ellipse of the normal form Hamiltonian flow (neglecting the remainder effects) prevents the chaotic orbits from diffusing in directions of the action space normal to the ellipse. For instance, when reaching the doubly resonant chaotic domain at the center of Fig.13a, the chaotic orbit appears for a long time to be nearly completely confined along a circular arc. This arc is part of an invariant ellipse of the form $Z_0 = E_Z$ corresponding to the particular resonances into play in this example. In fact, as shown below the orbit can only escape from this type of motion due to remainder effects, as will be shown below. Thus, *the size of the remainder determines the rate at which the diffusion of chaotic orbits can proceed.* This property holds in the neighborhood of every resonance junction in the action space of a system satisfying convexity conditions similar to the one set at the beginning of the present subsection. But since the doubly resonant points are dense in the action space, the entire chaotic motion in the resonance web can be modeled as a sequence of crossings of different resonant junctions, where, in each junction, a local resonant normal form construction can be made leading to a set of invariant ellipses as above.

We now show why the convexity conditions of the original Hamiltonian imply that the relation of the form $Z_0 = E_Z = const$ defines an invariant ellipse in the plane (J_{R_1}, J_{R_2}). This is equivalent to showing that the quadratic form

$$\zeta_{0,2} = \frac{1}{2} \sum_{i,j=1}^{3} M_{ij*}(k_i^{(1)} J_{R_1} + k_i^{(2)} J_{R_2})(k_j^{(1)} J_{R_1} + k_j^{(2)} J_{R_2}) \qquad (169)$$

is positive definite. The latter can be written as

$$\zeta_{0,2} = (J_{R_1}, J_{R_2}) \cdot k^{(1,2)} \cdot M_* \cdot (k^{(1,2)})^T \cdot (J_{R_1}, J_{R_2})^T \qquad (170)$$

where $k^{(1,2)}$ is a 2×3 matrix whose first and second line are given by $(k_1^{(1)}, k_2^{(1)}, k_3^{(1)})$ and $(k_1^{(2)}, k_2^{(2)}, k_3^{(2)})$ respectively. Since the matrix M_* is real symmetric, it can be written in the form $M_* = X \cdot \mu_* \cdot X^T$, where $\mu_* = diag(\mu_1, \mu_2, \mu_3)$, with $\mu_i =$ the eigenvalues of M_*, while X is an orthogonal matrix with columns equal to the normalized eigenvectors of M_*. Using the above expression for M_*, Eq.(170) resumes the form

$$\zeta_{0,2} = (J_{R_1}, J_{R_2}) \cdot Y \cdot \mu_* \cdot Y^T (J_{R_1}, J_{R_2})^T$$

where $Y = k^{(1,2)} \cdot X$ is a 2×3 matrix. Writing Z_{02} as $Z_{02} = Q J_{R_1}^2 + V J_{R_1} J_{R_2} + P J_{R_2}^2$, and denoting by y_{ij} the elements of Y, the discriminant $\Delta = 4QP - V^2$ is given by:

$$\begin{aligned} \Delta &= -[(y_{11}y_{22} - y_{12}y_{21})^2 \mu_1 \mu_2 + (y_{11}y_{23} - y_{13}y_{21})^2 \mu_1 \mu_3 \\ &+ (y_{12}y_{23} - y_{13}y_{22})^2 \mu_2 \mu_3] \ . \end{aligned} \qquad (171)$$

However, by the convexity condition assumed at the beginning of this subsection, we have that either all three eigenvalues μ_i have the same sign, or two of them have the same sign and one is zero, by Eq.(171) we have that $\Delta < 0$. That is, the quadratic form $\zeta_{0,2}$ is positive definite.

After demonstrating the geometric consequences of convexity for the chaotic motion along resonances, we now examine separately the chaotic diffusion in the case of *simple resonance*, i.e. when as discussed above, $|k^{(2)}| >> |k^{(1)}|$, and in the case of *double resonance*, i.e. when $|k^{(2)}|$ and $|k^{(1)}|$ are of comparable size.

4.3. Diffusion along simple resonances. Comparison with Chirikov's estimates

In order to implement resonant normal form theory in the case of simply resonant dynamics, we define first local resonant canonical variables around the center of the ellipses $Z_0 = E_Z$, which is given by:

$$
\begin{aligned}
J_{R_1,0} &= \frac{\left(k^{(1)} \cdot M_* k^{(2)}\right)\left(m \cdot M_* k^{(2)}\right) - \left(k^{(2)} \cdot M_* k^{(2)}\right)\left(m \cdot M_* k^{(1)}\right)}{\left(k^{(1)} \cdot M_* k^{(1)}\right)\left(k^{(2)} \cdot M_* k^{(2)}\right) - \left(k^{(1)} \cdot M_* k^{(2)}\right)^2} J_F \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (172) \\
J_{R_2,0} &= \frac{\left(k^{(1)} \cdot M_* k^{(2)}\right)\left(m \cdot M_* k^{(1)}\right) - \left(k^{(1)} \cdot M_* k^{(1)}\right)\left(m \cdot M_* k^{(2)}\right)}{\left(k^{(1)} \cdot M_* k^{(1)}\right)\left(k^{(2)} \cdot M_* k^{(2)}\right) - \left(k^{(1)} \cdot M_* k^{(2)}\right)^2} J_F
\end{aligned}
$$

We then define:

$$
\Delta J_{R_i} = J_{R_i} - J_{R_i,0}, \quad a_{ij} = k^{(i)} \cdot M_* k^{(j)}, \quad i,j = 1,2 \ .
$$

The key remark is that, if we have $|k^{(2)}| >> |k^{(1)}|$, for all values of ϵ of practical interest we also find $|k^{(2)}| > K_{opt}(\epsilon) > |k^{(1)}|$, i.e. *for all values of ϵ of practical interest, all resonant terms containing the angle ϕ_{R_2} are of Fourier order higher than the optimal K–truncation order*. This implies that such terms can be considered as *part of the remainder*. [42]

---

[42] A concrete example is here in order. Let as consider again, as at the beginning of the previous subsection, the resonant normal form construction for $I_{1*} = 0.355$, $I_{2*} = 0.1775$. We saw that in that case $|k^{(1)}| = 3$, while $|k^{(2)}| = 271$. However, if we roughly set $K_{opt} = K' r_{opt} = K'(\epsilon_0/\epsilon)^a$ (in view of Eq.(163)), then we see that if we take $a = 0.25$ (simple resonance, see Eq.(240)), giving some typical values to K' and ϵ_0, e.g. $K' = 3$, $\epsilon_0 \sim 0.1$ we have $K_{opt} < 10^2$ in the range $10^{-4} \leq \epsilon \leq 10^{-1}$. Thus, in the same range, we have $|k^{(2)}| > K_{opt}$, implying that the construction can be considered simply resonant in practically the entire range of values of ϵ that would be encountered in realistic applications (cf. Table 1). We note also that the latter result is not sensitive to the choice of the constants K' and ϵ_0.
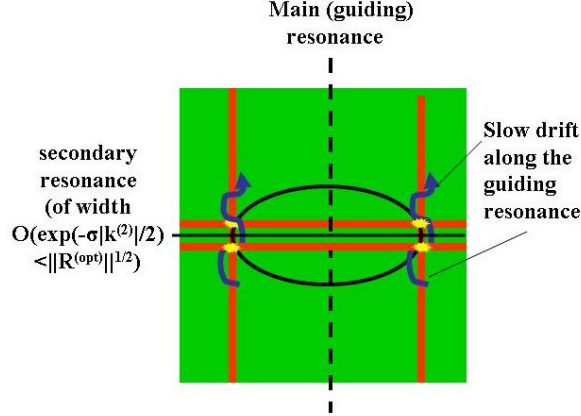
Figure 14.   Schematic representation of the diffusion along a simple resonance. Any resonance crossing transversally the main (guiding) resonance has an exponentially small width and acts as a 'driving' resonance for diffusion.

With the above considerations, the transformed Hamiltonian reads:

$$h(J^{(r_{opt})}, \phi^{(r_{opt})}) = Z(J^{(r_{opt})}, \phi^{(r_{opt})}) + R(J^{(r_{opt})}, \phi^{(r_{opt})})$$

$$= \omega_* \cdot J^{(r_{opt})} + \epsilon^{1/2} \sum_{i=1}^{3} \sum_{j=1}^{3} \frac{1}{2} M_{ij*} J_i^{(r_{opt})} J_j^{(r_{opt})} + \dots \tag{173}$$

$$+ \epsilon^{1/2} \sum_{n \in \mathcal{Z}^*} g_n(J^{(r_{opt})}) \exp(i(nk^{(1)} \cdot \phi^{(r_{opt})}) + \dots$$

$$+ R(J^{(r_{opt})}, \phi^{(r_{opt})}) \quad .$$

Taking into account the canonical transformations (166), after some algebra the Hamiltonian (173) takes the form

$$h \approx (m \cdot \omega_*) J_F + \epsilon^{1/2} \left[\frac{1}{2} a_{11} \Delta J_{R_1}^2 + a_{12} \Delta J_{R_1} \Delta J_{R_2} + \frac{1}{2} a_{22} \Delta J_{R_2}^2 \right. \tag{174}$$

$$\left. + \; 2 f_{R_1} \cos(\phi_{R_1}) + \dots + \sum_{|k| \geq K^{(opt)}} f_{k*} \exp[ik \cdot (\kappa_1 \phi_{R_1} + \kappa_2 \phi_{R_2} + \kappa_3 \phi_F)] + \dots \right]$$

where i) the (non-integer) vectors κ_i, $i = 1, 2, 3$ come from the solution of the right set of Eqs.(166) for the angles $\phi_i^{(r_{opt})}$ in terms of the angles ϕ_{R_1}, ϕ_{R_2}, and ϕ_F, and ii) we approximate all the Fourier coefficients in the remainder series by their constant values f_{k*} at the points $\Delta J_{R_1} = \Delta J_{R_2} = 0$ (we set $f_{R_1} = f_{k*}$ for $k = k^{(1)}$).

We note that, neglecting the non diagonal term $a_{12} \Delta J_{R_1} \Delta J_{R_2}$, the Hamiltonian (174) is of the form discussed at the beginning of the present section, i.e. the Hamiltonian (143).

The 'pendulum' part of the above Hamiltonian is given by

$$
\begin{aligned}
Z_{res} & = \frac{1}{2}a_{11}\Delta J_{R_1}^2 + a_{12}\Delta J_{R_1}\Delta J_{R_2} + \frac{1}{2}a_{22}\Delta J_{R_2}^2 + ... \\
& + \epsilon^{1/2}\left(g_n e^{in\phi_{R_1}} + g_{-n}e^{-in\phi_{R_1}}\right) + \ldots
\end{aligned} \tag{175}
$$

Since the angle ϕ_{R_2} is ignorable, the action J_{R_2} (or ΔJ_{R_2}) is an integral of the flow of Z_{res}, in addition to J_F. Thus, Z_{res} defines an integrable Hamiltonian. A pair of constant values $J_F = c_1$, $\Delta J_{R_2} = c_2$ defines a straight line

$$
\Delta J_{R_1} = -\frac{a_{12}}{a_{11}}c_2 \tag{176}
$$

which corresponds to the unique resonance $\omega_{R_1}(J^{(r_{opt})}) = k^{(1)} \cdot \omega(J^{(r_{opt})}) = 0$. This will be called 'main resonance' (= the 'guiding resonance' in Chirikov's theory; see Chirikov (1979), Cincotta (2002)).

In Figure 14 (schematic), the domain of the main resonance is delimited by two vertical thick red lines corresponding to the separatrix-like thin chaotic layers at the boundary of the resonance similarly to Fig.15.

Under the normal form dynamics, motions are allowed only across the resonance, i.e. in the direction $\Delta J_{R_2} = const$. In Fig.14 this is the horizontal direction. The thin strip delimited by two horizontal red lines corresponds to the resonance with resonant wavevector $k^{(2)}$, which, since $k^{(2)} > K(\epsilon)$, is now of width exponentially small ($O(\epsilon^{1/2}e^{-\sigma|k^{(2)}|/2})$). Thus, it will be called a 'secondary' resonance.

We now ask the following question: since the diffusion along the resonance is only possible because of the remainder influence on dynamics, can we estimate the *speed* of diffusion (or the value of the diffusion coefficient), by knowing the *size of the remainder of the optimal normal form construction*?

We can see that the approximation of Eq.(174) is sufficient for estimates regarding the speed of diffusion. The key remark is that for all the coefficients f_{k*} the bound $|f_{k*}| < ||R_{opt}||$ holds, while, for the leading Fourier term $\exp(ik_l \cdot \phi^{(r_{opt})})$ in the remainder we have $|f_{k_l*}| \sim ||R_{opt}||$. In fact, we typically find that the size of the leading term is larger from the size of the remaining terms by several orders of magnitude, since this term contains a repeated product of small divisors of the form $k_l \cdot \omega_*$ (see Appendix). Furthermore, using an analysis as in Efthymiopoulos et al. (2004), we readily find $|k_l| = (1 - d)K_{opt}$, where $0 < d < 1$ is a so-called (in Efthymiopoulos et al. (2004)) 'delay' constant. We note in passing that the Fourier terms of the form $\exp(ik_l \cdot \phi^{(r_{opt})})$ are called 'resonant' in Morbidelli and Giorgilli (1997).

The value of the diffusion coefficient can now be estimated by applying the basic theory of Chirikov (1979). This theory is reviewed for the purpose of applications in dynamical astronomy by Cincotta (2002), and a recent application in the so-called three body resonance problem in solar system dynamics was given in Cachucho et al. (2010). Briefly, the theory uses the Arnold-Melnikov integral technique to estimate the speed of diffusion along simple resonances by examining the so-called variations in the pendulum energy as a chaotic orbit moves in the separatrix-like chaotic layer of a simply resonant domain. It is found that

the following estimate holds for the diffusion coefficient:

$$D \sim \frac{\epsilon}{2\Omega_G^2 T} |f_{k_l*}|^2 A(|\kappa_l|)^2 \quad . \tag{177}$$

In Eq.(177):

i) $\Omega_G = \epsilon^{1/4} f_{R_1}^{1/2}$, where $T = ln(32e/w)/\Omega_G$ is an average period of motion within the main resonance separatrix-like thin chaotic layer, of width w.

ii) A is the *Melnikov function* with argument $|\kappa_l|$ (see Appendix B of Ferraz-Mello (2007)). The vector κ_l is defined by the relation

$$\kappa_{l,1}\phi_{R_1} + \kappa_{l,2}\phi_{R_2} + \kappa_{l,3}\phi_F = k_l \cdot \phi^{(r_{opt})}$$

and the estimate

$$A(|\kappa_l|) \sim 8\pi|\kappa_l|e^{-\pi|\kappa_l|/2}$$

holds.

The key point in connecting Chirikov's formula (177) with the estimates based on the optimal remainder function is that, in view of Eq.(166), we have that

$$|\kappa_l| = O((1-d)K^{(opt)}/|k^{(1)}|) \quad . \tag{178}$$

However, it was pointed out in subsection 4.2 that the optimal K–truncation order K_{opt} depends on ϵ as an inverse power, i.e. we have $K_{opt} \sim \epsilon^{-1/4}$ in simply resonant domains (see Appendix), yielding also $||R_{opt}|| \sim e^{-\sigma K^{(opt)}}$. Substituting these expressions into Eq.(178) it follows that

$$A(|\kappa_l|) \sim \epsilon^{3/4}||R_{opt}||^b$$

for an exponent $b > 0$. Putting, finally, these estimates together in Eq.(177), we arrive at an estimate of the dependence of the diffusion coefficient D on the optimal remainder $||R_{opt}||$ in the case of simple resonances, namely:

$$D \sim \frac{\epsilon}{2\Omega_G^2 T}\epsilon^{3/4}||R_{opt}||^{2(1+b)} \quad . \tag{179}$$

We emphasize that the precise value of b is an open issue, which appears to be hardly tractable to address on the basis exclusively of the behavior of the Melnikov integrals discussed above. We note, however, that the quantity $A(\kappa_l)$ yields the size of the 'splitting' S of the separatrix of the main (guiding) resonance due to the effects of the leading term in the remainder function. The relation between the separatrix splitting and the size of the optimal remainder has been examined in Neishtadt (1984), and later in Morbidelli and Giorgilli (1997). In the latter work, the estimate $S \sim \mu^{1/2}$ was predicted and probed by numerical experiments, where μ (in the notation of Morbidelli and Giorgilli 1997) is the effective size of the perturbation to the normal form pendulum dynamics caused by the remainder. Setting thus $\mu \sim ||R_{opt}||$ suggests the scaling $A(\kappa_l) \sim S \sim ||R_{opt}||^{1/2}$, whereby the constant b can be estimated as $b \simeq 1/2$. Hence (in view of Eq.(179)) we find

$$D \sim ||R_{opt}||^3$$

in simply resonant domains. In Efthymiopoulos (2008) the diffusion coefficient D along a simple resonance was compared directly to the size of the optimal normal form remainder in a concrete example referring to the numerical data reported in the work of Lega et al. (2003). It was found that $D \propto ||R_{opt}||^{2.98}$, essentially confirming that $p = 2(1 + b) \simeq 3$, or $b \simeq 0.5$. However, in Lega et al. (2010a) a different exponent was found $p \simeq 2.56$ regarding the same resonance, while it was found that $p = 2.1$ in the case of a very low order simple resonance (with $|k^{(1)} < K'|$). These numerically defined exponents, however, depend on the chosen definition of numerical measure used to estimate both S and $||R_{opt}||$. Thus, a detailed quantitative comparison of the various estimates given in the literature is an open problem.

4.4. Diffusion along double resonances

We now pass to analyzing, via a resonant normal form, the properties of weakly-chaotic diffusion when double resonance condition is fulfilled, the latter being defined by the requirement that $|k^{(2)}| < K_{opt}(\epsilon)$, ensuring that both resonances due to $k^{(1)}$ and $k^{(2)}$ are important.

In order to analyze the normal form dynamics in double resonances, we first write the normal form constant energy condition as:

$$E' = (Z_0 - (\omega_* \cdot m)J_F) = const. \tag{180}$$

If higher order terms in the action variables of the development of Eq.(168) are taken into account, the constant energy condition of Eq.(180) yields deformed ellipses on the plane (J_{R_1}, J_{R_2}). If $J_{R_1} \neq J_{R_{1,0}}$ or $J_{R_2} \neq J_{R_{2,0}}$, we can define two slow frequencies for the resonant angles, namely $\dot{\phi}_{R_1} \equiv \omega_{R_1}$, $\dot{\phi}_{R_2} \equiv \omega_{R_2}$. We have

$$\omega_{R_1} = \left(k^{(1)} \cdot M_* k^{(1)}\right)(J_{R_1} - J_{R_{1,0}}) + \left(k^{(1)} \cdot M_* k^{(2)}\right)(J_{R_2} - J_{R_{2,0}}) + \dots$$
$$\tag{181}$$
$$\omega_{R_2} = \left(k^{(1)} \cdot M_* k^{(2)}\right)(J_{R_1} - J_{R_{1,0}}) + \left(k^{(2)} \cdot M_* k^{(2)}\right)(J_{R_2} - J_{R_{2,0}}) + \dots$$

On the other hand, due to the definition (166) one has

$$\omega_{R_1} = k^{(1)} \cdot \omega(J^{(r_{opt})}), \quad \omega_{R_2} = k^{(2)} \cdot \omega(J^{(r_{opt})}), \quad \omega_F = m \cdot \omega(J^{(r_{opt})})$$

which is valid for any value of (J_{R_1}, J_{R_2}, J_F) in the domain of convergence of the series (168). It follows that all the resonant manifolds defined by relations of the form $(n_1 k^{(1)} + n_2 k^{(2)}) \cdot \omega(J^{(r_{opt})}) = 0$ intersect any of the planes (J_{R_1}, J_{R_2}) corresponding to a fixed value of J_F. Using again the notation

$$\Delta J_{R_i} = J_{R_i} - J_{R_{i,0}}, \quad a_{ij} = k^{(i)} \cdot M_* k^{(j)}, \quad i, j = 1, 2$$

the intersection of one resonant manifold with the plane (J_{R_1}, J_{R_2}) is a curve given, in the linear approximation, by

$$(n_1 a_{11} + n_2 a_{12})\Delta J_{R_1} + (n_1 a_{12} + n_2 a_{22})\Delta J_{R_2} + \dots = 0 \quad .$$

The above equation defines a 'resonant line', which is the local linear approx-
imation to a 'resonant curve'. All resonant lines (or curves) pass through the
point $(J_{R_1,0}, J_{R_2,0})$, which, therefore, belongs to the resonant junction defined by
the wavevectors $k^{(1)}, k^{(2)}$. To each resonant curve we can associate a resonant
strip in action space whose width is proportional to the separatrix width for that
resonance. If, for a single pair of integers (n_1, n_2), we only isolate the resonant
terms $g_{\pm n_1, \pm n_2} e^{\pm i(n_1\phi_{R_1} + n_2\phi_{R_2})}$ in the normal form Z (Eq.(165)), we obtain a
simplified resonant normal form $Z_{res(n_1,n_2)}$ corresponding to the limiting case of
a single resonance. In a strict sense, Z_{res} describes well the dynamics far from
the resonant junction. However, it can also be used in order to obtain estimates
of the resonance width along the whole resonant curve defined by the integer pair
(n_1, n_2). To this end, the leading terms of $Z_{res(n_1,n_2)}$ are (apart from constants):

$$
Z_{res(n_1,n_2)} = \epsilon^{1/2}\left[\frac{1}{2}a_{11}\Delta J_{R_1}^2 + a_{12}\Delta J_{R_1}\Delta J_{R_2} + \frac{1}{2}a_{22}\Delta J_{R_2}^2 + ... \right.
$$

$$
\left. + \left(g_{n_1,n_2}e^{i(n_1\phi_{R_1}+n_2\phi_{R_2})} + g_{-n_1,-n_2}e^{-i(n_1\phi_{R_1}+n_2\phi_{R_2})}\right)\right] + \ldots \qquad (182)
$$

The coefficients $g_{\pm n_1, \pm n_2}$ satisfy the estimate

$$
|g_{n_1,n_2}| \approx A e^{-(|n_1||k^{(1)}|+|n_2||k^{(2)}|)\sigma} \quad . \qquad (183)
$$

After still another transformation $\Delta J_{R_1} = n_1 J_R + n_2 J_F$, $\Delta J_{R_2} = n_2 J_R - n_1 J_F$,
$\phi_R = n_1\phi_{R_1} + n_2\phi_{R_2}$, J_F becomes a second integral of motion of $Z_{res(n_1,n_2)}$,
which takes the form

$$
Z_{res(n_1,n_2)} = \epsilon^{1/2}\left[c(J_F) - \frac{1}{2}(a_{11}n_{11}^2 + 2a_{12}n_1 n_2 + a_{22}n_2^2)(J_R - J_{R,0}(J_F))^2\right.
$$

$$
\left. + \left(g_{n_1,n_2}e^{i\phi_R} + g_{-n_1,-n_2}e^{-i\phi_R}\right) + \ldots\right] \qquad (184)
$$

where $c(J_F)$ and $J_{R,0}(J_F)$ are constants of the Hamiltonian flow of (184). Com-
bining (183) and (184), the separatrix width can be estimated as

$$
\Delta J_R \approx \sqrt{\frac{32A e^{-(|n_1||k^{(1)}|+|n_2||k^{(2)}|)\sigma}}{a_{11}n_{11}^2 + 2a_{12}n_1 n_2 + a_{22}n_2^2}} \quad . \qquad (185)
$$

Eq.(185) allows to estimate the width of a resonant strip in the direction normal
to a resonant curve on the plane (J_{R_1}, J_{R_2}). Using the relations $\Delta(\Delta J_{R_i}) =
n_i \Delta J_R$ (for $\Delta J_F = 0$), this estimate takes the form

$$
\Delta J_{R,width} =\approx \left(\frac{32A(n_1^2 + n_2^2)}{a_{11}n_{11}^2 + 2a_{12}n_1 n_2 + a_{22}n_2^2}\right)^{1/2} e^{-\frac{1}{2}(|n_1||k^{(1)}|+|n_2||k^{(2)}|)\sigma} \quad . \quad (186)
$$

The outcome of the above analysis can be visualized with the help of Figure
15 (schematic). The left panel shows the structure of a doubly-resonant domain
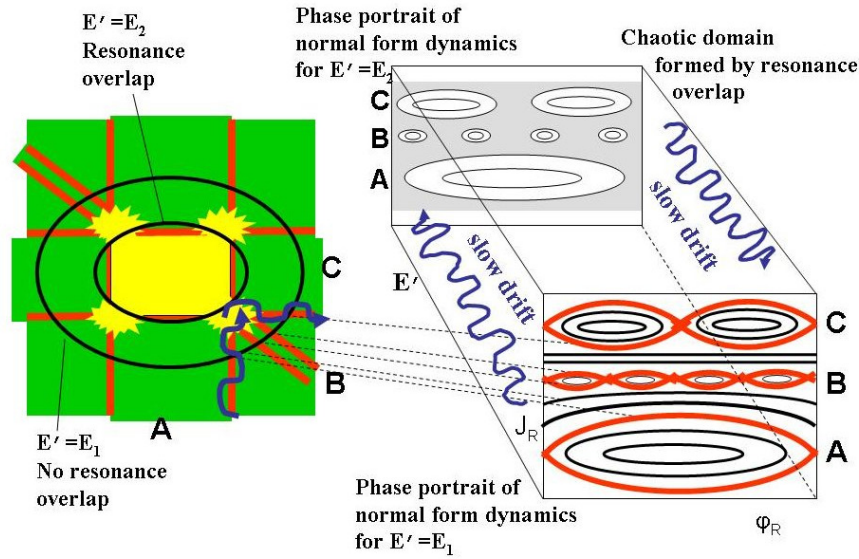
Figure 15. Schematic representation of the normal form and remainder dynamics in a domain of double resonance. Left panel: the resonant structure formed in the action plane of the variables (J_{R_1}, J_{R_2}) by the overlapping of various resonant strips whose limits (pairs of parallel red lines) correspond to separatrix-like thin chaotic domains around each resonance. Two constant normal form energy ellipses $E' = E_1$ and $E' = E_2$ are also shown. Right: The front and back panels show the phase portraits corresponding to a surface of section (in one of the pairs (ϕ_{R_1}, J_{R_1}) or (ϕ_{R_2}, J_{R_2})) under the normal form dynamics alone, for the energies $E' = E_1$ (front panel) and $E' = E_2$ (back panel). The blue curly arrows in both panels indicate the directions of a possible 'drift' motion (=slow change of the value of E') due to the influence of the remainder on dynamics.

in the plane of the resonant action variables (J_{R_1}, J_{R_2}). The two bold ellipses correspond to the constant energy condition for two different values of E', namely $E' = E_1$ and $E' = E_2$ with $E_1 > E_2$. Their common center is the point $(J_{R_1,0}, J_{R_2,0})$ defined in Eq.(172). The three pairs of parallel red lines depict the borders of the separatrix-like thin chaotic layers of three resonances passing through the center. Infinitely many such resonances exist, corresponding to different choices of integer vectors $n \equiv (n_1, n_2)$; however, their width decreases as $|n|$ increases, according to Eq.(186). We thus show schematically only three resonances with a relatively low value of $|n|$, named by the letters 'A', 'B' and 'C'.

With the help of Figure 15, the influence of the normal form terms on dynamics, by considering the Hamiltonian flow under the approximation $H \simeq Z$, can now be understood as follows:

- For any fixed value of E', and a fixed section in the angles, the motion is confined on one ellipse.

- For E' large enough ($E' = E_1$, outermost ellipse in the left panel of Fig.15), the various resonant strips intersect the ellipse $E' = E_1$ at well distinct arcs, i.e. there is no resonance overlap. The right front panel in Fig.15 shows schematically the expected phase portrait, which can be obtained by evaluating an appropriate Poincaré surface of section, e.g. in the variables (J_{R_1}, ϕ_{R_1}) or (J_{R_2}, ϕ_{R_2}). The dashed lines show the correspondence between the limits of various resonant domains depicted in the left and right panels. In particular, the intersection of each resonant strip in the left panel with the ellipse $E' = E_1$ corresponds to the appearance of an associated *island chain* in the right panel. The size of islands is given essentially by the separatrix width estimate of Eq.(186). Hence, the size of the islands decreases exponentially with the order of the resonance $n = |n_1| + |n_2|$. However, the main effect to note is that, since all resonant strips are well separated on the ellipse, the thin separatrix-like chaotic layers marking the borders of each of their respective island chains do not overlap. As a result the local chaos around one resonance is isolated from the local chaos around the other resonances. In fact, the normal form dynamics induces the presence of rotational KAM tori which, in this approximation ($H \simeq Z$), completely obstruct the communication among the resonances.

- While the size of the islands is nearly independent of the energy E', their separation is reduced as the energy *decreases*. Thus, below a critical energy E'_c, significant resonance overlap takes place, leading to the communication of the chaotic layers of the various resonances and an overall increase of chaos. This is shown in the left panel of Fig.15 for an ellipse $E' = E_2 < E'_c$, with the corresponding phase portrait shown in the right back panel. We note in particular the 'melting' of all three resonant domains one into the other, which produces a large connected chaotic domain surrounding all three island chains (and many other smaller chains, not visible in this scale).

The value of the critical energy E'_c marking the onset of large scale resonance overlap can be estimated as follows: Each resonant strip intersects one fixed energy ellipse on one arc segment. Also, Eq.(186) can be replaced by the estimate

$$\Delta J_{R,width} \approx \frac{(32A)^{1/2}}{M_h k_{1,2}} e^{-\frac{1}{2} n k_{1,2} \sigma} \qquad (187)$$

where $n = |n_1| + |n_2|$, $k_{1,2} = (|k^{(1)}| + |k^{(2)}|)/2$, and M_h is a measure of the size of the eigenvalues of the Hessian matrix M_*. The total length S_{res} of all segments can be now estimated by summing, for all n, the estimate (187), namely

$$S_{res} \approx \frac{(32A)^{1/2}}{M_h k_{1,2}} \sum_{n=1}^{\infty} e^{-\frac{1}{2} n k_{1,2} \sigma} \approx \frac{(128A)^{1/2}}{M_h k_{1,2} \sigma} e^{-\frac{1}{2} k_{1,2} \sigma} \quad . \qquad (188)$$

On the other hand, the total circumference of the ellipse for the energy E' is estimated as $S_{E'} = \pi R(E')^2$ where $R(E')$ is the geometric mean of the ellipse's major and minor semi-axes. For $R(E')$ one has the obvious estimate $R(E') \sim (2E'/(\epsilon^{1/2} M_h))^{1/2}$, whence

$$S_{E'} \sim \frac{2\pi E'}{\epsilon^{1/2} M_h} \quad . \qquad (189)$$

The critical energy $E' = E'_c$ can now be estimated as the value where $S(E') \approx S_{res}$, implying that the associated ellipse is fully covered by segments of resonant strips. Thus

$$E'_c \approx \frac{32(\epsilon A)^{1/2}}{\pi k_{1,2}\sigma} e^{-\frac{1}{2}k_{1,2}\sigma} \quad . \tag{190}$$

Eq.(190) implies that E'_c is a $O(\epsilon^{1/2}e^{-\frac{1}{2}k_{1,2}\sigma})$ quantity.

So far, we have neglected the role of the remainder in dynamics. In Fig.15, the drift in action space caused by the remainder is shown schematically by the blue curly curves in both the left and right panels. Their significance is the following: The energy $E = h$ corresponding to the total Hamiltonian $h = Z + R^{(r_{opt})}$ of Eq.(227) is an exactly preserved quantity. Thus, the doubly-resonant normal form energy E' as well as J_F cannot be preserved exactly, but they are approximate integrals, i.e. they undergo time variations bounded by an $O(||R^{(r_{opt})}||)$ quantity. In Fig.15, such variations will in general lead to a very slow change of the value of E', i.e. a very slow drift of the chaotic orbits from one ellipse to another. We seek to estimate the time required for the remainder to induce a transition between two ellipses with an energy difference of the same order as E'_c, namely

$$E'_2 - E'_1 = O(\epsilon^{1/2}e^{-\frac{1}{2}k_{1,2}\sigma}) \tag{191}$$

assuming that this effect can be described as a *random walk* in the value of E'. Let T be an average period of the oscillations of the resonant variables. By Eqs.(182) and (183), the estimate $T \sim (\epsilon A)^{-1/2}e^{n_{eff}k_{1,2}\sigma/2}$ holds, for a constant $n_{eff} \sim 1$ marking the order of the most important resonances in (182). In consecutive steps, dE' can be either positive or negative, while its typical size is $|dE'| \sim ||R_{opt}||$. Then, after N steps of a random walk (in the values of E'), we find an rms spread of these values given by

$$\Delta E \approx N^{1/2}||R_{opt}|| \tag{192}$$

Using (191) and (192), the number of steps required for the spread ΔE to become equal to $E'_2 - E'_1$ (given by (191) is $N \sim \epsilon e^{-n_{eff}k_{1,2}\sigma}||R_{opt}||^{-2}$. The diffusion coefficient can be estimated as

$$D \sim \frac{\Delta E^2}{NT} \sim \left(\epsilon A e^{-n_{eff}k_{1,2}\sigma}\right)^{1/2}||R_{opt}||^2 \tag{193}$$

i.e. the diffusion coefficient scales as the square of the size of the optimal remainder function.

Visualization of Arnold diffusion using normal forms

The possibility to use the resonant variables (ϕ_{R_1}, J_{R_1}), as well as the normal form energy E' in order to study the phenomenon of Arnold diffusion in systems satisfying the necessary conditions for the holding of the Nekhoroshev theorem, was first pointed out in the work of Benettin and Gallavotti (1986). Recently, however, it has become possible to obtain *real* (non-schematic) examples of this visualization, by computing a doubly-resonant normal form at a quite high order. More precisely, in Efthymiopoulos and Harsoula (2012) we have given a detailed
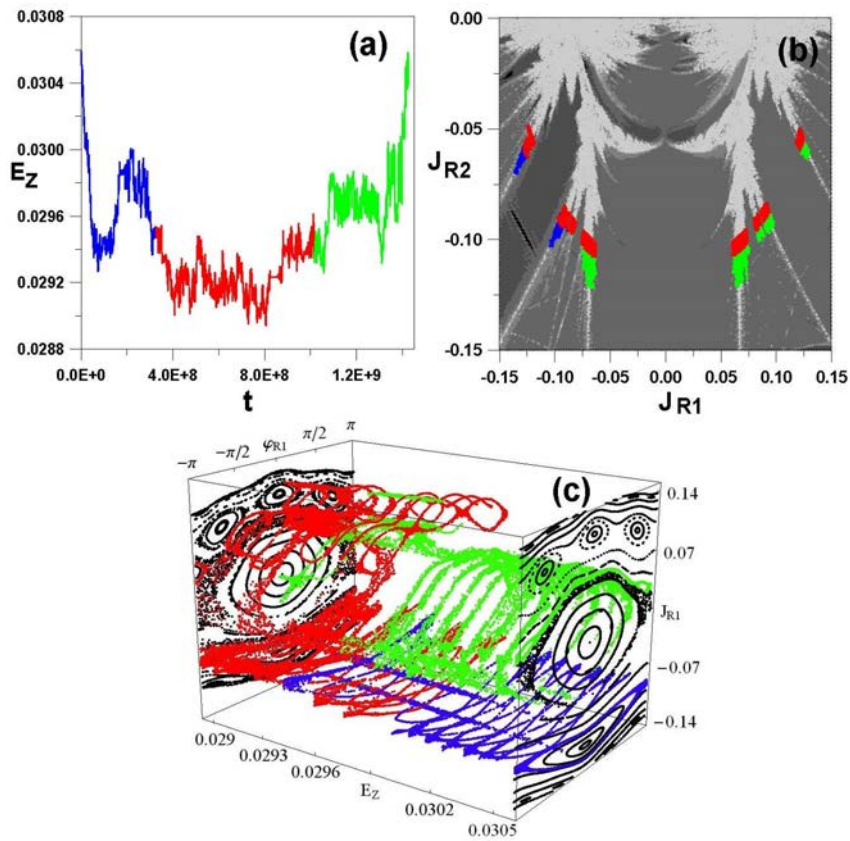
Figure 16.   Visualization of Arnold diffusion in a numerical example of high-order doubly resonant normal form computation in the Hamiltonian model (141) (After Efthymiopoulos and Harsoula (2012), see text for details).

numerical example of the above, using the Hamiltonian model (141) in the case of the double resonance defined by $I_{1*} = 0.4$, $I_{2*} = 0.2$. Figure 16 shows the visualization of Arnold diffusion in the above model in appropriate variables of the doubly-resonant normal form, for a numerical orbit in the Hamiltonian (141) for $\epsilon = 0.008$. After computing the optimal normal form, we find, via the Lie canonical transformations, the values of all transformed variables $J_{R_1}(t), J_{R_2}(t), J_F(t)$ and $\phi_{R_1}(t), \phi_{R_2}(t), \phi_F(t)$ corresponding to particular values of the old variables $J_1(t), J_2(t), J_3(t)$ and $\phi_1(t), \phi_2(t), \phi_3(t)$ stored at many different times t within an interval $0 \leq t \leq 1.5 \times 10^9$ along the numerical run. Using the numerical values of the computed transformed variables, (a) shows the variation of the normal form energy $E_Z(t)$ as a function of t in the intervals $0 \leq t \leq 3 \times 10^8$ (blue), $3 \times 10^8 \leq t \leq 10^9$ (red), and $10^9 \leq t \leq 1.5 \times 10^9$ (green). The initial and final values are equal to $E_Z(t = 0) = E_Z(t = 1.5 \times 10^9) = 0.0306$, while the minimum value, occurring around $t = 8 \times 10^8$ is $E_Z = 0.029$. On the other hand, the evolution of the orbit in the action space (J_{R_1}, J_{R_2}), using the same colors as in Fig.16a for the corresponding time intervals, is shown in Fig.16b. In the first time interval (blue), the orbit wanders in the thin chaotic layer of the resonance $\omega_1 + 3\omega_2 - \omega_3 = 0$. In the second time interval (red) it jumps first to the resonance $3\omega_1 - \omega_2 - \omega_3 = 0$, and then to the resonance $\omega_1 - 2\omega_2 = 0$. In the third time interval (green) the orbit recedes from the doubly-resonant domain along the resonance $\omega_1 - 2\omega_2 = 0$.

The main result, now, is shown in Fig.16c representing a 3D plot in the variables $(\phi_{R_1}, J_{R_1}, E_Z)$, which visualizes Arnold diffusion for the same orbit. Taking 20 equidistant values of $E_{Z,i}$, $i = 1, 2, \ldots 20$ in the interval $0.029 \leq E_Z \leq 0.0306$, we first find the times t_i in the interval $0 \leq t \leq 9 \times 10^8$ when the normal form energy value $E_Z(t)$ of the numerical orbit approaches closest to the values $E_{Z,i}$. For each i, starting with the momentary values of all resonant variables at t_i, we then compute 1000 Poincaré consequents of the normal form flow on a surface of section defined by $(\phi_2 mod 2\pi) = 0$. The same procedure is repeated in a second interval $9 \times 10^8 \leq t \leq 1.5 \times 10^9$. As a net result, the orbit at the beginning and end of the calculation is found on the same section (corresponding to $E_Z = 0.0306$), but in a different resonant layer, having bypassed the barriers (invariant tori of the normal form dynamics) via a third dimension (here parameterized by the time-varying value of E_Z).

As an overall conclusion, we see that high order normal form calculations lead to practically useful results, i.e. the possibility to obtain good canonical variables in which the phenomenon of Arnold diffusion can be conveniently studied, but also visualized. In fact, the use of good canonical variables obtained via a normal form is crucial for a proper study of all main structures of the phase space that play a role in diffusion phenomena. Such structures are, for example, the unstable periodic orbits of the normal form flow, whose asymptotic manifolds (and their heteroclinic intersections) have been conjectured to provide the mechanism by which the diffusion progresses over the entire Arnold web.

4.5. Arnold's example and its relation to normal forms

We close this section with some reference to the fundamental work of Arnold (1964), in which the topological features of diffusion along resonances in systems of three degrees of freedom were first discussed. The Hamiltonian model

considered by Arnold has the form:

$$H = \frac{I_1^2}{2} + \epsilon(\cos\phi_1 - 1) + \frac{I_2^2}{2} + \mu(\cos\phi_1 - 1)(\sin\phi_2 + \cos t) \ . \qquad (194)$$

A review of the properties of this model is given in Lega et al. (2008). Notice in particular Fig.2.1 of that review, showing a chain of the invariant manifolds corresponding to a set of *foliated pendula* along the axis I_2, for $\mu = 0$. Using Melnikov's theory, when $\mu \neq 0$, Arnold demonstrates that there are heteroclinic connections between the stable and unstable manifolds (called 'whiskers') of the chain of unstable 2D tori existing for different values of I_2. These connections can then cause long excursions of the chaotic orbits along the I_2 axis, even if ϵ is infinitesimally small.

We now show that the motivation for choosing a Hamiltonian model of the form (194) for the study of Arnold diffusion stems from a similarity with the optimal Hamiltonian arising in resonant normal form theory developed in the previous subsections. In fact, one can see that Arnold's model is a simplified version of the *simply-resonant* normal form (174). A comparison of the Hamiltonians (174) and (194) shows that the action angle pair (I_1, ϕ_1) in Arnold's model corresponds to the resonant pair $(\Delta J_{R_1}, \phi_{R_1})$ in the Hamiltonian (174), while I_2 corresponds to action ΔJ_{R_2}. Finally, the time t in Arnold's model corresponds to the angle ϕ_F in the Hamiltonian (174), since, neglecting the appearance of J_F in the remainder of Eq.(174), the angle ϕ_F varies essentially linearly in time and independently of the initial datum, i.e. we have $\phi_F = (m \cdot \omega)t$.

Finally, we notice that the term $\mu(\cos\phi_1 - 1)(\sin\phi_2 + \cos t)$ in Arnold's model provides a simplified model for the *remainder term* $R(\Delta J_{R_1}, \Delta J_{R_2}, J_F, \phi_{R_1}, \phi_{R_2}, \phi_F)$ in the Hamiltonian (174). It follows that *the parameter μ in Arnold's model (194) should have a value reflecting the size of the optimal remainder in a generic simply resonant normal form construction* of the type discussed in subsections 3.3 to 3.5.

However, at this point, precisely, lies the difficulty in generalizing Arnold's results in a generic Hamiltonian of the type considered above. The difficulty is the following: in Arnold's example, the possibility to implement Melnikov's theory in order to establish the existence of the whiskers' heteroclinic intersections referred to above is based on the fact that *the parameter μ is allowed to vary independently of the parameter ϵ*. However, in a generic model, *the optimal remainder size depends on ϵ*, i.e. one should consider a model in which $\mu \sim ||R_{opt}(\epsilon)||$. This is a theoretical difficulty, due to which a demonstration of Arnold's result in more general Hamiltonian systems of three degrees of freedom still remains an important open problem of Hamiltonian dynamical systems' theory.

5. HAMILTONIAN FORMALISM FOR BASIC SYSTEMS IN DYNAMICAL ASTRONOMY

In the present section, we present some basic Hamiltonian models in action - angle variables which find applications in the two main fields of dynamical astronomy, namely celestial mechanics and galactic (or stellar) dynamics. In every case considered, we give the basic starting formulae as well as some hints on how

to implement book-keeping in these models. Finally, in subsection 5.4 we present a 'case study', i.e. the resonant theory at the Inner Lindblad Resonance in spiral galaxies. In fact, using a technique analogous to the one used in subsection 3.3, in order to locate the position of a periodic orbit in this resonance, we are lead to a main conclusion regarding the orientation of these periodic orbits, which forms the basis for the orbital version of the so-called density wave theory of spiral structure in galaxies.

5.1. Action - angle variables in the restricted three body problem. The Hamiltonian at mean motion resonances

A basic formalism of Celestial Mechanics regards the derivation of a Hamiltonian function in action-angle variables describing the so-called *restricted three body problem* (RTBP). This refers to the dynamics of a small body (e.g. an asteroid or a small planet) under the combined gravitational effects of a star (Sun) and a giant planet (Jupiter) in orbit around the star. In fact, various extensions of this formalism (also to more than three bodies) can be used to study problems of greater complexity, such as the stability of planetary and/or satellite systems.

The definition of action angle variables in Hamiltonian models of Celestial Mechanics is based on the use of the so-called *elements* (of a Keplerian elliptic orbit). The elements are quantities allowing to fit the instantaneous motion of a test particle to a temporary ellipse, which is the ellipse that the particle would constantly move on if, at a certain time t, all interactions except for the one with the central star were instantaneously 'turned off'. Since in the real problem, however, these interactions always exist, it is the values of the elements that change in time, i.e. the elements are *osculating*. However, in many cases the use of canonical perturbation theory allows one to define approximate action integrals for the Hamiltonian under study. These are quite useful, since they serve as labels for the motions of small bodies in the solar system. Such integrals are called *proper elements* (see e.g. Milani and Knežević (1990), Knežević et al. (2002)). Their computation for an extended catalogue of minor bodies has been a central subject in studies of asteroidal dynamics, since they allow to identify the so-called *families* of asteroids, i.e. present-day groups of asteroids which are assumed to have originated from one bigger body through, e.g., some past collision event. Finally, the diffusion of such bodies in the action space, i.e. the space of proper elements, allows one to deduce information on the history, or even the *chronology* of a family (Milani and Farinella (1994), Nesvorny et al. (2002, 2003), Tsiganis et al. (2007)). In fact, the study of the long-term evolution of the various asteroidal zones (like the main or Kuiper's belt in our solar system) provides clues to restructuring *the history of the solar system itself*, and it is a quite modern subject of study (see, for example, Knežević and Milani 2005, Lazzaro et al. 2006).

The derivation of the Hamiltonian formulation of the RTBP is a classical topic discussed in all books of Celestial Mechanics (see, for example, Murray and Dermott (1999), Morbidelli (2002), or Ferraz-Mello (2007) for some modern expositions). Here, we only give some basic formulae, in order to demonstrate that this formalism is amenable to the usual book-keeping approach discussed in sections 3.1 and 3.2 for analytic Hamiltonian systems containing a perturbation with infinitely many Fourier harmonics.

The starting formula for the description of a test particle's motion in the combined gravitational potential of a star of mass M and a planet of mass m moving around the star is

$$H = \frac{\mathbf{p}^2}{2} - \frac{GM}{r} - Gm \left(\frac{1}{|\mathbf{r} - \mathbf{r'}|} - \frac{\mathbf{r} \cdot \mathbf{r'}}{\mathbf{r'^3}} \right) \qquad (195)$$

where $\mathbf{p}$ is the momentum per unit mass, $\mathbf{r}$, $\mathbf{r'}$ are the positions of the test body and of the planet respectively in a fixed heliocentric (=centered around the star) system of reference, r is the modulus of $\mathbf{r}$, and G is Newton's gravitational constant. In this formula, the terms $-GM/r$ and $-Gm/|\mathbf{r} - \mathbf{r'}|$ give the contribution to the gravitational potential due to the star and to the planet respectively. On the other hand, the term $-Gm\mathbf{r} \cdot \mathbf{r'}/\mathbf{r'^3}$ describes the effects caused by considering a frame of reference is fixed on the star, which, however, itself moves around the common barycenter of the star-planet system.

The elements are now defined as follows: we consider a fixed Cartesian frame (x, y, z) with origin on the star, as well as the plane of the temporary ellipse along which on which the test body would constantly move in the Keplerian approximation (i.e. if $m = 0$ in Eq.(195)). The plane of the ellipse intersects the plane (x, y) along a line. This is called *line of nodes*. The nodes (ascending and descending) are the points where the ellipse intersects the line of nodes. The oriented angle Ω between the positive direction of the x-axis and the direction of the line of nodes pointing to the ascending node, is called the *longitude of the nodes*. The angle ω formed between the direction of the ascending node and a line starting from the origin and directed to the perihelion of the ellipse is called *argument of the perihelion*, while the angle $\varpi = \omega + \Omega$ is called *longitude of the perihelion*. Finally, the angle u formed between the direction from the origin to the perihelion and the direction pointing to the actual position of the moving testing particle on the ellipse is called *true anomaly*, while the angle $M = n(t - t_0)$, where $n = 2\pi/T$ is the *mean motion* frequency corresponding to the period T of motion on the ellipse, and t_0 is a time when the test body passes from the perihelion, is called the *mean anomaly*. We define also the *mean longitude* $\lambda = \varpi + M$. The relation allowing to pass from the true to the mean anomaly is:

$$\cos(u) \quad = \quad \frac{2(1 - e^2)}{e} \left(\sum_{\nu=1}^{\infty} J_\nu(\nu e) \cos(\nu M) \right) - e \qquad (196)$$

$$\sin(u) \quad = 2 \quad \sqrt{1 - e^2} \sum_{\nu=1}^{\infty} \frac{1}{2}[J_{\nu-1}(\nu e) - J_{\nu+1}(\nu e)] \sin(\nu M)$$

where J_ν are Bessel functions, while e is the *eccentricity* of the ellipse defined by $(1-e)/(1+e) = r_p/r_a$, where r_p/r_a is the ratio of the distances to the perihelion and to the aphelion. On the other hand, the distance from the origin to the test body is given by:

$$r = a \left(1 + \frac{e^2}{2} - 2e \sum_{\nu=1}^{\infty} \frac{[J_{\nu-1}(\nu e) - J_{\nu+1}(\nu e)] \cos(\nu M)}{2\nu} \right) \qquad (197)$$

where a is the *major semi-axis* of the ellipse, [43] i.e. the distance from the center of the ellipse to the aphelion. The set of elements is completed by the *inclination* i, i.e the angle formed between the plane (x, y) and the plane of the ellipse.

The set of *modified Delaunay action-angle variables* (Λ, Γ, Z) (actions) and (λ, γ, ζ) (angles) is now defined by:

$$
\begin{aligned}
\Lambda &= \sqrt{(1-\mu)a}, & \lambda & \\
\Gamma &= \sqrt{(1-\mu)a}(1 - \sqrt{1-e^2}), & \gamma &= -\varpi \\
Z &= \sqrt{(1-\mu)a}(1 - \sqrt{1-e^2})(1 - \cos i), & \zeta &= -\Omega
\end{aligned}
\tag{198}
$$

where $\mu = m/M$ is the *mass parameter*. The action variables (Λ, Γ, Z) are a measure of the major semi-axis, the eccentricity and the inclination respectively.

In order to find the form of the Hamiltonian (195) in the modified Delaunay action-angle variables, we use Eqs.(196), (197), (198) for the test body and for the perturbing planet, as well as an equation relating the angle ϕ between the position vectors $\mathbf{r}$ and $\mathbf{r}'$, namely

$$
\cos(\phi) = (u - u' + \varpi - \varpi') \ .
$$

After all substitutions in Eq.(195), and a number of algebraic manipulations (described in every detail, e.g. in Murray and Dermott (1999)), we arrive at:

$$
H = -\frac{(1-\mu)^2}{2\Lambda^2} + n'\Lambda' - \mu R(\Lambda, \Gamma, Z, \lambda, \gamma, \zeta; a', e', i', \lambda', \gamma', \zeta')
\tag{199}
$$

where the function R is called the *disturbing function* of the RTBP and n' is the mean motion of the disturbing planet.

The form of the disturbing function, as well as the question of the existence of efficient methods to compute its coefficients, is a classical subject in Celestial Mechanics. Briefly, the disturbing function has the form

$$
\begin{aligned}
R =& \sum_{k_1, k_1', k_2, k_2', k_3, k_3'} \Big(R_{k_1, k_1', k_2, k_2', k_3, k_3'}(\Lambda, \Gamma, Z; a', e', i') \\
\times & \cos(k_1\lambda + k_1'\lambda' + k_2\gamma + k_2'\gamma' + k_3\zeta + k_3'\zeta') \Big)
\end{aligned}
\tag{200}
$$

where the integers $(k_1, k_1', k_2, k_2', k_3, k_3')$ satisfy the following so-called *D'Alembert rules*: i) k_3, k_3' are even, and ii) $k_1 + k_1' + k_2 + k_2' + k_3 + k_3' = 0$. Furthermore, we have that the coefficient $R_{k_1, k_1', k_2, k_2', k_3, k_3'}$ is of the form:

$$
\begin{aligned}
R_{k_1, k_1', k_2, k_2', k_3, k_3'}(\Lambda, \Gamma, Z; a', e', i') = & \\
\sum_{s_2 \geq |k_2|, s_2' \geq |k_2'|, s_3 \geq |k_3|, s_3' \geq |k_3'|} & Y(\Lambda; a')\Gamma^{s_2/2}(e')^{s_2'} Z^{s_3/2}(i')^{s_3'}
\end{aligned}
\tag{201}
$$

---

[43]Called, sometimes, also 'semi-major axis'. But an axis of an ellipse cannot be 'semi-major', either it is the major or the minor one.

where the coefficients $Y(\Lambda; a')$ are determined in terms of the so-called *Laplace coefficients* (see Murray and Dermott 1999).

A case of particular interest that can be studied with the help of the Hamiltonian model (199) is the case of *mean motion resonance* (MMR). This refers to a resonance between the periods T, T', or the mean motions $n = 2\pi/T$, $n' = 2\pi/T'$ of an asteroid and of Jupiter respectively. The main elements of dynamics in mean motion resonances are reviewed e.g. in Tsiganis (2008). The basic formulae are as follows: If we consider an asteroid moving on a Keplerian ellipse, according to third Kepler's law we have a mean motion resonance

$$kn - (k+q)n' = 0 \tag{202}$$

with k, q positive integers when the asteroid's ellipse has major semi-axis:

$$a_{k,q} = a' \left(\frac{k}{k+q} \right)^{2/3} (1 - \mu)^{1/3} \ .$$

The integer k is called the order of the resonance. The value of $a_{k,q}$ corresponds to a distance (from the Sun) which defines the region of a particular MMR. If we define the *resonant* action-angle variables by:

$$\psi = k\lambda - (k+q)\lambda', \quad \Lambda = k\Psi \tag{203}$$

the Hamiltonian can be shown to take the form (assuming the inclination of Jupiter equal to zero, and averaging over short period terms):

$$
\begin{aligned}
H_{MMR} \ = \ & -\frac{(1-\mu)^2}{2k^2\Psi^2} - n'(k+q)\Psi \\
& - \mu \left[c_1(\Psi)\Gamma + c_2(\Psi)e'\Gamma^{1/2}\cos(\gamma) + c_3(\Psi)Z \right] \\
& - \mu \sum_{m_1,m_2,m_3} C_{m_1,m_2,m_3}(\Psi, \Gamma, Z; e')\cos(m_1\psi + m_2\gamma + m_3\zeta) \ .
\end{aligned}
\tag{204}
$$

Under this basic form, the Hamiltonian (204) has been used in many classical studies of e.g. the origin of the Kirkwood gaps or diffusion in the main asteroidal belt (e.g. Wisdom (1980), (1983), Murray and Holman (1997), Neishtadt (1987), Nesvorny et al. (2002)). From the Hamiltonian (204) we see immediately that the major semi-axis of the asteroid does not vary significantly, since $\dot{\Lambda} = k\dot{\Psi} = O(\mu)$. Thus, in simplified models we quite often substitute Ψ in the coefficients c_1, c_2, c_3 as well as in the disturbing function by the constant value $\Psi_{k,q} = (1/k)\sqrt{(1-\mu)a_{k,q}}$.

Under this latter form, the Hamiltonian (204) represents a case of the second fundamental resonance model (see footnote 7). In particular, denoting $c_1^* = c_1(\Psi_{k,q}), c_2^* = c_2(\Psi_{k,q})$, the term $c_1^*\Gamma$ can be thought of as representing a harmonic oscillator Hamiltonian. However, due to the presence of also a term $\sim \Gamma^{1/2}$, the 'equilibrium' position of this oscillator is shifted with respect to the value $\Gamma = 0$. This is seen most easily if we pass to the so-called 'Poincaré' (or cartesian) variables $X = \sqrt{2\Gamma}\cos\gamma$, $Y = \sqrt{2\Gamma}\sin\gamma$, in view of which the oscillator part of the Hamiltonian takes the form

$$H_\Gamma = -\mu \left[\frac{c_1^*}{2}(X^2 + Y^2) + \frac{c_2^*e'}{\sqrt{2}}X \right]$$

implying that the equilibrium point is at $X_0 = -c_2^* e'/(c_1^* \sqrt{2})$. This non-zero equilibrium position represents now a 'forced eccentricity', i.e. the orbit of the asteroid at the equilibrium position has a certain eccentricity induced essentially by the effect of Jupiter's eccentricity. The value of the forced eccentricity is $e_f \simeq |X_0|/((1-\mu)a_{k,q})^{1/4}$.

It is now possible to define book-keeping rules for a treatment of the Hamiltonian (204). The key point is that the analyticity properties of the Hamiltonian result in that the coefficients C_{m_1,m_2} exhibit exponential decay with respect to the modulus $|m| = |m_1| + |m_2| + |m_3|$. In fact, one can see that trigonometric terms of the form $\cos(m_1\psi + m_2\gamma + m_3\zeta)$ can only contain powers $\Gamma^{s/2}Z^{p/2}$ with $s \geq |m_2|$, $p \geq |m_3|$. Thus, in the former case we have powers of a small quantity proportional to the asteroid's eccentricity, while in the latter case we have powers of a small quantity proportional to the asteroid's inclination. In summary, a simple book keeping rule for the Hamiltonian (204) is:

Book-keeping rule for the MMR Hamiltonian: i) Introduce a factor λ^0 in the first two lines of Eq.(204). ii) introduce a book-keeping factor λ^{1+s+p} in front of any trigonometric term of the third line of Eq.(204) with a monomial coefficient $\Gamma^{s/2}Z^{p/2}$.

5.2. Hamiltonian models of axisymmetric galaxies (or other axisymmetric gravitating bodies)

The gravitational field generated by a number of different astronomical or astrophysical objects can be well approximated by an *axisymmetric* gravitational potential, i.e., a potential of the form $V \equiv V(r,z)$ in cylindrical coordinates (r,ϕ,z) (z =axis of symmetry). Examples of such bodies are: i) axisymmetric galaxies, and ii) oblate stars or planets. The orbits around such bodies can, to some extent, be described by common forms of canonical perturbation theory.

We will refer, in particular, to the case of the orbits of stars in an axisymmetric galaxy. The orbits in the equatorial plane are described by a Hamiltonian of the form

$$H_0 = \frac{p_r^2}{2} + \frac{p_\theta^2}{2r^2} + V_0(r) \tag{205}$$

where $V_0(r) = V(r, z = 0)$, and $p_\theta = L_z = $ const is the component of the angular momentum along the axis of symmetry, which is a preserved quantity. Due to this, the Hamiltonian (205) can be considered as of one degree of freedom, with $p_\theta = L_z$ as a parameter. All the orbits are *rosettes* moving in an annulus between a pericentric and an apocentric radius (see Efthymiopoulos et al. (2008) for a review of the various types of orbits in axisymmetric galaxies). The quantity

$$V_{eff}(r; L_z^2) = L_z^2/2r^2 + V_0(r)$$

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 $E = V_{eff}(r, L_z^2)$. These are joined at the radius r_c of the circular orbit which corresponds to the minimum of the effective potential, i.e. the root, for r_c of the equation

$$-\frac{L_z^2}{r_c^3} + \frac{dV_0(r_c)}{dr} = 0 \quad . \tag{206}$$

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

$$\kappa = \sqrt{\frac{d^2V_{eff}(r_c)}{dr_c^2}} = \sqrt{\frac{d^2V_0(r_c)}{dr_c^2} + \frac{3}{r_c}\frac{dV_0(r_c)}{dr_c}} \quad . \tag{207}$$

On the other hand, the angular velocity of circular orbits is given by

$$\Omega(r_c) = \left(\frac{1}{r_c}\frac{dV_0(r_c)}{dr}\right)^{1/2} \quad . \tag{208}$$

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 a circular motion of the *guiding center*, with frequency $\Omega(r_c)$, and an oscillation in *both* the radial and angular directions, with frequency $\kappa(r_c)$.

If we now consider orbits off the equatorial plane, the Hamiltonian takes the form

$$H = \frac{p_z^2}{2} + \frac{p_r^2}{2} + \frac{L_z^2}{2r^2} + V(r, z) \quad . \tag{209}$$

A basic form of canonical perturbation theory arises from developing the Hamiltonian (209) around the coordinates $r = r_c$, $z = 0$ characterizing the circular orbit with given angular momentum L_z. We introduce the variables $q_1 = r - r_c$, $q_2 = z$, and notice that

$$-\frac{L_z^2}{r_c^3} + \frac{\partial V(r_c, 0)}{\partial r} = 0$$

(condition (206) of the circular orbit), and

$$\frac{\partial V(r_c, 0)}{\partial z} = 0$$

(condition of zero perpendicular force on stars whose orbits lie in the equatorial plane). The series expansion of (209) around $r = r_c$, $z = 0$ now yields (apart from a constant)

$$H = \frac{p_1^2}{2} + \frac{p_2^2}{2} + \frac{1}{2}\omega_1^2 q_1^2 + \frac{1}{2}\omega_2^2 q_2^2 + \sum_{s=3}^{\infty}\sum_{d=0}^{s} V_{s,d} q_1^d q_2^{s-d} \tag{210}$$

$$V_{s,d} = \frac{1}{(s-d)!d!}\frac{\partial^s}{\partial^{s-d}r\partial^d z}\left(\frac{L_z^2}{2r^2} + V(r, z)\right)\Bigg|_{r=r_c, z=0}$$

where $\omega_1 = \kappa$ as given by Eq.(207), and $\omega_2 = (\partial^2 V(r_c, 0)/\partial z^2)^{1/2}$. The Hamiltonian (210) is a *polynomial series* in the variables (q_1, q_2). In fact, the third

order truncation of the Hamiltonian (210) has a special place in the historical development of galactic dynamics, since it is the Hamiltonian function for which a 'third integral' (besides the energy and the angular momentum) was computed by Contopoulos (1960). Also, after a linear transformation this Hamiltonian reduces to the celebrated Hénon - Heiles model, on which the applicability of the 'third integral' was checked for the first time by numerical integrations (Hénon and Heiles, (1964)).

We can show now that the Hamiltonian (210), when transformed into action - angle variables, displays exponential decay of its associated Fourier coefficients. In order to introduce action - angle variables, we observe that the lowest order approximation of the Hamiltonian (210) corresponds to a 2D harmonic oscillator motions with frequencies ω_1 and ω_2. The harmonic oscillator action - angle variables are introduced via the canonical transformation

$$q_i = \sqrt{\frac{2I_i}{\omega_i}} \sin \phi_i, \quad p_i = \sqrt{2\omega_i I_i} \cos \phi_i, \quad ,i = 1,2 \tag{211}$$

in view of which the Hamiltonian (210) takes the form

$$
\begin{aligned}
H =& \ \omega_1 I_1 + \omega_2 I_2 \\
&+ \sum_{s=3}^{\infty} \sum_{d=0}^{s} V_{s,d} \left(\sqrt{\frac{2I_1}{\omega_1}} \frac{(e^{i\phi_1} - e^{-i\phi_1})}{2i} \right)^d \left(\sqrt{\frac{2I_2}{\omega_2}} \frac{(e^{i\phi_2} - e^{-i\phi_2})}{2i} \right)^{s-d} .
\end{aligned}
\tag{212}
$$

Consider a Fourier term $e^{i(k_1\phi_1 + k_2\phi_2)}$. It is straightforward to see that in the sum of the r.h.s. of Eq.(212) such a term appears first at the order $s = |k_1| + |k_2|$, with coefficient $V_{|k_1|+|k_2|,|k_1|}(1/2^{3/2})^{|k_1|+|k_2|}\omega_1^{-|k_1|/2}\omega_2^{-|k_2|/2}I_1^{|k_1|/2}I_2^{|k_2|/2}$. Since higher order contributions are smaller in size, we may use the above expression as an estimate of the total size of the coefficient of the Fourier term $e^{i(k_1\phi_1 + k_2\phi_2)}$. For further simplification, we may consider the two frequencies ω_1, ω_2 of similar size and substitute them by their mean ω. Finally, we note that the original Hamiltonian expansion (210) is, in general, valid within a domain of convergence given by the inequalities $|q_i| < \rho_0$, $|p_i| < \rho_0$, $i = 1,2$, for some constant $\rho_0 > 0$. Thus, there is a positive constant A such that

$$\sum_{s=3}^{\infty} \sum_{d=0}^{s} |V_{s,d}||q_1|^d|q_2|^{s-d} < A \sum_{s=3}^{\infty} \sum_{d=0}^{s} \frac{|q_1|^d|q_2|^{s-d}}{\rho_0^s} .$$

This implies that for the Fourier coefficient H_{k_1,k_2} of the term $e^{i(k_1\phi_1+k_2\phi_2)}$ in the Hamiltonian (212) we have an upper bound estimate

$$|H_{k_1,k_2}| < G_{k_1,k_2}$$

with

$$G_{k_1,k_2} \sim A \left(\frac{1}{\rho_0(8\omega)^{1/2}} \right)^{|k|} I_1^{|k_1|/2} I_2^{|k_2|/2}$$

where $|k| = |k_1| + |k_2|$. However, recalling that the action variables themselves represent oscillations with respect to an equilibrium point of the Hamiltonian

(212) (which corresponds to the equatorial circular orbit), the expansion in this Hamiltonian is valid for orbits whose action variables I_1 and I_2 remain always small enough so that the original canonical variables q_1,q_2,p_1,p_2 remain within the domain or convergence defined by ρ_0. This happens in a polydisc defined by two inequalities of the form $|I_i| < I_{0,i}$, $i = 1, 2$, where $I_{0,i}$ are positive constants. Thus, defining $I_0 = \min\{I_{0,1}, I_{0,2}\}$ we arrive at

$$G_{k_1,k_2} \sim A \left(\frac{I_0}{8\rho_0^2\omega} \right)^{|k|/2}$$

or

$$|H_{k_1,k_2}| < A e^{-\sigma|k|} \ \ \text{with} \ \ \sigma \sim \frac{1}{2} \ln \left(\frac{8\rho_0^2\omega}{I_0} \right) \ \ . \tag{213}$$

Eq.(213) clearly shows that the exponential decay of the Fourier coefficients of the Hamiltonian (212) is steeper in *smaller* polydiscs in the action space, i.e. for smaller values of I_0. As a result, the most natural choice of book-keeping in the Hamiltonian (212) is in ascending powers of the action variables, namely:

$$\begin{aligned} H \ = \ & \omega_1 I_1 + \omega_2 I_2 \\ + \ & \sum_{s=3}^{\infty} \lambda^{s-2} \sum_{d=0}^{s} V_{s,d} \left(\sqrt{\frac{2I_1}{\omega_1}} \frac{(e^{i\phi_1} - e^{-i\phi_1})}{2i} \right)^d \left(\sqrt{\frac{2I_2}{\omega_2}} \frac{(e^{i\phi_2} - e^{-i\phi_2})}{2i} \right)^{s-d} \ . \end{aligned} \tag{214}$$

We note that the terms of degree $s/2$ in the actions (or s in the original 'Poincaré' variables (q,p)) acquire a book-keeping factor λ^{s-2}. The reduction of the exponent by two is done for algorithmic convenience, i.e., in this way the harmonic oscillator terms are of order 0 in λ, the cubic terms of order 1, etc. In fact, the same rule can be applied in more general polynomial Hamiltonian models appearing in various contexts besides galactic dynamics.

5.3. Hamiltonian models in barred-spiral galaxies

Consider a disc galaxy with a rotating 'pattern' figure, i.e. a set of spiral arms, or a bar, rotating with uniform angular speed Ω_p. In cylindrical coordinates (r,θ,z), the gravitational potential on the disc plane takes the form

$$V(r,\theta, z = 0) = V_0(r) + V_1(r,\theta) \tag{215}$$

where V_0 accounts for the gravitational effects of the axisymmetric disc component, and $V_1(r,\theta)$ for the gravitational effects of all non axisymmetric features in the disc (i.e. the spiral arms or the bar) of the galaxy. The reader is deferred to (Efthymiopoulos (2010)) for a tutorial presentation of how, starting from observations of a disc galaxy, the decomposition of the potential in components of the form (215) can be realized in practice.

It is convenient to express the component $V_1(r,\theta)$ in terms of its Fourier decomposition

$$V_1(r,\theta) = \sum_{m=-\infty}^{\infty} V_m(r)e^{im\theta} \ \ . \tag{216}$$

One has $Im(V_m) = -Im(V_{-m})$ for V_1 to be a real-valued function of its (real-valued) arguments.

Let r_c be a fixed radius. We 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:

$$H(r, \theta, p_r, p_\theta) = \frac{p_r^2}{2} + \frac{p_\theta^2}{2r^2} - \Omega_p p_\theta + V_0(r) + \sum_{m=-\infty}^{\infty} V_m(r) e^{im\theta} \quad . \qquad (217)$$

In (217), 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 (217) describes motions in the rotating frame; p_θ defined as such preserves its canonical conjugate relation with θ).

We now define action-angle variables for the above Hamiltonian to describe local motions around r_c. The pair (θ, p_θ) in (217) is already an angle-action pair. As in subsection 3.4, a second pair (θ_r, J_r) can be defined for epicyclic oscillations via the canonical transformation $(r, p_r) \to (\theta_r, J_r)$ given by

$$r - r_c = \left(\frac{2J_r}{\kappa(r_c)}\right)^{1/2} \sin\theta_r, \qquad p_r = (2\kappa(r_c)J_r)^{1/2} \cos\theta_r \quad . \qquad (218)$$

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.(217), and Taylor-expanding with respect to r_c brings the Hamiltonian to the form (apart from a constant)

$$\begin{aligned} H(\theta_r, \theta, J_r, J_\theta) &= \kappa_c J_r + \frac{1}{2r_c^2} J_\theta^2 + (\Omega_c - \Omega_p) J_\theta & (219) \\ &+ \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} F_{m,n}(J_r, J_\theta) e^{i(n\theta_r + m\theta)} \end{aligned}$$

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

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

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

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

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

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

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

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

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

The key point to notice in Eq.(219) is that the basic frequencies κ_c and $\Omega_c - \Omega_p$ depend on the reference radius r_c in a continuous way. Thus, as r_c increases, infinitely many different resonant combinations $k_1\kappa_c + k_2(\Omega_c - \Omega_p) = 0$ are encountered.

In subsection 5.4 we will examine an example of implementation of resonant normal form theory in the Hamiltonian (219), in the case of the so-called *inner Lindblad resonance*, where $\Omega(r_c) - \Omega_p = \kappa(r_c)/2$. We will see that the theory is able to predict the form and orientation of a family of elliptic-like periodic orbits, whose superposition provides the basis of the so-called orbital *density wave theory* (Kalnajs (1971), Contopoulos (1975)).

Regarding the book-keeping rules for the Hamiltonian (219), following an analogous procedure as in subsection 5.2 it is straightforward to show that the coefficients of the Fourier development in the second line of Eq.(219) exhibit exponential decay with respect to the Fourier order $|k| = |n| + |m|$. It follows that the most natural choice of book keeping in (219) goes with powers of the action variables, i.e. we have the following book-keeping rule:

Book-keeping rule for the Hamiltonian 219: introduce a factor λ^{1+s+p} in front of any Fourier term in the second line of Eq.(219) with coefficient depending on the monomial $J_\theta^s J_r^{p/2}$.

The term $J_\theta^2/2r_c^2$, on the other hand, can be book-kept either as $O(\lambda^0)$ or as $O(\lambda^1)$, according to whether or not, depending on the application, we want its inclusion in the Hori kernel of the corresponding homological equation.

In a number of applications, and in particular in the investigation of the so-called 'edge-on' profiles of disc galaxies, we are interested also in examining motions perpendicularly to the disc. The 3D motion of stars is accounted for by the Hamiltonian:

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

$$(220)$$

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

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 (220) around $r = r_c$, $z = 0$. Defining

$$\kappa_z(r_c) = \left(\frac{\partial^2 V_0(r_c, z = 0)}{\partial z^2} \right)^{1/2} \tag{221}$$

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

$$p_z = (2\kappa_z(r_c) J_z)^{1/2} \cos\theta_z, \quad z = \left(\frac{2J_z}{\kappa_z(r_c)} \right)^{1/2} \sin\theta_z, \tag{222}$$

the Hamiltonian (220) resumes the form

$$\begin{aligned}
H(\theta_r, \theta, \theta_z, J_r, J_\theta, J_z) &= \kappa_c J_r + (\Omega_c - \Omega_p) J_\theta + \kappa_{zc} J_z + \frac{1}{2r_c^2} J_\theta^2 \\
&\quad + \sum_{|m|+|n|+|q|=0}^{\infty} F_{m,n,q}(J_r, J_\theta, J_z) e^{i(n\theta_r + m\theta + q\theta_z)}
\end{aligned} \tag{223}$$

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.(219) hold, while the bookkeeping is implemented by the same rules as for the Hamiltonian (219).

5.4. An example: Resonant dynamics in the inner Lindblad resonance and the density wave theory of spiral arms

The stellar dynamics of disc galaxies is a classical topic of dynamical astronomy (see Contopoulos (2002) for a review). Here, we will discuss one particular aspect of this topic, namely the study of the motions of stars in the neighborhood of the so-called *inner Lindblad resonance*. This study is essential in understanding the form and structure of rotating spiral arms in normal disc galaxies, or of the inner parts of a rotating bar in barred galaxies.

In the previous subsection we presented the main features of the Hamiltonian formalism of the so-called epicyclic theory of motions in a disc galaxies. The epicyclic approximation is justified when the motions of stars in a disc galaxy are not very far from circular motions. If we assume that a galaxy contains a rotating figure, like a set of spiral arms, revolving with constant angular speed Ω_p, then we can define *disc resonances* by commensurability relations between the basic frequencies of the epicyclic theory. In a frame rotating with speed Ω_p, the disc resonances are defined by

$$\Omega - \Omega_p = \frac{n}{m}\kappa \quad , \tag{224}$$

where Ω and κ are the angular velocity of the circular orbit and the epicyclic frequency respectively, defined in subsection 3.6.

The following are the most important disc resonances: for $n = 1$, and $m = 2$ we have the *inner Lindblad resonance (ILR)*. For $n = -1$ and $m = 2$ we have the *outer Lindblad resonance* (OLR). Finally, for $n = 1$ and $m \to \infty$ we have *corotation*, otherwise defined by $\Omega = \Omega_p$. It is important to recall that: i) Ω and κ are functions of the radial distance r on the disc, thus, these resonances occur at particular distances from the center (plus or minus some 'width' Δr associated with resonance width, see subsection 3.3). ii) The location of all resonances depends on the pattern speed Ω_p. However, the pattern speed is one of the most difficult quantities to determine observationally. Thus, one may use information from resonant theory in order to proceed the other way around, i.e., *estimate the pattern speed* from other observable features of resonances. This is one of the central questions of research in galactic dynamics regarding disc galaxies.

In the basic formulation of resonant perturbation theory for disc galaxies (see e.g. Contopoulos (1975, 1978) and (2002, pp.436-460)) we start from a Hamiltonian of the form (219) and aim to transform it into a *resonant normal form* using a procedure of canonical transformations similar to those examined in subsection 2.9 and 3.3. The new Hamiltonian $Z(\theta'_r, \theta', I'_r, I'_\theta)$ 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), where (k_1, k_2) is a resonant wave-vector corresponding to the particular resonance under study. Under the form of the new Hamiltonian it is possible to explain the main features of the orbits in the associated resonant domain by analytical means.

In the case of the Inner Lindblad resonance, using resonant theory we can find the form and orientation of the basic *stable* periodic orbits existing in a large domain of the disc. In the case of spiral galaxies, we find that the main periodic orbits have a shape of *elongated ellipses* with a major axis changing orientation when considering orbits further and further away from the center. Then, the superposition of these orbits creates a *response density* which accounts for the observed spiral arms. This configuration, of 'precessing ellipses', conceived in the early works of Lindblad (1940, 1956, 1961), was given a concrete form by Kalnajs (1973), and serves as the basis of the orbital version of the so-called *density wave theory* (Lin and Shu (1964), see Bertin (2000) and Binney and Tremaine (2008)).

A numerical example of this mechanism is shown in Figure (17a). The details of this figure are explained in (Efthymiopoulos (2010)), and a brief summary is given below. Essentially, the figure shows the predictions of resonant normal form theory regarding the shape of the periodic orbits in a particular model (see Efthymiopoulos (2010)). We see that the orientation of the orbits changes in a way so as to support a density wave, i.e. a local enhancement of the density traveling within the disc (when viewed in a fixed frame). In fact, the stars go in and out of the wave as they move with much higher angular velocities than the pattern itself. However, as there are many stars distributed all along any of the periodic orbits of Fig.(17a), the pattern is maintained invariant in time, being at every snapshot composed by different stars.

The main question regards the so-called *self-consistency* of such a mechanism, namely whether the response density produced by the combination of
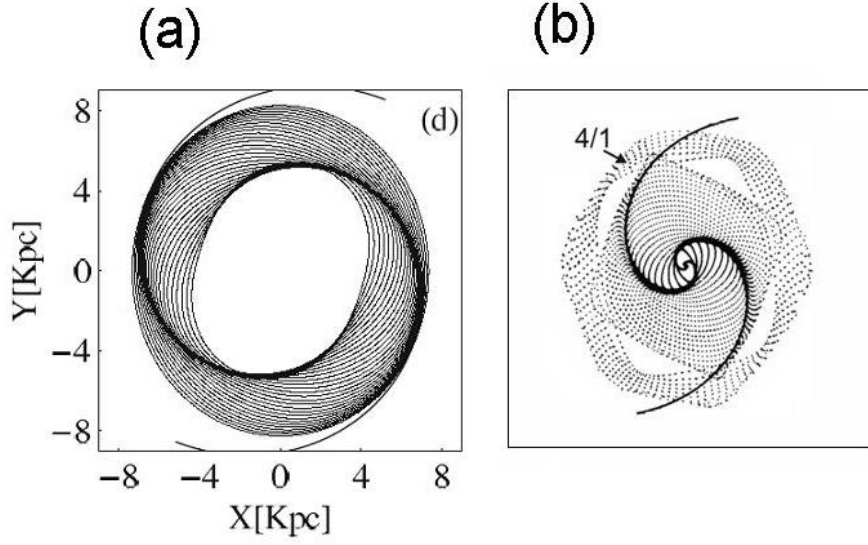
Figure 17. (a) Spiral density wave formed by periodic orbits beyond the Inner Lindblad Resonance in the model considered in Efthymiopoulos (2010). (b) Termination of spirals at the 4:1 resonance due to nonlinear effects which imply that the periodic orbits do not support the spiral arms beyond this resonance (after Contopoulos and Grosbol 1986).

'precessing ellipses' can match the imposed density upon which the calculation of periodic orbits was based. This question was examined in (Contopoulos (1985), Contopoulos and Grosbol (1986), and Patsis et al. (1991)). The main conclusion is that a self-consistent model of the above type can be maintained up to a distance a little before corotation, i.e. up to the 4/1 resonance, (see Fig.17b).

We now give the main features of the resonant normal form theory leading to a figure like Fig.17a. Our purpose will now be to compute a resonant normal form, and employ it in order to identify periodic orbits which are the *continuations of circular orbits* of the axisymmetric problem in the 2:1 resonant domain. We implement the following steps:

i) *Hamiltonian expansion.* We fix first a value of the radius r_c of a circular orbit of the axisymmetric problem near the ILR. The main features of the orbits are found by Taylor-expanding the Hamiltonian around r_c up to fourth degree in $r - r_c$, and by using the expressions (218). Then, starting from (219), the following Hamiltonian is arrived at:

$$
\begin{aligned}
H &= H_c + \omega_r J_r + \omega J_\theta + c_{20} J_r^2 + c_{11} J_r J_\theta \\
&+ c_{02} J_\theta^2 + c_{21} J_r^2 J_\theta + c_{22} J_r^2 J_\theta^2 + \ldots + H_{0F}(J_\theta; J_r, \theta_r) \\
&+ \left(d_{c1} J_r^{1/2} + d_{c3} J_r^{3/2} + \ldots \right) \cos(\theta_r - 2\theta) \\
&+ \left(d_{s1} J_r^{1/2} + d_{s3} J_r^{3/2} + \ldots \right) \sin(\theta_r - 2\theta) \\
&+ H_{1F}(J_r; \theta, \theta_r) \ .
\end{aligned}
\tag{225}
$$

The first two lines in (225) 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.(219)). 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', since under the dynamics induced by this Hamiltonian term all angles rotate with a fast frequency.

An important remark 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', i.e. $O(J_r^{1/2})$, terms can appear in the first two lines of (225).

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

$$d_{c1} = \frac{1}{(2\kappa_c)^{1/2}}[A_s'(r_c)\sin\phi_2(r_c) + A_s(r_c)\phi_2'(r_c)\cos\phi_2(r_c)] \qquad (226)$$

$$d_{s1} = \frac{1}{(2\kappa_c)^{1/2}}[A_s'(r_c)\cos\phi_2(r_c) - A_s(r_c)\phi_2'(r_c)\sin\phi_2(r_c)]$$

and we notice that the coefficients d_{c1} and d_{s1} are non-zero only if the non-axisymmetric perturbation is non-zero. We note here, in passing, that due to the form of Eq.(225), the dynamics in the inner Lindblad resonance represent a case of the 'second fundamental resonance model' (see footnote 7 and references therein).

Finally, 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 (225) 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. As explained in Efthymiopoulos (2010), this phenomenon of 'forced ellipticity' (a term introduced first by Kalnajs (1973)) is analogous to the phenomenon of 'forced eccentricity' in motions at mean motion resonances in the restricted elliptic three body problem.

In the case of the Hamiltonian (225), 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 will 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)

$$
\begin{aligned}
H_{res} &= \omega_r J_r + \omega J_\theta + c_{20} J_r^2 + c_{11} J_r J_\theta + c_{02} J_\theta^2 + c_{21} J_r^2 J_\theta + c_{22} J_r^2 J_\theta^2 + \ldots \\
&+ \left(d_{c1} J_r^{1/2} + d_{c3} J_r^{3/2} + \ldots \right) \cos(\theta_r - 2\theta) \\
&+ \left(d_{s1} J_r^{1/2} + d_{s3} J_r^{3/2} + \ldots \right) \sin(\theta_r - 2\theta) \quad .
\end{aligned}
\tag{227}
$$

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 will look for periodic orbits in the full Hamiltonian (227) 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 $(\theta_r, \theta, J_r, J) \to (\psi, \theta, J_r, J_F)$). The resonant Hamiltonian in new variables reads:

$$
\begin{aligned}
H_{res} &= (\omega_r - 2\omega) J_r + \omega J_F + (c_{20} - 2c_{11} + 4c_{02}) J_r^2 + (c_{11} - 4c_{02}) J_r J_F \\
&+ c_{02} J_F^2 - 2c_{21} J_r^3 + c_{21} J_r^2 J_F + 4c_{22} J_r^4 + c_{22} J_r^2 J_F^2 - 4c_{22} J_r^3 J_F \\
&+ \left(d_{c1} J_r^{1/2} + d_{c3} J_r^{3/2} \right) \cos \psi + \left(d_{s1} J_r^{1/2} + d_{s3} J_r^{3/2} \right) \sin \psi \quad .
\end{aligned}
\tag{228}
$$

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

$$
H_{res} = 0, \quad \dot\psi = \frac{\partial H_{res}}{\partial J_r} = 0, \quad \dot J_r = -\frac{\partial H_{res}}{\partial \psi} = 0 \quad .
\tag{229}
$$

The following simplifications allow to analytically approximate the roots J_r^*, J_F^*, ψ^* of the system (229):

1) to estimate J_{F*} we use the lowest order terms of the first of equations (229). That is

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

or

$$
J_F^* \simeq -\frac{1}{\omega} \left[(\omega_r - 2\omega) J_r^* + J_r^{*1/2} \left(d_{c1} \cos \psi^* + d_{s1} \sin \psi^* \right) \right] \quad .
\tag{230}
$$

2) We substitute the above expression into the second and third of equations (229) and solve simultaneously for J_r^*, ψ^*. It can be shown (Contopoulos (1975)) 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). We use a different nomenclature 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 (228), the roots for ψ^* are defined by:

$$\tan\psi^* = \frac{d_{s1}}{d_{c1}} = \frac{A_s'(r_c)\cos\phi_2(r_c) - A_s(r_c)\phi_2'(r_c)\sin\phi_2(r_c)}{A_s'(r_c)\sin\phi_2(r_c) + A_s(r_c)\phi_2'(r_c)\cos\phi_2(r_c)} \tag{231}$$

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

$$\left|\frac{A_s(r_c)\phi_2'(r_c)}{A_s'(r_c)}\right| = \left|\frac{1}{\tan i_0(1 - r_c\epsilon_s)}\right| >> 1$$

for i_0 equal to a few degrees, thus the second terms in both the numerator and denominator of the r.h.s. of (231) 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.(231) has two solutions, which differ by π. Substituting one of them into the third of Eqs.(229), 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 outermost 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 $J_r^{*1/2}$ 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 an in-and-out motion 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. However, it is possible to obtain the form of the periodic orbits by an alternative way, namely by computing a local *non-resonant* normal form, setting $J_{r*} = 0$ and $J_{\theta*}$ equal to the value of the angular momentum corresponding to the circular orbit in the axisymmetric case. Then, the form of the periodic orbits in the perturbed problem is found directly from the canonical transformation

connecting the action variables before and after the computation of the non-resonant normal form.

Figure 17a shows the combined effect of all previous phenomena, which will be called the 'precessing ellipses' mechanism of generation of spiral density waves. This figure shows a precise calculation of periodic orbits as described in the above steps for the model introduced in Efthymiopoulos (2010), and for the adopted parameter set in Fig. 4c of that reference. 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 $r_{ILR} < r_c \leq 1.6 r_{ILR}$. We see that the gradual variation of ψ^* essentially follows the phase of the imposed spirals. This fact causes a gradual re-orientation of the elliptical periodic orbits in a way so as to produce a 'response density' appearing enhanced exactly on the locus of the imposed spiral arms.

Regarding the termination of spiral arms supported by the above mechanism, the main effect is due to the presence of higher order terms in the expansions of the resonant Hamiltonian (227). Due to such terms, 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 (Contopoulos and Grosbol (1986, 1988), Patsis et al (1991). A numerical calculation shows that the resulting periodic orbits should have a rectangular shape (Fig.17b). This fact prevents the response density from supporting the self-consistency of the spiral arms beyond the 4:1 resonance, 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 Patsis et al. (1991)).

6. SUMMARY

The present article deals with some topics of canonical perturbation theory, as well as with applications of the latter in problems of stability and (weakly) chaotic diffusion in systems appearing in the framework of dynamical astronomy. The focus is on giving concrete examples of implementation of various normal form techniques, stressing the practical and/or computational aspects, and providing (in most examples) sufficient detail in order to facilitate self-study. We summarize below, in form of a 'practical guideline list', the main methods, techniques, and applications of normal forms discussed in the present text.

1) In subsection 2.3 we introduced the main algebraic technique used throughout this tutorial for performing near-identity canonical transformations needed in normal form computation. This is the technique of *Lie series*, whose merits are explained in the same subsection.

In subsection 2.4 we discussed an algorithmic technique, called *book-keeping* which greatly facilitates the practical computation of normal forms and the development of computer-algebraic programs for this purpose. It was explained how the book-keeping process, which reflects an evaluation of the importance

of various terms in the original Hamiltonian to the dynamics, influences also purely algorithmic aspects of a normal form construction. It was discussed how to develop book-keeping schemes maximizing algorithmic convenience.

2) In subsections 2.4 to 2.7 we presented the main elements of a *Birkhoff normal form*, using , as an example, the theoretical computation of rotational tori in a perturbed pendulum model. We described (subsection 2.6) the practical aspects of this computation, which allows to find i) approximate integrals of motion, and ii) the form of the invariant tori (or invariant curves) in *open domains* of the action space. We stressed however (subsection 2.7) the asymptotic character of the Birkhoff series, implying that there is an optimal normalization order beyond which the method yields worse instead of better results. We then introduced the notion of the *remainder* of a normal form Hamiltonian, which gives a measure of the deviations of the true dynamics from the theoretical dynamics computed via a normal form.

3) In subsection 2.8 we presented the *Kolmogorov normal form*, making again reference to a concrete example. We demonstrated that, contrary to the Birkhoff normal form, the Kolmogorov normal form is convergent for sufficiently small values of the perturbation parameter. This convergence forms, in fact, the basis for proving Kolmogorov theorem for the existence of invariant tori in nearly-integrable Hamiltonian systems. We explained how this convergence arises by examining the effects of the accumulation of small divisors in this case. Finally, we emphasized that in practical computations the Kolmogorov normal form can be used in order to find a convergent series representation of the motion on *just one* invariant torus (for some fixed choice of frequencies), instead of an open domain, as in the case of the Birkhoff normal form. This limitation notwithstanding, the Kolmogorov normal form allows to recover the form of quasi-periodic solutions on a torus with an arbitrary accuracy, while Birkhoff's method cannot have an accuracy better than the size of the remainder at the optimal normalization order, which is finite.

4) In subsection 2.9 we examined the construction of a *resonant normal form*. This was used in order to obtain the local phase portraits in domains containing *islands of stability*. Also, the resonant normal form was used to estimate how the size of an island scales with ϵ.

5) In subsection 2.10 we introduced the *hyperbolic normal form* due to Moser (1958), using again a concrete example. It was shown that by this normal form it is possible to compute the position and a periodic series representation of an *unstable* periodic orbit, and also to compute (up to some extent) the stable and unstable invariant manifolds of the same orbit. We then discussed an extension of the method, allowing to compute the position of homoclinic points where the stable and unstable manifolds intersect each other. We finally explained why the method continues to be successful in the regime of strong chaos.

6) Section 3 deals with systems whose Hamiltonian contains infinitely many Fourier harmonics already at first order in the small parameter ϵ. In this case, it was explained how to perform book-keeping, by splitting the Fourier series of the

original Hamiltonian in groups of Fourier terms of different order of smallness. It was shown that such a splitting is naturally suggested by the consequences of the Fourier theorem on analytic functions, which implies that the Fourier coefficients decay exponentially with increasing Fourier order. Finally, we gave two concrete examples of how to implement book-keeping in this case, one for a resonant normal form, and one for the Kolmogorov normal form. As an application, in the former case we computed a critical value of ϵ at which a *resonance overlap* criterion is fulfilled for the system under study.

7) Section 4 presents the relation between resonant normal form theory in systems of three degrees of freedom and the phenomenon called Arnold diffusion. This includes a discussion of the role of convexity in the relevant phenomena, as well as some reference to the relation between the original model introduced by Arnold (1964) and the Hamiltonians found by resonant normal form theory. However, we also give a concrete example of visualization of Arnold diffusion using appropriate variables constructed via a resonant normal form.

8) Finally, in section 5 we provide some basic formulations of Hamiltonian functions in action-angle variables for three problems of central interest in dynamical astronomy. These are i) the restricted three body problem (with a particular application in the case of mean motion resonances in the solar system), ii) the Hamiltonian in axisymmetric non-rotating galaxies, and iii) the Hamiltonian in rotating barred-spiral galaxies. In all cases we give explicit formulae for the corresponding Hamiltonian functions, supplemented by a discussion of how to perform book-keeping in normal form computations in each case. The section closes with one example from resonant normal form theory in the inner Lindblad resonance in disc galaxies, which leads to the derivation of a theoretical model for the orbits supporting a 'density wave' form of spiral arms.

References

Aquilano, R.O., Muzzio, J.C., Navone, H.D., & Zorzi, A.F. 2007, Cel. Mech. Dyn. Astron. 99, 307

Arnold, V.I. 1963, Russ. Math. Surveys, 18, 9

Arnold, V.I. 1964, Sov. Math. Dokl., 6, 581

Arnold, V.I., & Avez, A. 1968, Ergodic Problems in Classical Mechanics, Benjamin, New York.

Arnold, V.I. 1978, Mathematical Methods of Classical Mechanics, Springer-Verlag, Berlin.

Arnold, V.I., Kozlov, V.V., & Neishtadt, A.I. 1993, Mathematical Aspects of Classical and Celestial Mechanics (2rd Edition), Springer, Berlin - Heidelberg.

Benest, D., Froeschlé, C., & Lega, E. 2008, Topics in Gravitational Dynamics, Lect. Notes Phys., 729, Springer.

Benest, D., Froeschlé, C., & Lega, E. 2010, Diffusion and Dissipation in Quasi-Integrable Systems. Theory and Application to Gravitational Systems, Eur. Phys. J. Spec. Topics, 186.

Benettin, G., Galgani, L., & Giorgilli, A. 1985, Cel. Mech., 37, 1.

Benettin, G., & Gallavotti, G. 1986, J. Stat. Phys., 44, 293.

Benettin, G. 1999, in A. Giorgilli (Ed.) Hamiltonian Dynamics. Theory and Applications, *Lect. Notes Math.* 1861, 1.

Bertin, G. 2000, Dynamics of Galaxies, Cambridge Univ. Press, Cambridge.

Binney, J., & Tremaine, S. 2008, Galactic Dynamics (2nd edition), Princeton University Press, New Jersey.

Boccaletti, D. & Pucacco, G. 1996, Theory of Orbits, Springer, Berlin.

Cachucho, F., Cincotta, P.M., & Ferraz-Mello, S. 2010, Cel. Mech. Dyn. Astron., 108, 35.

Celletti, A., Giorgilli, A., & Locatelli, U. 2000, Nonlinearity, 13, 397.

Cherry, T.M. 1924, Proc. Camb. Phil. Soc. 22, 510.

Chierchia, L. 2008, Reg. Cha. Dyn., 13, 130.

Chirikov, B.V. 1979, Phys. Rep., 52, 263.

Cincotta, P. 2002, New Astronomy Reviews, 46, 13.

Cincotta, P., Giordano, C., & Pérez, M.J. 2006, Astron. Astrophys. 455, 499.

Cincotta, P., & Giordano, C. 2012, in Cincotta, P., Giordano, C., & Efthymiopoulos, C. (eds), Proceedings of the Third La Plata School on Astronomy and Geophysics, AAA.

Contopoulos, G. 1960, Z. Astrophys., 49, 273.

Contopoulos, G. 1963, Astron. J., 68, 763.

Contopoulos, G., & Moutsoulas, M. 1965, Astron. J., 70, 817.

Contopoulos, G. 1966, Bull. Astron. CNRS, 2, 223.

Contopoulos, G, 1975, Astrophys. J, 201, 566.

Contopoulos, G, 1978, Astron. Astrophys., 64, 323.

Contopoulos, G., & Grosbol, P. 1986, Astron. Astrophys., 155, 11.

Contopoulos, G. & Grosbol, P. 1988, Astron. Astrophys., 197, 83.

Contopoulos, G. & Polymilis, H. 1993, Phys. Rev. E, 47, 1546.

Contopoulos, G. 2002, Order and Chaos in Dynamical Astronomy, Springer, Berlin.

Cordeiro, R.R., & Mendes de Souza, L.A. 2005, Astron. Astrophys. 439, 375.

Cordeiro, R.R. 2006, Astron. J., 132, 2114.

Da Silva Ritter, G.I., Ozorio de Almeida, A.M., & Douady, R. 1987, Physica D, 29, 181.

Delshams, A., & Gutiérrez, P. 1996, J. Diff. Eq., 128, 415.

Deprit, A., 1969, Cel. Mech., 1, 12.

Dumas, H.S., & Laskar, J. 1993, Phys. Rev. Lett., 70, 2975.

Efthymiopoulos, C., Contopoulos, G., & Voglis, N. 1998, in Benest, D., and Froeschlé, C. (Eds), Discrete Dynamical Systems, Gordon and Breach Science Publishers, pp.91-106.

Efthymiopoulos, C., Giorgilli, A., & Contopoulos, G. 2004, J. Phys. A Math. Gen., 37, 10831.

Efthymiopoulos, C. 2008, Cel. Mech. Dyn. Astron., 102, 49.

Efthymiopoulos, C., Voglis, N., & Kalapotharakos, C. 2008, in D. Benest, C. Froeschlé and E. Lega (eds), Topics in Gravitational Dynamics, Lect. Notes Phys., 729, Springer, Berlin, pp.297-389.

Efthymiopoulos, C. 2010, Eur. Phys. J. Spec. Topics, 186, 91.

Efthymiopoulos, C., & Harsoula, M. 2012, "The speed of Arnold diffusion", submitted.

Ferraz-Mello, S. 2007, Canonical Perturbation Theories. Degenerate Systems and Resonance, Springer (New York).

Froeschlé, C. 1970a, Astron. Astrophys. 4, 115.

Froeschlé, C. 1970b, Astron. Astrophys. 5, 177.

Froeschlé, C. 1972, Astron. Astrophys. 16, 172.

Froeschlé, C., & Scheidecker, J.P. 1973, Astron. Astrophys. 22, 431.

Froeschlé, C., Lega, E., & Gonczi, R. 1997, Cel. Mech. Dyn. Astron., 67, 41.

Froeschlé, C., Guzzo, M., & Lega, E. 2000, Science, 289, (5487), 2108.

Froeschlé, C., Guzzo, M., & Lega, E. 2005, Cel. Mech. Dyn. Astron., 92, 243.

Giordano, C.M., & Cincotta, P.M. 2004, Astron. Astrophys., 423, 745.

Giorgilli A., & Locatelli, U. 1997, ZAMP, 48, 220.

Giorgilli, A., & Locatelli, U. 1999, in C. Simó (ed.), Hamiltonian Systems with Three or More Degrees of Freedom, NATO ASI Series C, 533, Kluwer, Dordrecht, pp. 72 - 89.

Giorgilli, A 2001, Disc. Cont. Dyn. Sys., 7, 855.

Giorgilli, A 2002, Notes on exponential stability of Hamiltonian systems, in Dynamical Systems. Part I: Hamiltonian Systems and Celestial Mechanics, Pubblicazioni della Classe di Scienze, Scuola Normale Superiore, Pisa.

Giorgilli, A., Locatelli, U., & Sansoterra, M.: 2009, Cel. Mech. Dyn. Astron., 104, 159.

Gómez, G., & Barrabés, E. 2011, Scholarpedia 6(2), 10597.

Grobman, D.M. 1959, Dokl. Akad. Nauk SSSR, 128, 880.

Gustavson, F., 1966, Astron. Astrophys., 71, 670.

Guzzo, M., Lega, E., & Froeschlé, C. 2005, Discr. Con. Dyn. Sys. B, 5, 687.

Guzzo, M., Lega, E., & Froeschlé, C. 2006, Nonlinearity, 19, 1049.

Hartman, P. 1960, Proc. Amer. Math. Soc., 11, 610.

Hénon, M., & Heiles, C. 1964, Astron. Astrophys., 69, 73.

Hénon, M. 1966, Bulletin Astronomique 1, 3eme série.

Henrard, J., & Lemaitre, A. 1983, Cel.Mech., 30, 218.

Holman, M., & Murray, N. 1996, Astron. J. 112, 1278.

Hori, G.I. 1966, Publ. Astron. Soc. Japan, 18, 287.

Kalnajs, A. 1971, Astrophys. J., 166, 275.

Kalnajs, A. 1973, Proc. Astron. Soc. Australia, 2, 174.

Kaneko, K., & Konishi, T. 1989, Phys. Rev. A, 40, 6130.

Knežević, Z., Lemaitre, A., & Milani, A. 2002, in Bottke, W.F. et al. (eds), Asteroids III, University of Arizona Press, Tucson, pp. 603-612.

Knežević, Z., & Milani, A. 2005, Dynamics of Populations of Planetary Systems, Proceedings of the IAU Colloquium 197, Cambridge University Press.

Kolmogorov, A.N. 1954, Dokl. Akad. Nauk SSSR, 98, 527.

Laskar, J. 1993, Physica D, 67, 257.

Lazzaro, D., Ferraz-Mello, S., & Fernandez, J.A. 2006, Asteroids, Comets and Meteors, Proceedings of the IAU Symposium 229, Cambridge University Press.

Lega, E., Guzzo, M., & Froeschlé, C. 2003, Physica D, 182, 179.

Lega, E., Froeschlé, C., & Guzzo, M. 2008, Lect. Notes Phys., 729, 29.

Lega, E., Guzzo, M., & Froeschlé, C. 2009, Cel. Mech. Dyn. Astron., 104, 191.

Lega, E., Guzzo, M., & Froeschlé, C. 2010a, Cel. Mech. Dyn. Astron., 107, 129.

Lega, E., Guzzo, M., & Froeschlé, C. 2010b, Cel. Mech. Dyn. Astron., 107, 115.

Levison, H.F., Shoemaker, E.M., & Shoemaker, C.S. 1997, Nature 385, 42.

Lhotka, Ch., Efthymiopoulos, C., & Dvorak, R. 2008, Mon. Not. R. Astron. Soc., 384, 1165.

Lin, C.C., & Shu, F.H. 1964, Astrophys. J., 140, 646.

Lindblad, B. 1940, Astrophys. J., 92, 1.

Lindblad, B. 1956, Stockholm Obs. Ann., 19, 1.

Lindblad, B., Stockholm Obs. Ann. 1961, 21, 8.

Lochak, P. 1992, Russ. Math. Surv., 47, 57.

Lochak, P. 1999, in C. Simó (ed.), Hamiltonian Systems with Three or More Degrees of Freedom, NATO ASI Series C, 533, Kluwer, Dordrecht, 168.

Maffione, N.P., Darriba, L.A., Cincotta, P.M., & Giordano, C.M. 2011, Cel. Mech. Dyn. Astron. 111, 285.

Marzari, F., Tricarico, P., & Scholl, H. 2003, Mon. Not. Royal. Astron. Soc. 345, 1091.

Mather, J. 2004, J. Math. Sci., 124, 5275.

Mestre, M.F., Cincotta, P., & Giordano, C. 2012, in Cincotta, P., Giordano, C., & Efthymiopoulos, C. (eds), Proceedings of the Third La Plata School on Astronomy and Geophysics, AAA.

Mestre, M.F. 2012, Difusión caótica en sistemas Hamiltonianos casi-integrables, Ph.D Thesis, Universidad Nacional de La Plata.

Milani, A., & Knežević, Z. 1990, Cel. Mech. Dyn. Astron. 49, 347.

Milani, A., & Farinella, P. 1994, Nature 370, 40.

Morbidelli, A., & Guzzo, M. 1997, Cel. Mech. Dyn. Astron., 65, 107.

Morbidelli, A., & Giorgilli, A. 1997, Physica D, 102, 195.

Morbidelli, A., & Nesvorny, D 1999, Icarus 139, 295.

Morbidelli, A. 2002, Modern Celestial Mechanics. Aspects of Solar System Dynamics, Taylor and Francis, London.

Moser, J 1958,Commun. Pure Applied Math., 11, 1958.

Moser, J 1962, Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl II, 1-20.

Murray, C.D., & Dermott, S.F. 1999, Solar System Dynamics, Cambridge University Press.

Murray, N., & Holman, M. 1997, Astron. J. 114, 1246.

Murray, N., Holman, M., & Potter, M. 1998, Astron. J. 116, 2583.

Muzzio, J.C., Carpintero, D.D., & Wachlin, F.C. 2005, Cel. Mech. Dyn. Astron. 91, 173.

Neishtadt, A.I. 1984, J. Appl. Math. Mech., 48, 133.

Nekhoroshev, N.N. 1977, Russ. Math. Surv., 32(6), 1.

Nesvorny, D., & Morbidelli, A. 1998, Astron. J. 116, 3029.

Nesvorny, D., Bottke, W.F., Dones, L. et al. 2002, Nature 417, 720.

Nesvorny, D., Bottke, W.F., Levison, H. et al. 2003, Astrophys. J. 591, 486.

Ozorio de Almeida, A.M. 1988, Hamiltonian Systems: Chaos and Quantization, Cambridge University Press.

Ozorio de Almeida, A.M., & Viera, W.M. 1997, Phys. Lett. A, 227, 298.

Patsis, P.A., Contopoulos, G., & Grosbol, P. 1991, Astron. Astrophys., 243, 373.

Perozzi, E., & Ferraz-Mello, S. 2010, Space Manifold Dynamics, Springer.

Pinsky, M. 2002, Introduction to Fourier Analysis and Wavelets, Grad. Studies Math., Am. Math. Soc.

Poincaré, H 1892, Méthodes Nouvelles de la Mécanique Céleste, Gautier-Vilard, Paris.

Pöshel, J. 1993, Math. Z., 213, 187.

Robutel, P., & Gabern, F. 2006, Mon. Not. Royal Astron. Soc. 372, 1463.

Romero-Gomez, M., Masdemont, J.J., Athanassoula, E.M., & Garcia-Gomez, C., 2006, Astron. Astrophys, 453, 39.

Romero-Gomez, M., Athanassoula, E.M., Masdemont, J.J., & Garcia-Gomez, C. 2007, Astron. Astrophys., 472, 63.

Rosenbluth, M., Sagdeev, R., Taylor, J., & Zaslavskii, M. 1966, Nucl. Fusion, 6, 217.

Sansottera, M., Locatelli, U., & Giorgilli, A. 2011, Cel. Mech. Dyn. Astron. 111, 337.

Siegel, C.L., & Moser, J. 1991, Lectures on Celestial Mechanics, Springer, Heidelberg, 1991.

Simó, C., & Valls, C. 2001, Nonlinearity, 14, 1707.

Skokos, C., Contopoulos, G., & Polymilis, C. 1997, Cel. Mech. Dyn. Astr., 65, 223.

Tennyson, J. 1982, Physica D, 5, 123.

Tsiganis, K., Varvoglis, H., & Dvorak, R. 2005, Cel. Mech. Dyn. Astron. 92, 71.

Tsiganis, K., knežević, Z., & Varvoglis, H. 2007, Icarus, 186, 484.

Tsiganis, K. 2008, Lect. Notes Phys. 729, 111.

Tsiganis, K. 2010, Eur. Phys. J. Special Topics 186, 67.

Tsoutsis, P., Efthymiopoulos, C., & Voglis, N. 2008, Mon. Not. R. Astr. Soc., 387, 1264.

Tsoutsis, P., Kalapotharakos, C., Efthymiopoulos, C., & Contopoulos, G. 2009, Astron. Astrophys., 495, 743.

Valluri, M., Debattista, V.P., Quinn, T.R., Roškar, R., & Wadsley, J. 2012, Mon. Not. Royal Astron. Soc. 419, 1951.

Vieira, W.M., and A.M. Ozoiro de Almeida 1996, Physica D, 90, 9.

Voglis, N., Tsoutsis, P., & Efthymiopoulos, C. 2006, Mon. Not. R. Astr. Soc., 373, 280.

Whittaker, E.T. 1916, Proc. Roy Soc. Edinburgh, Sect. A, 37, 95.

Wisdom, J. 1980, Astron. J. 85, 1122.

Wisdom, J. 1983, Icarus 56, 51.

Wood, B.P., Lichtenberg, A.J., & Lieberman, M.A. 1990, Phys. Rev. A, 42, 5885.

Appendix: Optimal normalization order

In Efthymiopoulos et al. (2004) a method was proposed on how to practically estimate the growth of the size of the remainder function in a Birkhoff normalization procedure with Lie generating functions. This method relies upon carefully examining the accumulation of small divisors in the series terms; see also classical "constructive" proofs of either the Kolmogorov or the Nekhoroshev theorems (e.g. Giorgilli 1999, and references therein for a review, or Giorgilli and Locatelli 1997) based on a similar approach. We now examine how the same method applies in the case of the normalization algorithm developed in subsection 2.2.

We start by defining first the sequence a_r of smallest divisors appearing at any normalization order r. This is given by:

$$a_r = \min\{|k \cdot \omega_*| : \ |k| < K'r, \ k \notin \mathcal{M}\} \tag{232}$$

where K' is given by Eq.(155) and $\mathcal{M}$ is the resonant module defined in Eq.(152). In view of the homological equation defined in Eq.(160, the generating function χ_r acquires a term with divisor $k \cdot \omega_*$ for any Fourier term $\exp(ik \cdot \phi)$ of $\tilde{H}_r^{(r-1)}$. Then the bound

$$||\chi_r|| \leq \frac{||\tilde{H}_r^{(r-1)}||}{a_r} \tag{233}$$

holds, where norms are defined as explained in subsection 4.2.

We now examine how a divisor a_r appearing first at the normalization order r propagates at subsequent orders. The propagation of divisors is due to Eq.(159), implying that the divisor a_r appears also in $H^{(r+1)}$, since

$$H^{(r+1)} = H^{(r)} + \{H^{(r)}, \chi_r\} + \frac{1}{2}\{\{H^{(r)}, \chi_r\}, \chi_r\} + \dots \tag{234}$$

By repeating the two previous steps, we then find that a_r propagates at all subsequent orders after r. This, in turn, implies that $\tilde{H}_{r+1}^{(r+1)}$ contains divisors accumulated along all previous steps, i.e. by contributions from the generating functions $\chi_r, \chi_{r-1}, \dots$ via Poisson brackets like in Eq.(234). The same holds true for χ_{r+1} defined via the homological equation $\{\omega_* \cdot J, \chi_r\} + \tilde{H}_{r+1}^{(r+1)} = 0$. Define now the operators $\mathcal{G}_{j_{r,s}}^{(r)}$ acting on analytic functions f in the domains considered in subsection 2.1 via the recursive formulae

$$
\begin{aligned}
\mathcal{G}_{j_{r,s}}^{(r)} f &= \left(\prod_{q=1}^{r} \frac{1}{j_q!} \mathcal{L}_{\psi_q}^{j_q}\right) f, \quad j_{r,s} \equiv (j_1, j_2, \dots, j_r) \text{ with } j_q \geq 0, \quad \sum_{q=1}^{r} j_q = s \\
\psi_q &= -\mathcal{L}_{\omega_* \cdot J}^{-1} < \mathcal{G}_{j_{q-1,s}}^{(q-1)} f >_q
\end{aligned}
\tag{235}
$$

where $< \cdot >_q$ denotes the terms of order q in the book-keeping parameter of the included function belonging to the range of the operator $\mathcal{L}_{\omega_* \cdot J}$. The product in the first of Eqs.(235) is considered to be normally ordered, i.e. the operators

$\mathcal{L}_{j_q}$ are ordered from right to left for increasing q. The hamiltonian after r normalization steps is then given by

$$H^{(r)} = \exp\mathcal{L}_{\chi_r} \circ \exp\mathcal{L}_{\chi_{r-1}} \circ \ldots \exp\mathcal{L}_{\chi_1} H^{(0)} = \sum_{s=0}^{\infty} \sum_{\text{all possible } j_{r,s}} \mathcal{G}^{(r)}_{j_{r,s}} H^{(0)} \tag{236}$$

Let $\nu(w; j_{r,s})$ represent the number of times that an integer j_q in $j_{r,s}$ is equal to w. The key remark is that the asymptotic character of the normalization process is only due to sequences of terms generated by operators $\mathcal{G}^{(r)}_{j_{r,s}}$ for which $\nu(1; j_{r,s}) \neq 0$. In fact, it can be shown that for any fixed positive integer u, all sequences $b_u^{(r)}$, $r = 2, 3, \ldots$ of the form:

$$b_u^{(r)} = \sum_{\substack{\text{all possible } j_{r,r+u} \\ \text{with}_{\nu(w; j_{r,s})=0} \text{ if } _{w<2}}} ||\mathcal{G}^{(r)}_{j_{r,r+u}} H^{(0)}|| \tag{237}$$

are absolutely convergent in a sub-domain of the original domain of analyticity of f. The proof follows the same arguments as in subsection 3.4. Namely, one shows that the accumulation of divisors in sequences of the form (237) is at worst quadratic, and this suffices to establish an upper bound for the terms $b_{r,u}$ given by a geometric series. However, this is no longer guaranteed for operations in Eq.(236) involving $\mathcal{G}^{(r)}_{j_{r,s}}$ $\nu(1; j_{r,s}) \neq 0$, because among the latter there are cases leading to asymptotic rather than absolutely convergent sequences. In particular, the worst accumulation of divisors appears in the sequence:

$$d^{(r)} = ||\mathcal{G}^{(r)}_{j_1=j_2=\ldots=j_r=1} H^{(0)}|| \quad . \tag{238}$$

Similarly to what was observed in Efthymiopoulos et al. (2004), this particular sequence causes *repetitions* of small divisors that eventually lead to an asymptotic growth of the series coefficients. Starting from any Fourier term $h_{k,r_0} = c_k(J)\exp(ik \cdot \phi)$ in the Hamiltonian (156), this term is first encountered at the normalization order $r_0 = [|k|/K'] + 1$. The repetitions are found by considering the terms produced in the generating functions χ_{r_0}, χ_{r_0+1}, $\cdots$ due to Poisson brackets with the $O(\lambda)$ normal form terms. If I_* is such that both resonant vectors satisfy the inequalities $|k^{(1)}| \geq K'$ and $|k^{(2)}| > K'$ (this excludes only some very low order resonances), the only $O(\lambda)$ normal form terms are given by $Z_1 = \epsilon^{1/2}\sum_{i,j=1}^{3}(1/2)M_{ij}J_iJ_j$. Taking the successive normalization steps after the order r_0 we find the terms

In χ_{r_0}:
$\xi_{k,r_0} = \frac{h_{k,r_0}}{ik\cdot\omega_*}$ whereby $||\xi_{k,r_0}||_{\mathcal{W}_{I_*,B}} \leq \frac{||h_{k,r_0}||_{\mathcal{W}_{I_*,B}}}{a_{r_0}}$
In $H^{(r_0)}_{r_0+1}$:
$h_{k,r_0+1} \equiv \{Z_1, \xi_{k,r_0}\}$ whereby $||h_{k,r_0+1}||_{\mathcal{W}_{I_*,B}} \leq ||\xi_{k,r_0}||_{\mathcal{W}_{I_*,B}}\epsilon^{1/2}|M|B|k|$
$\leq \frac{||h_{k,r_0}||_{\mathcal{W}_{I_*,B}}K'\epsilon^{1/2}|M|Br_0}{a_{r_0}}$

In χ_{r_0+1}:

$\xi_{k,r_0+1} = \frac{h_{k,r_0+1}}{ik\cdot\omega_*}$ whereby $||\xi_{k,r_0+1}||_{\mathcal{W}_{I_*,B}} \leq \frac{||\xi_{k,r_0}||_{\mathcal{W}_{I_*,B}} K'\epsilon^{1/2}|M|Br_0}{a_{r_0}}$

etc., where $|M| = \max|M_{ij}|$. In general, the following inequalities are satisfied:

$$\frac{||\xi_{k,r_0+u}||_{\mathcal{W}_{I_*,B}}}{||\xi_{k,r_0+u-1}||_{\mathcal{W}_{I_*,B}}} \leq \frac{K'\epsilon^{1/2}|M|Br_0}{a_{r_0}}, \quad \frac{||h_{k,r_0+u}||_{\mathcal{W}_{I_*,B}}}{||h_{k,r_0+u-1}||_{\mathcal{W}_{I_*,B}}} \leq \frac{K'\epsilon^{1/2}|M|Br_0}{a_{r_0}}$$

whereby it follows that both sequences $||\xi_{k,r_0+u}||_{\mathcal{W}_{I_*,B}}$ and $||h_{k,r_0+u}||_{\mathcal{W}_{I_*,B}}$, $u = 1, 2, \ldots$ are bounded by a geometric series with ratio $K'\epsilon^{1/2}|M|Br_0/a_{r_0}$. That is, after being first encountered at order r_0, the same term h_{k,r_0} causes a sequence of repetitions of the same Fourier terms in all orders after r_0, the size of the term growing essentially by a geometric factor $\sim (K'\epsilon^{1/2}|M|Br_0/a_{r_0})^u$ at the order $r_0 + u$. Since, now, the original terms $h_{k,r}$ are of order $\sim Ae^{-\sigma|k|} \sim A\epsilon^{r_0/2}$, we finally arrive at an estimate for the size of the terms $d^{(r)}$ of the sequence (238), which itself represents an estimator of the size of the remainder $R^{(r)}$, namely:

$$||d^{(r)}||_{\mathcal{W}_{I_*,B}} \sim ||R^{(r)}||_{\mathcal{W}_{I_*,B}} \sim A\epsilon^{r/2} \sum_{r_0=2}^{r} \left(\frac{K'\epsilon^{1/2}|M|Br_0}{a_{r_0}}\right)^{r-r_0} . \quad (239)$$

Eq.(239) implies an asymptotic growth of the size of $||R^{(r)}||_{\mathcal{W}_{I_*,B}}$ with r, since, e.g. close to the median of the summation values of r_0, i.e. $r_0 \simeq r/2$, one has terms of size $\sim (\epsilon^{1/2}Br/(2a_{r/2}))^{r/2} \sim [(\epsilon^{1/2}Be/(2a_{r/2}))^r r!]^{1/2}$, which grow at least as fast as $(B\epsilon^{1/2})^r r!$. The exact growth rate depends on whether $\mathcal{W}_{I_*,B}$ is a doubly, simply, or non-resonant domain with respect to a high enough K-truncation, since one has, $a_r = \min|\omega_*|_i$ in doubly-resonant domains, while $a_r \sim \gamma/r^\tau$ with γ a positive constant $\gamma = O(\min|\omega_*|_i)$, and $\tau = 1$ for simple resonance or $\tau = 2$ for non-resonance (in three degrees of freedom, assuming diophantine behavior of the frequencies). Accordingly, we arrive at the remainder estimate $||R^{(r)}||_{\mathcal{W}_{I_*,B}} \sim (B\epsilon^{1/2})^r r!^{1+\tau}$, whence, using standard approximations (e.g. Stirling's formula), we finally find

Double resonance: $\quad r_{opt} \sim \dfrac{1}{\epsilon^{1/2}}, \quad ||R^{(r_{opt})}|| \sim \exp\left((\epsilon_0/\epsilon)^{-1/2}\right)$

Simple resonance: $\quad r_{opt} \sim \dfrac{1}{\epsilon^{1/4}}, \quad ||R^{(r_{opt})}|| \sim \exp\left((\epsilon_0/\epsilon)^{-1/4}\right)$ $\quad (240)$

Non-resonance: $\quad r_{opt} \sim \dfrac{1}{\epsilon^{1/6}}, \quad ||R^{(r_{opt})}|| \sim \exp\left((\epsilon_0/\epsilon)^{-1/6}\right)$

that is, we recover Eq.(163), and its associated exponential estimates for the size of the remainder, with exponent a as indicated in Eq.(240).

LECTURE

Methods of algebraic manipulation in perturbation theory

A. Giorgilli[1,2], M. Sansottera[3]

(1) Dipartimento di Matematica, Università degli Studi di Milano,
    Via Saldini 50, 20133 — Milano, Italy
(2) Istituto Lombardo Accademia di Scienze e Lettere
(3) naXys, Namur Center for Complex Systems, FUNDP,
    Rempart de la Vierge 8, B5000 — Namur, Belgium

Abstract. We give a short introduction to the methods of representing polynomial and trigonometric series that are often used in Celestial Mechanics. A few applications are also illustrated.

1. Overview

Algebraic manipulation on computer is a tool that has been developed quite soon, about one decade after the birth of computers, the first examples dating back to the end of the fifties of the last century. General purpose packages began to be developed during the sixties, and include, e.g., Reduce (1968), Macsyma (1978), muMath (1980), Maple (1984), Scratchpad (1984), Derive (1988), Mathematica (1988), Pari/GP (1990) and Singular (1997) (the dates refer to the first release). However, most of the facilities of these general purpose manipulators are simply ignored when dealing with perturbation methods in Celestial Mechanics. For this reason, the job of developing specially devised manipulation tools has been undertaken by many people, resulting in packages that have limited capabilities, but are definitely more effective in practical applications. Producing a list of these packages is a hard task, mainly because most of them are not publicly available. A list of "old time" packages may be found in Henrard (1989) and Laskar (1989). In recent times a manipulator developed by J. Laskar and M. Gastineau has become quite known.

Finding references to the methods implemented in specially devised packages is as difficult as giving a list. We know only a few papers by Broucke and Garthwaite (1969), Broucke (1989), Rom (1970), Henrard (1986 and 1989), Laskar (1989), Jorba (1999) and Biscani (2009). A complete account of the existing literature on the subject goes beyond the limits of the present note. The present work introduces some ideas that have been used by the authors in order to implement a package named $X\varrho\acute{o}\nu o\varsigma$.

As a matter of fact, most of the algebraic manipulation used in Celestial Mechanics makes use of the so called "Poisson series", namely series with a general term of the form

$$x_1^{j_1} \cdot \ldots \cdot x_n^{j_n} \frac{\cos}{\sin}(k_1\varphi_1 + \ldots + k_m\varphi_m) ,$$

(with obvious meaning of the symbols). Thus, a very minimal set of operations is required, namely sums, products and derivatives of polynomials and/or trigonometric polynomials. Traditionally, also the operation of inversion of functions, usually made again via series expansion, was required. However, the expansion methods based on Lie series and Lie transforms typically get rid of the latter operation (see, e.g., Henrard 1974).

Writing a program doing algebraic manipulation on series of the type above leads one to be confronted with a main question, namely how to represent a polynomial, trigonometric polynomial or Poisson series on a computer. The papers quoted above actually deal with this problem, suggesting some methods. In these lectures we provide an approach to this problem, followed by a few examples of applications.

In sect. 2 we include a brief discussion about the construction of normal form for a Hamiltonian system in the neighborhood of an elliptic equilibrium. We do not attempt to give a complete discussion, since it is available in many papers. We rather try to orient the reader's attention on the problem of representing perturbation series.

In sect. 3–7 we introduce a method which turns out to be quite useful for the representation of a function as an array of coefficients. The basic idea has been suggested to one of the authors by the paper of Gustavson (1966) (who, however, just mentions that he used an indexing method, without giving any detail about its implementation). One introduces an *indexing function* which transforms an array of exponents in a polynomial (or trigonometric polynomial) in a single index within an array. The general scheme is described in sect. 3. The basics behind the construction of an indexing function are described in sect. 4. The details concerning the representation of polynomials and trigonometric polynomials are reported in sects. 5 and 6, respectively. In sect. 7 we include some hints about the case of sparse series, that may be handled by combining the indexing functions above with a tree representation. Finally, sect. 8 is devoted to three applications, by giving a short account of the contents of published papers.

2. A common problem in perturbation theory

A typical application of computer algebra is concerned with the construction of first integrals or of a normal form for a Hamiltonian system. A nontrivial example, which however may be considered as a good starting point, is the calculation of a normal form for the celebrated model of Hénon and Heiles (1964), which has been done by Gustavson (1966). Some results on this model are reported in sect. 8.

We assume that the reader is not completely unfamiliar with the concept of normal form for a (possibly Hamiltonian) system of differential equations. Thus, let us briefly illustrate the problem by concentrating our attention on the algorithmic aspect and by explaining how algebraic manipulation may be introduced.

2.1. Computation of a normal form

Let us consider a canonical system of differential equations in the neighborhood of an elliptic equilibrium. The Hamiltonian may typically be given the form of

a power series expansion

$$H(x,y) = H_0(x,y) + H_1(x,y) + \dots , \quad H_0(x,y) = \sum_{j=1}^{n} \frac{\omega_j}{2}(x_j^2 + y_j^2) , \qquad (1)$$

where $H_s(x,y)$ for $s \geq 1$ is a homogeneous polynomial of degree $s + 2$ in the canonical variables $(x,y) \in \mathbb{R}^{2n}$. Here $\omega \in \mathbb{R}^n$ is the vector of the frequencies, that are assumed to be all different from zero.

In such a case the system is said to be in Birkhoff normal form in case the Hamiltonian takes the form

$$H(x,y) = H_0(x,y) + Z_1(x,y) + Z_2(x,y) + \dots \quad \text{with} \quad L_{H_0}Z_s = 0 , \qquad (2)$$

where $L_{H_0} \cdot = \{H_0, \cdot\}$ is the Lie derivative with respect to the flow of H_0, actually the Poisson bracket with H_0.

The concept of Birkhoff normal form is better understood if one assumes also that the frequencies are non resonant, i.e., if

$$\langle k, \omega \rangle \neq 0 \quad \text{for all } k \in \mathbb{Z}^n , \ k \neq 0 ,$$

where $\langle k, \omega \rangle = \sum_j k_j \omega_j$. For, in this case the functions $Z_s(x,y)$ turn out to be actually function only of the n actions of the system, namely of the quantities

$$I_j = \frac{x_j^2 + y_j^2}{2} , \quad j = 1, \dots, n .$$

It is immediate to remark that $I_1, \dots, I_n$ are independent first integrals for the Hamiltonian, an that they are also in involution, so that, by Liouville's theorem, the system turns out to be integrable. The definition of normal form given in (2) is more general, since it includes also the case of resonant frequencies.

The calculation of the normal form may be performed using the Lie transform method, which turns out to be quite effective. We give here the algorithm without proof. A complete description may be found, e.g., in (Giorgilli 1978), and the description of a program implementing the method via computer algebra is given in (Giorgilli 1979). The corresponding `FORTRAN` program is available from the CPC library.

The Lie transform is defined as follows. Let a *generating sequence* $\chi_1(x,y)$, $\chi_2(x,y), \dots$ be given, and define the operator

$$T_\chi = \sum_{s \geq 0} E_s \qquad (3)$$

where the sequence $E_0, E_1, \dots$ of operators is recursively defined as

$$E_0 = 1 , \quad E_s = \sum_{j=1}^{s} \frac{j}{s} L_{\chi_j} E_{s-j} \qquad (4)$$

This is a linear operator that is invertible and satisfies the interesting properties

$$T_\chi \{f, g\} = \{T_\chi f, T_\chi g\} , \quad T_\chi(f \cdot g) = T_\chi f \cdot T_\chi g . \qquad (5)$$

Let now $Z(x, y) = H_0(x, y) + Z_1(x, y) + Z_2(x, y) + \dots$ be a function such that

$$T_\chi Z = H \;, \tag{6}$$

where H is our original Hamiltonian, and let Z possess a first integral Φ, i.e., a function satisfying $\{Z, \Phi\} = 0$. Then one has also

$$T_\chi \{Z, \Phi\} = \{T_\chi Z, T_\chi \Phi\} = \{H, T_\chi \Phi\} = 0 \;,$$

which means that *if Φ is a first integral for Z then $T_\chi \Phi$ is a first integral for H.*

The question now is: *can we find a generating sequence $\chi_1, \chi_2, \dots$ such that the function Z satisfying* (6) *is in Birkhoff normal form?*

The answer to this question is in the positive, and the generating sequence may be calculated via an explicit algorithm that can be effectively implemented via computer algebra. We include here the algorithm, referring to, e.g., (Giorgilli 1978) for a complete deduction. Here we want only to stress that all operations that are required may be actually implemented on a computer.

The generating sequence is determined by solving for χ and Z the equations

$$Z_s - L_{H_0} \chi_s = H_s + Q_s \;, \quad s \geq 1 \;, \tag{7}$$

where Q_s is a known homogeneous polynomial of degree $s + 2$ given by $Q_1 = 0$ and

$$Q_s = -\sum_{j=1}^{s-1} \left(E_j Z_{s-j} + \frac{j}{s} \{\chi_j, E_{s-j} H_0\} \right) \;, \quad s > 1 \;.$$

In order to solve (7) it is convenient to introduce complex variables ξ, η via the canonical transformation

$$x_j = \frac{1}{\sqrt{2}} (\xi_j + i\eta_j) \;, \quad y_j = \frac{i}{\sqrt{2}} (\xi_j - i\eta_j)$$

which transforms $H_0 = i \sum_j \omega_j \xi_j \eta_j$. In these variables the operator L_{H_0} takes a diagonal form, since

$$L_{H_0} \xi^j \eta^k = i \langle k - j, \omega \rangle \xi^j \eta^k \;,$$

where we have used the multi-index notation $\xi^j = \xi_1^{j_1} \cdot \dots \cdot \xi_n^{j_n}$, and similarly for η. Thus, writing the r.h.s. of (7) as a sum of monomials $c_{j,k} \xi^j \eta^k$ the most direct form of the solution is found by including in Z all monomials with $\langle k - j, \omega \rangle = 0$, and adding $\frac{c_{j,k}}{i \langle k-j, \omega \rangle} \xi^j \eta^k$ to χ_s for all monomials with $\langle k - j, \omega \rangle \neq 0$. This is the usual way of constructing a normal form for the system (1).

Let us now examine in some more detail the algebraic aspect. With a little amount of patience one can verify that (7) involves only homogeneous polynomials of degree $s + 2$. Thus, one should be able to manipulate this kind of functions. Moreover, a careful examination of the algorithm shows that there are just elementary algebraic operations that are required, namely:

(i) sums and multiplication by scalar quantities;

(ii) Poisson brackets, which actually require derivatives of monomials, sums and products;

(iii) linear substitution of variables, which may still be reduced to calculation of sums and products without affecting the degree of the polynomial;

(iv) solving equation (7), which just requires a division of coefficients.

These remarks should convince the reader that implementing the calculation of the normal form via algebraic manipulation on a computer is just matter of being able of representing homogeneous polynomials in many variables and performing on them a few elementary operations, such as sum, product and derivative.

2.2.   A few elementary considerations

In order to have an even better understanding the reader may want to consider the elementary problem of representing polynomials in one single variable. We usually write such a polynomial of degree s (non homogeneous, in this case) as

$$f(x) = a_0 + a_1 x + \ldots + a_s x^s \ .$$

A machine representation is implemented by storing the coefficients $a_0, a_1, \ldots, a_n$ as a one-dimensional array of floating point quantities, either real or complex. E.g., in `FORTRAN` language one can represent a polynomial of degree 100 by just saying, e.g., `DIMENSION F(101)` and storing the coefficient a_j as `F(j+1)` (here we do not use the extension of `FORTRAN` that allows using zero or even negative indices for an array). Similarly in a language like `C` one just says, e.g., `double f[101]` and stores a_j as `f[j]`.

The operation of sum is a very elementary one: if f, g are two polynomials and the coefficients are stored in the arrays `f,g` (in `C` language) then the sum h is the array `h` with elements `h[j] = f[j] + g[j]`. The derivative of f is the array `fp` with elements `fp[j] = (j+1)*f[j+1]`. In a similar way one can calculate the product, by just translating in a programming language the operations that are usually performed by hand.

The case of polynomials in two variables is just a bit more difficult. A homogeneous polynomial of degree s is usually written as

$$f(x, y) = a_{s,0} x^s + a_{s-1,1} x^s y + \ldots + a_{0,s} y^s \ .$$

The naive (not recommended) representation would use an array with two indices (a matrix), by saying, e.g., `DIMENSION F(101,101)` and storing the coefficient $a_{j,k}$ as `F(j+1,k+1)`. Then the algebra is just a straightforward modification with respect to the one-dimensional case.

Such a representation is not recommended for at least two reasons. The first one is that arrays with arbitrary dimension are difficult to use, or even not allowed, in programming languages. The second and more conclusive reason is that such a method turns out to be very effective in wasting memory space. E.g., in the two dimensional case a polynomial of degree up to s requires a matrix with $(s+1)^2$ elements, while only $(s+1)(s+2)/2$ are actually used. Things go much worse in higher dimension, as one easily realizes.

The arguments above should have convinced the reader that an effective method of representing polynomials is a basic tool in order to perform computer algebra for problems like the calculation of normal form. Once such a method

is available, the rest is essentially known algebra, that needs to be translated in a computer language.

The problem for Poisson series is a similar one, as the reader can easily imagine. The following sections contains a detailed discussion of indexing methods particularly devised for polynomials and for Poisson series. The underlying idea is to represent the coefficients as a one-dimensional array by suitably packing them in an effective manner, so as to avoid wasting of space.

3.   General scheme

The aim of this section is to illustrate how an appropriate algebraic structure may help in representing the particular classes of functions that usually appear in perturbation theory. We shall concentrate our attention only on polynomials and trigonometric polynomials, which are the simplest and most common cases. However, the reader will see that most of the arguments used here apply also to more general cases.

3.1.   Polynomials and power series

Let $\mathcal{P}$ denote the vector space of polynomials in the independent variables $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. A basis for this vector space is the set $\{u_k(x)\}_{k \in \mathbb{Z}_+^n}$, where

$$u_k(x) = x^k \equiv x_1^{k_1} \cdot \ldots \cdot x_n^{k_n} \ . \tag{8}$$

In particular, we shall consider the subspaces $\mathcal{P}_s$ of $\mathcal{P}$ that contain all homogeneous polynomials of a given degree $s \geq 0$; the subspace $\mathcal{P}_0$ is the one-dimensional space of constants, and its basis is $\{1\}$. The relevant algebraic properties are the following:

(i) every subspace $\mathcal{P}_s$ is closed with respect to sum and multiplication by a number, i.e., if $f \in \mathcal{P}_s \ \wedge \ g \in \mathcal{P}_s$ then $f + g \in \mathcal{P}_s$ and $\alpha f \in \mathcal{P}_s$;

(ii) the product of homogeneous polynomials is a homogeneous polynomial, i.e., if $f \in \mathcal{P}_r \ \wedge \ g \in \mathcal{P}_s$ then $fg \in \mathcal{P}_{r+s}$;

(iii) the derivative with respect to one variable maps homogeneous polynomials into homogeneous polynomials, i.e., if $f \in \mathcal{P}_s$ then $\partial_{x_j} f \in \mathcal{P}_{s-1}$; if $s = 0$ then $\partial_{x_j} f = 0$, of course.

These three properties are the basis for most of the algebraic manipulations that are commonly used in perturbation theory.

A power series is represented as a sum of homogeneous polynomials. Of course, in practical calculations the series will be truncated at some order. Since every homogeneous polynomial $f \in \mathcal{P}_s$ can be represented as

$$f(x) = \sum_{|k|=s} f_k u_k(x) \ ,$$

it is enough to store in a suitable manner the coefficients f_k. A convenient way, particularly effective when most of the coefficients are different from zero, is

Table 1.   Illustrating the function representation for power series. A memory block is assigned to the function $f(x)$. The coefficient f_k of $u_k(x)$ is stored at the address resulting by adding the offset $I(k)$ to the starting address of the memory block.

$f_{(0,0,\ldots,0)}$	$\leftarrow 0 = I(k) \leftarrow k = (0,0,\ldots,0)$
$f_{(1,0,\ldots,0)}$	$\leftarrow 1 = I(k) \leftarrow k = (1,0,\ldots,0)$
$\cdots$ $\cdots$ $\cdots$	
$f_{(k_1,k_2,\ldots,k_n)}$	$\leftarrow I(k) \leftarrow k = (k_1,k_2,\ldots,k_n)$
$\cdots$ $\cdots$ $\cdots$	

based on the usual lexicographic ordering of polynomials (to be pedantic, inverse lexicographic). E.g., a homogeneous polynomial of degree s in two variables is ordered as

$$a_{s,0}x_1^s + a_{s-1,1}x_1^{s-1}x_2 + \ldots + a_{0,s}x_2^s \ .$$

The idea is to use the position of a monomial x^k in the lexicographic order as an index $I(k_1,\ldots,k_n)$ in an array of coefficients. We call I and *indexing function*. Here we illustrate how to use it, deferring to sect. 5 the actual construction of the function.

The method is illustrated in table 1. Let f be a power series, truncated at some finite order s. A memory block is assigned to f. The size of the block is easily determined as $I\big((0,\ldots,0,s)\big)$. For, $(0,\ldots,0,s)$ is the last vector of length s. The starting address of the block is assigned to the coefficient of $u_{(0,0,\ldots,0)}$; the next address is assigned to the coefficient of $u_{(1,0,\ldots,0)}$, because $(1,0,\ldots,0)$ is the first vector of length 1, and so on. Therefore, the address assigned to the coefficient of $u_{(k_1,\ldots,k_n)}$ is the starting address of the block incremented by $I\big((k_1,\ldots,k_n)\big)$. If f is a homogeneous polynomial of degree s the same scheme works fine with a few minor differences: the length of the block is $I\big((0,\ldots,0,s)\big) - I\big((0,\ldots,0,s-1)\big)$, the starting address of the block is associated to the coefficient of $u_{(s,0,\ldots,0)}$, and the coefficient of $u_{(k_1,\ldots,k_n)}$ is stored at the relative address $I\big((k_1,\ldots,k_n)\big) - I\big((0,\ldots,0,s-1)\big)$. This avoids leaving an empty space at the top of the memory block.

In view of the form above of the representation a function is identified with a set of pairs (k, f_k), where $k \in \mathbb{Z}_+^n$ is the vector of the exponents, acting as the label of the elements of the basis, and f_k is the numerical coefficient. Actually the vector k is not stored, since it is found via the index. The algebraic operations of sum, product and differentiation can be considered as operations on the latter set.

(i) If $f, g \in \mathcal{P}_s$ then the operation of calculating the sum $f + g$ is represented as

$$\left.\begin{array}{c}(k,\, f_k) \\ (k,\, g_k)\end{array}\right\} \mapsto (k,\, f_k + g_k) \ ,$$

to be executed over all k such that $|k| = s$.

(ii) If $f \in \mathcal{P}_r$ and $g \in \mathcal{P}_s$ then the operation of calculating the product fg is represented as

$$\left.\begin{array}{c}(k,\, f_k) \\ (k',\, g_{k'})\end{array}\right\} \mapsto (k + k',\, f_k g_{k'}) \ ,$$

to be executed over all k, k' such that $|k| = r$ and $|k'| = s$.

(iii) If $f \in \mathcal{P}_s$ then the operation of differentiating f with respect to, e.g., x_1 is represented as

$$(k, f_k) \mapsto \begin{cases} \emptyset & \text{for } k_1 = 0 \ , \\ (k', k_1 f_k) & \text{for } k_1 \neq 0 \ , \end{cases}$$

where $k' = (k_1 - 1, k_2, \ldots, k_n)$.

It is perhaps worthwhile to spend a few words about how to make the vector k to run over all its allowed values. In the case of sum, we do not really need it: since the indexes of both addends and of the result are the same, the operation can actually be performed no matter which k is involved: just check that the indexes are in the correct range.[1] In order to perform product and differentiation it is essential to know the values of k and k'. To this end, we can either use the inverse of the indexing function, or generate the whole sequence by using a function that gives the vector next to a given k.

3.2.   Fourier series

Let us denote by $\varphi = (\varphi_1 \ldots, \varphi_n) \in \mathbb{T}^n$ the independent variables. The Fourier expansion of a real function on $\mathbb{T}^n$ takes the form

$$f(\varphi) = \sum_{k \in \mathbb{Z}^n} \left(a_k \cos\langle k, \varphi \rangle + b_k \sin\langle k, \varphi \rangle \right) \ , \tag{9}$$

where a_k and b_k are numerical coefficients. In this representation there is actually a lot of redundancy: in view of $\cos(-\alpha) = \cos\alpha$ and $\sin(-\alpha) = -\sin\alpha$ the modes $-k$ and k can be arbitrarily interchanged. On the other hand, it seems that we actually need two different arrays for the sin and cos components, respectively. A straightforward way out is to use the exponential representation $\sum_k a_k e^{i\langle k, \varphi \rangle}$, but a moment's thought leads us to the conclusion that the redundancy is not removed at all. However, we can at the same time remove the redundancy and reduce the representation to a single array by introducing a suitable basis $\{u_k(\varphi)\}_{k \in \mathbb{Z}^n}$. Let $k \in \mathbb{Z}^n$; we shall say that k is *even* if the first non zero

---

[1]For a homogeneous polynomial of degree s the first vector is $(s, 0, \ldots, 0)$, and the last one is $(0, \ldots, 0, s)$. The indexes of these two vectors are the limits of the indexes in the sum.

component of k is positive, and that k is *odd* if the first non zero component of k is negative. The null vector $k = 0$ is said to be even. Then we set

$$u_k(\varphi) = \begin{cases} \cos\langle k, \varphi \rangle & \text{for } k \text{ even} , \\ \sin\langle k, \varphi \rangle & \text{for } k \text{ odd} . \end{cases} \tag{10}$$

This makes the representation $f(\varphi) = \sum_{k \in \mathbb{Z}^n} \varphi_k u_k(\varphi)$ unique and redundancy free. It may be convenient to remark that the notation for the sin function may create some confusion. Usually, working with one variable, we write $\sin \varphi$. The convention above means that we should rather write $-\sin(-\varphi)$, which is correct, but a bit funny. This should be taken into account when, after having accurately programmed all the operations, we discover that our manipulator says, e.g., that $\frac{\mathrm{d}}{\mathrm{d}\varphi} \cos \varphi = -\sin(-\varphi)$.

In view of the discussion in the previous section it should now be evident that a truncated Fourier expansion of a function $f(\varphi)$ can easily be represented by storing the coefficient of $u_k(\varphi)$ at an appropriate memory address, as calculated by the indexing function $I(k)$ of sect. 6.

The considerations of the previous section can be easily extended to the problem of calculating the sum and/or product of two functions, and of differentiating a function. Let us identify any term of the Fourier expansion of the function f with the pair (k, f_k). Let us also introduce the functions $\mathrm{od}(k)$ and $\mathrm{ev}(k)$ as follows: if k is odd, then $\mathrm{od}(k) = k$ and $\mathrm{ev}(k) = -k$; else $\mathrm{od}(k) = -k$ and $\mathrm{ev}(k) = k$. That, is, force k to be odd or even, as needed, by possibly changing its sign.

(i) Denoting by (k, f_k) and (k, g_k) the same Fourier components of two functions f and g, respectively, the sum is computed as

$$\left. \begin{array}{c} (k, \, f_k) \\ (k, \, g_k) \end{array} \right\} \mapsto (k, f_k + g_k) . \tag{11}$$

(ii) Denoting by (k, f_k) and $(k', g_{k'})$ any two terms in the Fourier expansion of the functions f and g, respectively, the product is computed as

$$\left. \begin{array}{c} (k, \, f_k) \\ (k', \, g_{k'}) \end{array} \right\} \mapsto \begin{cases} \left(\mathrm{ev}(k+k'), \dfrac{f_k g_{k'}}{2} \right) \cup \left(\mathrm{ev}(k-k'), \dfrac{f_k g_{k'}}{2} \right) & \text{for } k \text{ even} \\ & \text{and } k' \text{ even} , \\[2mm] \left(\mathrm{od}(k+k'), \dfrac{f_k g_{k'}}{2} \right) \cup \left(\mathrm{od}(k-k'), -\dfrac{f_k g_{k'}}{2} \right) & \text{for } k \text{ even} \\ & \text{and } k' \text{ odd} , \\[2mm] \left(\mathrm{od}(k+k'), \dfrac{f_k g_{k'}}{2} \right) \cup \left(\mathrm{od}(k-k'), \dfrac{f_k g_{k'}}{2} \right) & \text{for } k \text{ odd} \\ & \text{and } k' \text{ even} , \\[2mm] \left(\mathrm{ev}(k+k'), -\dfrac{f_k g_{k'}}{2} \right) \cup \left(\mathrm{ev}(k-k'), \dfrac{f_k g_{k'}}{2} \right) & \text{for } k \text{ odd} \\ & \text{and } k' \text{ odd} . \end{cases}$$
$$\tag{12}$$

Remark that the product always produces two distinct terms, unless $k = 0$ or $k' = 0$.

(iii) Denoting by (k, f_k) any term in the Fourier expansion of a function f, differentiation with respect to, e.g., φ_1 is performed as

$$(k, f_k) \mapsto \begin{cases} (-k, -k_1 f_k) & \text{for } k \text{ even}, \\ (-k, k_1 f_k) & \text{for } k \text{ odd}. \end{cases} \tag{13}$$

All these formulæ follow from well known trigonometric identities.

4.   Indexing functions

The basic remark for constructing an index function is the following. Suppose that we are given a countable set $\mathcal{A}$. Suppose also that $\mathcal{A}$ is equipped with a relation of complete ordering, that we shall denote by the symbols $\prec, \preceq, \succ$ and $\succeq$. So, for any two elements $a, b \in \mathcal{A}$ exactly one of the relations $a \prec b$, $a = b$ and $b \succ a$ is true. Suppose also that there is a minimal element in $\mathcal{A}$, i.e., there is $a_0 \in \mathcal{A}$ such that $a \succ a_0$ for all $a \in \mathcal{A}$ such that $a \neq a_0$. Then an index function I is naturally defined as

$$I(a) = \#\{b \in \mathcal{A} \,:\, b \prec a\} \,. \tag{14}$$

If $\mathcal{A}$ is a finite set containing N elements, then $I(\mathcal{A}) = \{0, 1, \ldots, N-1\}$. If $\mathcal{A}$ is an infinite (but countable) set, then $I(\mathcal{A}) = \mathbb{Z}_+$, the set of non negative integers. For instance, the trivial case is $\mathcal{A} = \mathbb{Z}_+$ equipped with the usual ordering relation. In such a case the indexing function is just the identity.

Having defined the function $I(a)$, we are interested in performing the following basic operations:

(i) for a given $a \in \mathcal{A}$, find the index $I(a)$;

(ii) for a given $a \in \mathcal{A}$, find the element next (or prior) to a, if it exists;

(iii) for a given $l \in I(\mathcal{A})$, find $I^{-1}(l)$, i.e., the element $a \in \mathcal{A}$ such that $I(a) = l$.

The problem here is to implement an effective construction of the index for some particular subsets of $\mathbb{Z}^n$ that we are interested in. In order to avoid confusions, we shall use the symbols $\prec, \preceq, \succ$ and $\succeq$ when dealing with an ordering relation in the subset of $\mathbb{Z}^n$ under consideration. The symbols $<, \leq, \geq$ and $>$ will always denote the usual ordering relation between integers.

As a first elementary example, let us consider the case $\mathcal{A} = \mathbb{Z}$. The usual ordering relation $<$ does not fulfill our requests, because there is no minimal element. However, we can construct a different ordering satisfying our requests as follows.

Let $k, k' \in \mathbb{Z}$. We shall say that $k' \prec k$ in case one of the following relations is true:

(i) $|k'| < |k|$;

(ii) $|k'| = |k| \,\wedge\, k' > k$.

The resulting order is $0, 1, -1, 2, -2, \ldots$, so that 0 is the minimal element.

Constructing the indexing function in this case is easy. Indeed, we have

$$I(0) = 0 \; , \quad I(a) = \begin{cases} 2a - 1 & \text{for } a > 0 \; , \\ -2a & \text{for } a < 0 \; . \end{cases} \tag{15}$$

The inverse function is also easily constructed:

$$I^{-1}(0) = 0 \; , \quad I^{-1}(l) = \begin{cases} (l+1)/2 & \text{for } l \text{ odd} \; , \\ -l/2 & \text{for } l \text{ even} \; . \end{cases} \tag{16}$$

In the rest of this section we show how an indexing function can be constructed for two particularly interesting cases, namely polynomials and trigonometric polynomials. However, we stress that the procedure we are using is a quite general one, so it can be extended to other interesting situations.

5. The polynomial case

Let us first take $\mathcal{A}_n = \mathbb{Z}_+^n$, i.e., integer vectors with non negative components; formally

$$\mathcal{A}_n = \{k = (k_1, \ldots, k_n) \in \mathbb{Z}^n \; : \; k_1 \geq 0, \ldots, k_n \geq 0\} \; .$$

The index n in $\mathcal{A}_n$ denotes the dimension of the space. This case is named "polynomial" because it occurs precisely in the representation of homogeneous polynomials, and so also in the Taylor expansion of a function of n variables: the integer vectors $k \in \mathcal{A}_n$ represent all possible exponents.

We shall denote by $|k| = k_1 + \ldots + k_n$ the length (or norm) of the vector $k \in \mathbb{Z}_+^n$. Furthermore, to a given vector $k = (k_1, \ldots, k_n) \in \mathcal{A}_n$ we shall associate the vector $t(k) \in \mathcal{A}_{n-1}$ (the *tail* of k) defined as $t(k) = (k_2, \ldots, k_n)$. This definition is meaningful only if $n > 1$, of course.

5.1. Ordering relation

Pick a fixed n, and consider the finite family of sets $\mathcal{A}_1 = \mathbb{Z}_+, \ldots, \mathcal{A}_n = \mathbb{Z}_+^n$.

Let $k, k' \in \mathcal{A}_m$, with any $1 \leq m \leq n$. We shall say that $k' \prec k$ in case one of the following conditions is true:

(i) $m \geq 1 \; \wedge \; |k'| < |k|$;

(ii) $m > 1 \; \wedge \; |k'| = |k| \; \wedge \; k_1' > k_1$;

(iii) $m > 1 \; \wedge \; |k'| = |k| \; \wedge \; k_1' = k_1 \; \wedge \; t(k') \prec t(k)$.

In table 2 the ordering resulting from this definition is illustrated for the cases $m = 2, 3, 4, 5$.

If $m = 1$ then only (i) applies, and this ordering coincides with the natural one in $\mathbb{Z}_+$. For $m > 1$, if (i) and (ii) are both false, then (iii) means that one must decrease the dimension n by replacing k with its tail $t(k)$, and retry the comparison. For this reason the ordering has been established for $1 \leq m \leq n$. Eventually, one ends up with $m = 1$, to which only (i) applies.

Table 2.   Ordering of integer vectors in $\mathbb{Z}_+^m$ for $m = 2, 3, 4, 5$.

$I(k)$	$m = 2$	$m = 3$	$m = 4$	$m = 5$
0	$(0,0)$	$(0,0,0)$	$(0,0,0,0)$	$(0,0,0,0,0)$
1	$(1,0)$	$(1,0,0)$	$(1,0,0,0)$	$(1,0,0,0,0)$
2	$(0,1)$	$(0,1,0)$	$(0,1,0,0)$	$(0,1,0,0,0)$
3	$(2,0)$	$(0,0,1)$	$(0,0,1,0)$	$(0,0,1,0,0)$
4	$(1,1)$	$(2,0,0)$	$(0,0,0,1)$	$(0,0,0,1,0)$
5	$(0,2)$	$(1,1,0)$	$(2,0,0,0)$	$(0,0,0,0,1)$
6	$(3,0)$	$(1,0,1)$	$(1,1,0,0)$	$(2,0,0,0,0)$
7	$(2,1)$	$(0,2,0)$	$(1,0,1,0)$	$(1,1,0,0,0)$
8	$(1,2)$	$(0,1,1)$	$(1,0,0,1)$	$(1,0,1,0,0)$
9	$(0,3)$	$(0,0,2)$	$(0,2,0,0)$	$(1,0,0,1,0)$
10	$(4,0)$	$(3,0,0)$	$(0,1,1,0)$	$(1,0,0,0,1)$
11	$(3,1)$	$(2,1,0)$	$(0,1,0,1)$	$(0,2,0,0,0)$
12	$(2,2)$	$(2,0,1)$	$(0,0,2,0)$	$(0,1,1,0,0)$
13	$(1,3)$	$(1,2,0)$	$(0,0,1,1)$	$(0,1,0,1,0)$
14	$(0,4)$	$(1,1,1)$	$(0,0,0,2)$	$(0,1,0,0,1)$
15	$(5,0)$	$(1,0,2)$	$(3,0,0,0)$	$(0,0,2,0,0)$
16	$(4,1)$	$(0,3,0)$	$(2,1,0,0)$	$(0,0,1,1,0)$
17	$(3,2)$	$(0,2,1)$	$(2,0,1,0)$	$(0,0,1,0,1)$
18	$(2,3)$	$(0,1,2)$	$(2,0,0,1)$	$(0,0,0,2,0)$
19	$(1,4)$	$(0,0,3)$	$(1,2,0,0)$	$(0,0,0,1,1)$
20	$(0,5)$	$(4,0,0)$	$(1,1,1,0)$	$(0,0,0,0,2)$
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$

It is convenient to define $\mathcal{P}_n(k)$ as the set of the elements which precede k; formally:

$$\mathcal{P}_n(k) = \{k' \in \mathcal{A}_n \ : \ k' \prec k\} \ .$$

With the latter notation the indexing function is simply defined as $I(k) = \#\mathcal{P}_n(k)$. The following definitions are also useful. Pick a vector $k \in \mathcal{A}_n$, and define the sets $\mathcal{B}_n^{(i)}(k)$, $\mathcal{B}_n^{(ii)}(k)$ and $\mathcal{B}_n^{(iii)}(k)$ as the subsets of $\mathcal{A}_n$ satisfying (i), (ii) and (iii), respectively, in the ordering algorithm above. Formally:

$$\mathcal{B}_n^{(i)}(0) = \mathcal{B}_n^{(ii)}(0) = \mathcal{B}_n^{(iii)}(0) = \mathcal{B}_1^{(ii)}(k) = \mathcal{B}_1^{(iii)}(k) = \emptyset \ ,$$
$$\mathcal{B}_n^{(i)}(k) = \{k' \in \mathcal{A}_n \ : \ |k'| < |k|\} \ ,$$
$$\mathcal{B}_n^{(ii)}(k) = \{k' \in \mathcal{A}_n \ : \ |k'| = |k| \ \wedge \ k_1' > k_1\} \ ,$$
$$\mathcal{B}_n^{(iii)}(k) = \{k' \in \mathcal{A}_n \ : \ |k'| = |k| \ \wedge \ k_1' = k_1 \ \wedge \ t(k') < t(k)\} \ . \qquad (17)$$

The sets $\mathcal{B}_n^{(i)}(k)$, $\mathcal{B}_n^{(ii)}(k)$ and $\mathcal{B}_n^{(iii)}(k)$ are pairwise disjoint, and moreover

$$\mathcal{B}_n^{(i)}(k) \cup \mathcal{B}_n^{(ii)}(k) \cup \mathcal{B}_n^{(iii)}(k) = \mathcal{P}_n(k) \ .$$

This easily follows from the definition.

5.2. Indexing function for polynomials

Let $k \in \mathcal{A}_n$. In view of the definitions and of the properties above the index function, defined as in (14), turns out to be

$$I(0) = 0 \ , \quad I(k) = \#\mathcal{B}_n^{(i)}(k) + \#\mathcal{B}_n^{(ii)}(k) + \#\mathcal{B}_n^{(iii)}(k) \ . \tag{18}$$

Let us introduce the functions

$$J(n,s) = \#\{k \in \mathcal{A}_n \ : \ |k| = s\} \ ,$$

$$N(n,s) = \sum_{j=0}^{s} J(n,j) \qquad \text{for } n \geq 1, \ s \geq 0 \ . \tag{19}$$

These functions will be referred to in the following as *J-table and N-table*. We claim that the indexing function can be recursively computed as

$$I(0) = 0 \ ,$$

$$I(k) = \begin{cases} N(n, |k|-1) & \text{for } k_1 = |k| \ , \\ N(n, |k|-1) + I(t(k)) & \text{for } k_1 < |k| \ . \end{cases} \tag{20}$$

The claim follows from

$$\#\mathcal{B}_n^{(i)}(k) = N(n, |k|-1) \ ; \tag{21}$$

$$\#\mathcal{B}_n^{(ii)}(k) = \begin{cases} 0 & \text{for } k_1 = |k| \ , \\ N(n-1, |k|-k_1-1) & \text{for } k_1 < |k| \ ; \end{cases} \tag{22}$$

$$\#\mathcal{B}_n^{(iii)}(k) = \begin{cases} I(t(k)) & \text{for } k_1 = |k| \ , \\ I(t(k)) - N(n-1, |k|-k_1-1) & \text{for } k_1 < |k| \ . \end{cases} \tag{23}$$

The equality (21) is a straightforward consequence of the definition of the *N*-table. The equality (22) follows from (17). Indeed, for $|k| = k_1$ we have $\mathcal{B}_n^{(ii)}(k) = \emptyset$, and for $|k| > k_1$ we have

$$\mathcal{B}_n^{(ii)}(k) = \bigcup_{k_1 < j \leq |k|} \{k' \in \mathcal{A}_n \ : \ k_1' = j \ \wedge \ |t(k')| = |k| - j\}$$

$$= \bigcup_{0 \leq l < |k|-k_1} \{k' \in \mathcal{A}_n \ : \ k_1' = |k| - l \ \wedge \ |t(k')| = l\} \ ;$$

Coming to (23), first remark that

$$\mathcal{B}_n^{(iii)}(k) = \{k' \in \mathcal{A}_n \ : \ k_1' = k_1 \ \wedge \ |t(k')| = |k| - k_1 \ \wedge \ t(k') \prec t(k)\} \ ,$$

so that

$$\#\mathcal{B}_n^{(iii)}(k) = \#\{\lambda \in \mathcal{A}_{n-1} \ : \ |\lambda| = |k| - k_1 \ \wedge \ \lambda \prec t(k)\} \ .$$

Then, the equality follows by remarking that

$$\mathcal{P}_{n-1}(t(k)) = \{\lambda \in \mathcal{A}_{n-1} \ : \ |\lambda| = |k| - k_1 \ \wedge \ \lambda \prec t(k)\}$$
$$\cup \{\lambda \in \mathcal{A}_{n-1} \ : \ |\lambda| < |k| - k_1\} \ .$$

Adding up all contributions (20) follows.

5.3.   Construction of the tables

In view of (19) and (20) the indexing function is completely determined in explicit form by the J-table. We show now how to compute the J-table recursively. For $n = 1$ we have, trivially, $J(1, s) = 1$ for $s \geq 0$. For $n > 1$ use the elementary property

$$\{k \in \mathcal{A}_n \ : \ |k| = s\} = \bigcup_{0 \leq j \leq s} \{k \in \mathcal{A}_n \ : \ k_1 = s - j \ \wedge \ |t(k)| = j\} \ .$$

Therefore

$$J(1, s) = 1 \ ,$$
$$J(n, s) = \sum_{j=0}^{s} J(n - 1, j) \quad \text{for } n > 1 \ . \tag{24}$$

This also means that, according to (19), we have $N(n, s) = J(n + 1, s)$.

By the way, one will recognize that the J-table is actually the table of binomial coefficients, being $J(n, s) = \binom{n+s-1}{n-1}$.

5.4.   Inversion of the index function

The problem is to find the vector $k \in \mathbb{Z}_+^n$ corresponding to a given index l.

For $n = 1$ we have $I^{-1}(l) = l$, of course. Therefore, let us assume $n > 1$. We shall construct a recursive algorithm which calculates the inverse function by just showing how to determine k_1 and $I(t(k))$.

(i) If $l = 0$, then $k = 0$, and there is nothing else to do.

(ii) If $l > 0$, find an integer s satisfying $N(n, s - 1) \leq l < N(n, s)$. In view of (20) we have $|k| = s$ and $I(t(k)) = l - N(n, s - 1)$. Hence, by the same method, we can determine $|t(k)|$, and so also $k_1 = s - |t(k)|$.

5.5.   An example of implementation

We include here an example of actual implementation of the indexing scheme for polynomials. This is part of a program for the calculation of first integrals that is fully described in (Giorgilli 1979). The complete computer code is also available from the CPC program library.

We should mention that the `FORTRAN` code included here has been written in 1976. Hence it may appear a little strange to programmers who are familiar with the nowadays compilers, since it does not use many features that are available in `FORTRAN 90` or in the current versions of the compiler. It rather uses the standard of `FORTRAN II`, with the only exception of the statement `PARAMETER` that has been introduced later.

The `PARAMETER`s included in the code allow the user to control the allocation of memory, and may be changed in order to adapt the program to different needs. `NPMAX` is the maximum number of degrees of freedom
`NORDMX` is the maximal polynomial degree that will be used
`NBN1` and `NBN2` are calculated from the previous parameters, and are used in order to allocate the correct amount of memory for the table of binomial coefficients. Here are the statements:

```
      PARAMETER (NPMAX=3)
      PARAMETER (NORDMX=40)
      PARAMETER (NBN2=2*NPMAX)
      PARAMETER (NBN1=NORDMX+NBN2)
```

As explained in the previous sections, the indexing function for polynomials uses the table of binomial coefficients. The table is stored in a common block named `BINTAB` so that it is available to all program modules. In the same block there are also some constants that are used by the indexing functions and are defined in the subroutine `BINOM` below. Here is the statement that must be included in every source module that uses these data:

```
      COMMON  /BINTAB/  IBIN(NBN1,NBN2),NPIU1,NMEN1,NFAT,NBIN
```

Subroutine `BINOM` fills the data in the common block `BINTAB`. It must be called at the beginning of the execution, so that the constants become available. Forgetting this call will produce unpredictable results. The calling arguments are the following.
`NLIB` : the number of polynomial variables. In the Hamiltonian case considered in the present notes it must be set as $2n$, where n is the number of degrees of freedom. It must not exceed the value of the parameter `NPMAX`.
`NORD` : the wanted order of calculation of the polynomials, which in our case is the maximal order of the normal form. It must not exceed the value of the parameter `NFAT`.

The subroutine checks the limits on the calling arguments; if the limits are violated then the execution is terminated with an error message. The calculation of the part of the table of binomial coefficients that will be used is based on well known formulæ.

```
      SUBROUTINE BINOM(NLIB,NORD)
C
C     Compute the table of the binomial coefficients.
C
      COMMON  /BINTAB/  IBIN(NBN1,NBN2),NPIU1,NMEN1,NFAT,NBIN
C
      NFAT=NORD+NLIB
      NBIN=NLIB
      IF(NFAT.GT.NBN1.OR.NBIN.GT.NBN2) GO TO 10
```

```
          NPIU1 = NLIB+1
          NMEN1 = NLIB-1
          DO 1 I=1,NFAT
          IBIN(I,1) = I
          DO 1 K=2,NBIN
          IF(I-K) 2,3,4
   2      IBIN(I,K) = 0
          GO TO 1
   3      IBIN(I,K) = 1
          GO TO 1
   4      IBIN(I,K) = IBIN(I-1,K-1)+IBIN(I-1,K)
   1      CONTINUE
          RETURN
  10      WRITE(6,1000) NFAT,NBIN
          STOP
1000      FORMAT(//,5X,15HERROR SUB BINOM,2I10,//)
          END
```

Function `INDICE` implements the calculation of the indexing function for polynomials. The argument `J` is an integer array of dimension `NLIB` which contains the exponents of the monomial. It must contain non negative values with sum not exceeding the value `NORD` initially passed to the subroutine `BINOM`. These limits are not checked in order to avoid wasting time: note that this function may be called several millions of times in a program. The code actually implements the recursive formula (20) using iteration. Recall that recursion was not implemented in `FORTRAN II`.

```
          FUNCTION INDICE(J,NLIB)
   C
   C      Compute the relative address I corresponding to the
   C      exponents J.
   C
          COMMON  /BINTAB/  IBIN(NBN1,NBN2),NPIU1,NMEN1,NFAT,NBIN
          DIMENSION J(NLIB)
   C
          NP=NLIB+1
          INDICE = J(NLIB)
          M = J(NLIB)-1
          DO 1 I=2,NLIB
          IB=NP-I
          M = M + J(IB)
          IB=M+I
          INDICE = INDICE + IBIN(IB,I)
   1      CONTINUE
          RETURN
          END
```

Subroutine `ESPON` is the inverse of the indexing function. Given the index `N` it calculates the array `J` of dimension `NLIB` which contains the exponents. The value of `N` must be positive (not checked) and must not exceed the maximal index

implicitly introduced by the initial choice of `NLIB` and `NORD` passed to `BINOM`. The latter error is actually checked (this does not increase the computation time). The code implements the recursive algorithm described in sect. 5.4, again using iteration.

```
        SUBROUTINE ESPON(N,J,NLIB)
C
C       Compute the exponents J correponding to the
C       index N.
C
        COMMON  /BINTAB/  IBIN(NBN1,NBN2),NPIU1,NMEN1,NFAT,NBIN
        DIMENSION J(NLIB)
C
        NM=NLIB-1
        NP=NLIB+1
        DO 1 K=NLIB,NFAT
        IF (N.LT.IBIN(K,NLIB)) GO TO 2
  1     CONTINUE
        WRITE(6,1000)
        STOP
  2     NN = K-1
        M = N-IBIN(NN,NLIB)
        IF(NLIB-2) 8,6,7
  7     DO 3 I = 2,NM
        L = NP-I
        DO 4 K=L,NFAT
        IF(M.LT.IBIN(K,L)) GO TO 5
  4     CONTINUE
  5     IB=NLIB-L
        J(IB) = NN-K
        NN = K-1
        M = M - IBIN(NN,L)
  3     CONTINUE
  6     J(NM) = NN-M-1
        J(NLIB) = M
        RETURN
  8     J(1)=N
        RETURN
 1000   FORMAT(//,5X,15HERROR SUB ESPON,//)
        END
```

The code described here is the skeleton of a program performing algebraic manipulation on polynomial. Such a program must include a call to `BINOM` in order to initialize the table of binomial coefficients.

In order to store the coefficient of a monomial with exponents `J` (an integer array with dimension `NLIB` the user must include a statement like

```
        K = INDICE(J,NLIB)
```

and then say, e.g., `F(K)=`... which stores the coefficient at the address K of the array F.

Suppose instead that we must perform an operation on all coefficients of degree IORD of a given function F. We need to perform a loop on all the corresponding indices and retrieve the corresponding exponents. Here is a sketch of the code.

```
C       Compute the minimum and maximum index NMIN and NMAX
C       of the coefficients of order IORD.
C

        IB=IORD+NMEN1
        NMIN = IBIN(IB,NLIB)
        IB=IORD+NLIB
        NMAX = IBIN(IB,NLIB) - 1
C
C       Loop on all coefficients
C

        DO 1 N = NMIN,NMAX
        CALL ESPON(N,J,NLIB)
        ... more code to operate on the coefficient F(N) ...
    1   CONTINUE
```

Let us add a few words of explanation. According to (20), the index of the first coefficient of degree s in n variables is $I(s, 0, \ldots, 0) = N(n, s-1)$, and we also have $N(n, s-1) = \binom{n+s-1}{n}$ as explained at the end of sect. 5.3. This explains how the limits NMIN and NMAX are calculated as $N(n, s-1)$ and $N(n, s+1) - 1$, respectively. The rest of the code is the loop that retrieves the exponents corresponding to the coefficient of index N.

6. Trigonometric polynomials

Let us now consider the more general case $\mathcal{A}_n = \mathbb{Z}^n$. The index n in $\mathcal{A}_n$ denotes again the dimension of the space. The name used in the title of the section is justified because this case occurs precisely in the representation of trigonometric polynomials, as explained in sect. 3.2.

We shall now denote by $|k| = |k_1| + \ldots + |k_n|$ the length (or norm) of the vector $k \in \mathbb{Z}^n$. The tail $t(k)$ of a vector k will be defined again as $t(k) = (k_2, \ldots, k_n)$.

6.1. Ordering relation

Pick a fixed n, and consider the finite family of sets $\mathcal{A}_1 = \mathbb{Z}, \ldots, \mathcal{A}_n = \mathbb{Z}^n$.

Let $k, k' \in \mathcal{A}_m$, with any $1 \leq m \leq n$. We shall say $k' \prec k$ in case one of the following conditions is true:

(i) $m \geq 1 \ \wedge \ |k'| < |k|$;

(ii) $m > 1 \ \wedge \ |k'| = |k| \ \wedge \ |k_1'| > |k_1|$;

(iii) $m \geq 1 \ \wedge \ |k'| = |k| \ \wedge \ |k_1'| = |k_1| \ \wedge \ k_1' > k_1$;

(iv) $m > 1 \;\wedge\; |k'| = |k| \;\wedge\; k'_1 = k_1 \;\wedge\; t(k') \prec t(k)$.

Table 3.   Ordering of integer vectors in $\mathbb{Z}^m$ for $m = 2, 3, 4$.

$I(k)$	$m = 2$	$m = 3$	$m = 4$
0	(0, 0)	(0, 0, 0)	(0, 0, 0, 0)
1	(1, 0)	(1, 0, 0)	(1, 0, 0, 0)
2	(−1, 0)	(−1, 0, 0)	(−1, 0, 0, 0)
3	(0, 1)	(0, 1, 0)	(0, 1, 0, 0)
4	(0, −1)	(0, −1, 0)	(0, −1, 0, 0)
5	(2, 0)	(0, 0, 1)	(0, 0, 1, 0)
6	(−2, 0)	(0, 0, −1)	(0, 0, −1, 0)
7	(1, 1)	(2, 0, 0)	(0, 0, 0, 1)
8	(1, −1)	(−2, 0, 0)	(0, 0, 0, −1)
9	(−1, 1)	(1, 1, 0)	(2, 0, 0, 0)
10	(−1, −1)	(1, −1, 0)	(−2, 0, 0, 0)
11	(0, 2)	(1, 0, 1)	(1, 1, 0, 0)
12	(0, −2)	(1, 0, −1)	(1, −1, 0, 0)
13	(3, 0)	(−1, 1, 0)	(1, 0, 1, 0)
14	(−3, 0)	(−1, −1, 0)	(1, 0, −1, 0)
15	(2, 1)	(−1, 0, 1)	(1, 0, 0, 1)
16	(2, −1)	(−1, 0, −1)	(1, 0, 0, −1)
17	(−2, 1)	(0, 2, 0)	(−1, 1, 0, 0)
18	(−2, −1)	(0, −2, 0)	(−1, −1, 0, 0)
19	(1, 2)	(0, 1, 1)	(−1, 0, 1, 0)
20	(1, −2)	(0, 1, −1)	(−1, 0, −1, 0)
21	(−1, 2)	(0, −1, 1)	(−1, 0, 0, 1)
22	(−1, −2)	(0, −1, −1)	(−1, 0, 0, −1)
23	(0, 3)	(0, 0, 2)	(0, 2, 0, 0)
24	(0, −3)	(0, 0, −2)	(0, −2, 0, 0)
. . .	. . .	. . .	. . .

In table 3 the order resulting from this definition is illustrated for the cases $m = 2, 3, 4$.

If $m = 1$ this ordering coincides with the ordering in $\mathbb{Z}$ introduced in sect 4. For $m > 1$, if (i), (ii) and (iii) do not apply, then (iv) means that one must decrease the dimension n by replacing k with its tail $t(k)$, and retry the comparison. Eventually, one ends up with $m = 1$, falling back to the one dimensional case to which only (i) and (iii) apply.

The ordering in this section has been defined for the case $\mathcal{A}_n = \mathbb{Z}^n$. However, it will be useful to consider particular subsets of $\mathbb{Z}^n$. The natural choice will be to use again the ordering relation defined here. For example, the case of integer vectors with non negative components discussed in sect. 5.1 can be considered as a particular case: the restriction of the ordering relation to that case gives exactly the order introduced in sect. 5.1. Just remark that the condition (iii) above becomes meaningless in that case, so that it can be removed.

The set $\mathcal{P}_n(k)$ of the elements preceding $k \in \mathcal{A}^n$ in the order above is defined as in sect. 5.1. Following the line of the discussion in that section it is also convenient to give some more definitions. Pick a vector $k \in \mathcal{A}_n$, and define the sets $\mathcal{B}_n^{(i)}(k)$, $\mathcal{B}_n^{(ii)}(k)$, $\mathcal{B}_n^{(iii)}(k)$ and $\mathcal{B}_n^{(iv)}(k)$ as the subsets of $\mathcal{A}_n$ satisfying (i), (ii), (iii) and (iv), respectively, in the ordering algorithm above. Formally,

$$\mathcal{B}_n^{(i)}(0) = \mathcal{B}_n^{(ii)}(0) = \mathcal{B}_n^{(iii)}(0) = \mathcal{B}_n^{(iv)}(0) = \mathcal{B}_1^{(ii)}(k) = \mathcal{B}_1^{(iv)}(k) = \emptyset \ ,$$
$$\mathcal{B}_n^{(i)}(k) = \{k' \in \mathcal{A}_n \ : \ |k'| < |k|\} \ ,$$
$$\mathcal{B}_n^{(ii)}(k) = \{k' \in \mathcal{A}_n \ : \ |k'| = |k| \ \wedge \ |k_1'| > |k_1|\} \ ,$$
$$\mathcal{B}_n^{(iii)}(k) = \{k' \in \mathcal{A}_n \ : \ |k'| = |k| \ \wedge \ |k_1'| = |k_1| \ \wedge \ k_1' > k_1\} \ ,$$
$$\mathcal{B}_n^{(iv)}(k) = \{k' \in \mathcal{A}_n \ : \ |k'| = |k| \ \wedge \ k_1' = k_1 \ \wedge \ t(k') < t(k)\} \ . \qquad (25)$$

The sets $\mathcal{B}_n^{(i)}(k)$, $\mathcal{B}_n^{(ii)}(k)$, $\mathcal{B}_n^{(iii)}(k)$ and $\mathcal{B}_n^{(iv)}(k)$ are pairwise disjoint, and moreover

$$\mathcal{B}_n^{(i)}(k) \cup \mathcal{B}_n^{(ii)}(k) \cup \mathcal{B}_n^{(iii)}(k) \cup \mathcal{B}_n^{(iv)}(k) = \mathcal{P}(k) \ .$$

This easily follows from the definition.

6.2.   Indexing function for trigonometric polynomials

Let $k \in \mathcal{A}_n$. In view of the definitions and of the properties above the index function, defined as in (14), turns out to be

$$I(0) = 0 \ , \quad I(k) = \#\mathcal{B}_n^{(i)}(k) + \#\mathcal{B}_n^{(ii)}(k) + \#\mathcal{B}_n^{(iii)}(k) + \#\mathcal{B}_n^{(iv)}(k) \ . \qquad (26)$$

Let us introduce the J-table and the N-table as

$$J(n,s) = \#\{k \in \mathcal{A}_n \ : \ |k| = s\} \ ,$$
$$N(n,s) = \sum_{j=0}^{s} J(n,j) \qquad\qquad \text{for } n \geq 1 \, , \ s \geq 0 \ . \qquad (27)$$

We claim that the index function can be recursively computed as

$I(0) = 0$,

$$I(k) = \begin{cases} N(n, |k|-1) & \text{for } |k_1| = |k| \\ & \text{and } k_1 \geq 0 \text{ ,} \\ N(n, |k|-1) + 1 & \text{for } |k_1| = |k| \\ & \text{and } k_1 < 0 \text{ ,} \\ N(n, |k|-1) + N(n-1, |k|-|k_1|-1) + I(t(k)) & \text{for } |k_1| < |k|, \\ & \text{and } k_1 \geq 0 \text{ ,} \\ N(n, |k|-1) + N(n-1, |k|-|k_1|) + I(t(k)) & \text{for } |k_1| < |k| \\ & \text{and } k_1 < 0 \text{ ,} \end{cases}$$

(28)

This formula follows from

$$\#\mathcal{B}_n^{(i)}(k) = N(n, |k|-1) \text{ ;}$$

(29)

$$\#\mathcal{B}_n^{(ii)}(k) = \begin{cases} 0 & \text{for } |k_1| = |k| \text{ ,} \\ 2N(n-1, |k|-|k_1|-1) & \text{for } |k_1| < |k| \text{ ;} \end{cases}$$

(30)

$$\#\mathcal{B}_n^{(iii)}(k) = \begin{cases} 0 & \text{for } |k_1| \leq |k| \ \wedge \ k_1 \geq 0 \text{ ,} \\ J(n-1, |k|-|k_1|) & \text{for } |k_1| \leq |k| \ \wedge \ k_1 < 0 \text{ ;} \end{cases}$$

(31)

$$\#\mathcal{B}_n^{(iv)}(k) = \begin{cases} I(t(k)) & \text{for } |k_1| = |k| \text{ ,} \\ I(t(k)) - N(n-1, |k|-|k_1|-1) & \text{for } |k_1| < |k| \text{ .} \end{cases}$$

(32)

The equality (29) is a straightforward consequence of the definition (25). The equality (30) follows by remarking that for $|k_1| = |k|$ we have $\mathcal{B}_n^{(ii)}(k) = \emptyset$, and for $|k_1| < |k|$ we have

$$\mathcal{B}_n^{(ii)}(k) = B_n^+(k) \cup B_n^-(k) \text{ ,} \quad B_n^+(k) \cap B_n^-(k) = \emptyset \text{ ,}$$

with

$$B_n^+(k) = \bigcup_{0 \leq l < |k|-|k_1|} \{k' \in \mathcal{A}_n \ : \ k_1' = |k| - l \ \wedge \ |t(k)| = l\} \text{ ,}$$

$$B_n^-(k) = \bigcup_{0 \leq l < |k|-|k_1|} \{k' \in \mathcal{A}_n \ : \ k_1' = l - |k| \ \wedge \ |t(k)| = l\} \text{ ;}$$

use also $\#B_n^+(k) = \#B_n^-(k)$. The equality (31) follows from

$$\mathcal{B}_n^{(iii)}(k) = \begin{cases} \emptyset & \text{for } |k_1| = |k| \\ & \text{and } k_1 \geq 0 \text{ ,} \\ \{k' \in \mathcal{A}_n \ : \ k_1' = |k_1| \ \wedge \ |t(k')| = |k| - |k_1|\} & \text{for } |k_1| = |k| \\ & \text{and } k_1 < 0 \text{ .} \end{cases}$$

Coming to (32), remark that

$$\mathcal{B}_n^{(iv)}(k) = \{k' \in \mathcal{A}_n \ : \ |t(k')| = |k| - |k_1| \ \wedge \ t(k') \prec t(k)\} \ .$$

Proceeding as in the polynomial case we find again

$$\#\mathcal{B}_n^{(iv)}(k) = \#\{\lambda \in \mathcal{A}_{n-1} \ : \ |\lambda| = |k| - |k_1| \ \wedge \ \lambda \prec t(k)\} \ ,$$

and (32) follows by remarking that

$$\mathcal{P}_{n-1}(t(k)) = \{\lambda \in \mathcal{A}_{n-1} \ : \ |\lambda| = |k| - |k_1| \ \wedge \ \lambda \prec t(k)\}$$
$$\cup \{\lambda \in \mathcal{A}_{n-1} \ : \ |\lambda| < |k| - |k_1|\} \ .$$

Adding up all contributions (28) follows.

6.3.   Construction of the tables

We show now how to construct recursively the J-table, so that the N-table can be constructed, too. For $n = 1$ we have, trivially, $J(1,0) = 1$ and $J(1,s) = 2$ for $s > 0$. For $n > 1$ use the elementary property

$$\{k \in \mathcal{A}_n \ : \ |k| = s\} = \bigcup_{-s \le j \le s} \{k \in \mathcal{A}_n \ : \ k_1 = j \ \wedge \ |t(k)| = s - |j|\} \ .$$

Therefore

$$\begin{aligned}
J(1,0) &= 1 \ , \\
J(1,s) &= 2 \ , \\
J(n,s) &= \sum_{j=-s}^{s} J(n-1, s - |j|) \quad \text{for } n > 1 \ .
\end{aligned} \qquad (33)$$

This completely determines the J-table.

6.4.   Inversion of the index function

The problem is to find the vector k of given dimension n corresponding to the given index l. For $n = 1$ the function $I(k)$ and its inverse $I^{-1}(l)$ are given by (15) and (16). Therefore in the rest of this section we shall assume $n > 1$. We shall give a recursive algorithm, showing how to determine k_1 and $I(t(k))$.

(i) If $l = 0$ then $k = 0$, and there is nothing else to do.

(ii) Assuming that $l > 0$, determine s such that

$$N(n, s-1) \le l < N(n, s) \ .$$

From this we know that $|k| = s$.

(iii) Define $l_1 = l - N(n, s - 1)$, so that $I\big(t(k)\big) \leq l_1$ by (28). If $l_1 = 0$ set $s_1 = 0$; else, determine s' such that

$$N(n - 1, s' - 1) \leq l_1 < N(n - 1, s') \ ;$$

and let $s_1 = \min(s', s)$. In view of $I\big(t(k)\big) \leq l_1$ we know that $|t(k)| \leq s_1$. Remark also that $s_1 = 0$ if and only if $l_1 = 0$. For, if $s_1 \geq 1$ then we have $l_1 \geq N(n - 1, 0) = 1$.

(iv) If $l_1 = 0$, then by the first of (28) we conclude

$$k_1 = |k| = s \ , \quad t(k) = 0 \ ,$$

and there is nothing else to do.

(v) If $l_1 = 1$, then by the second of (28) we conclude

$$k_1 = -|k| = -s \ , \quad t(k) = 0 \ ,$$

and there is nothing else to do.

(vi) If $l_1 > 1$ and $s_1 > 0$, we first look if we can set $0 \leq k_1 < |k|$. In view of the third of (28) we should have

$$|k| - k_1 = s_1 \ , \quad \big|t(k)\big| = s_1 \ , \quad I\big(t(k)\big) = l_1 - N(n - 1, s_1 - 1) \ .$$

This can be consistently made provided the conditions

$$s_1 > 0 \quad \text{and} \quad I\big(t(k)\big) \geq N(n - 1, s_1 - 1)$$

are fulfilled. The condition $s > 0$ is already satisfied. By (28), the second condition is fulfilled provided $l_1 \geq 2N(n-1, s_1-1)$. This has to be checked.

> **(vi.a)** If the second condition is true, then set $k_1 = |k| - s_1$, and recall that $\big|t(k)\big| = s_1$. Hence, we can replace n, l, and s by $n - 1$, $l_1 - N(n - 1, s_1 - 1)$ and s_1, respectively, and proceed by recursion restarting again from the point (iii).
>
> **(vi.b)** If the second condition is false, then we proceed with the next point.

(vii) Recall that $l_1 > 1$, and remark that we have also $s_1 > 1$. Indeed, we already know $s_1 > 0$, so we have to exclude the case $s_1 = 1$. Let, by contradiction, $s_1 = 1$. Then we have $l_1 \geq 2 = 2N(n - 1, s_1 - 1)$, which is the case already excluded by (vi). We conclude $s_1 > 1$. We look now for the possibility of setting $|k_1| < |k|$ and $k_1 < 0$. In view of the fourth of (28) we should have

$$|k| + k_1 = s_1 - 1 \ , \quad \big|t(k)\big| = s_1 - 1 \ , \quad I\big(t(k)\big) = l_1 - N(n - 1, s_1 - 1) \ .$$

This can be consistently made provided the conditions

$$s_1 > 1 \quad \text{and} \quad I\big(t(k)\big) \geq N(n - 1, s_1 - 2)$$

are fulfilled. The condition $s_1 > 1$ is already satisfied. As to the second condition, by (28) it is fulfilled provided $l_1 > N(n - 1, s_1 - 1) + N(n - 1, s_1 - 2)$. This has to be checked.

(vii.a) If the second condition is true, then set $k_1 = -|k| + s_1 - 1$, and recall that $\left|t(k)\right| = s_1 - 1$. Hence, we can replace n, l, and s by $n-1$, $l_1 - N(n-1, s_1 - 2)$ and $s_1 - 1$, respectively, and proceed by recursion restarting again from the point (iii).

(vii.b) If the second condition is false we must decrease s_1 by one and start again with the point (vi); remark that $s_1 > 1$ implies $s_1 - 1 > 0$, which is the first of the two conditions to be satisfied at the point (vi), hence the recursion is correct.

Since $l_1 > 1$ we have $l_1 > 2N(n-1, 0)$, so that the conditions of point (vi) are satisfied for $s = 1$. Hence the algorithm above does not fall into an infinite loop between points (vi) and (vii). On the other hand, for $n = 1$ either (iii) or (iv) applies, so that the algorithm stops at that point.

7. Storing the coefficients for sparse functions

The method of storing the coefficient using the index, as illustrated in sect. 3, is the most direct one, but reveals to be ineffective when most of the coefficients of a function are zero (sparse function). For, allocating memory space for all coefficients results in a wasting of memory.

A method that we often use is to store the coefficients using a tree structure based on the index. However we should warn the reader that the method described here has the advantage of being easily programmed, but does not pretend to be the most effective one. Efficient programming of tree structure is described, e.g., in the monumental books *The art of computing programming*, by D.E. Knuth (1968).

7.1. The tree structure

The first information we need is how many bits are needed in order to represent the maximum index for a function. We shall refer to this number as the *length of the index*. In the scheme that we are presenting here this is actually the length of the path from the root of the tree to its leave, where the coefficient is found.

In fig. 1 we illustrate the scheme assuming that 4 bits are enough, i.e., there are at most 16 coefficients indexed from 0 to 15. The case is elementary, of course, but the method is the general one, and is extended to, e.g., several millions of coefficients (with a length a little more than 20) in a straightforward manner. The bits are labeled by their position, starting from the less significant one (choosing the most significant one as the first bit is not forbidden, of course, and sometimes may be convenient). The label of the bit corresponds to a level in the tree structure, level 0 being the root and level 3 being the last one, in our case. At level zero we find a cell containing two pointers, corresponding to the digit 0 and 1, respectively. To each digit we associate a cell of level 1, which contains a pair of pointers, and so on until we reach the last level (3 in our case). Every number that may be represented with 4 bits generates a unique path along the tree, and the last cell contains pointers to the coefficient. The example in the figure represents the path associated with the binary index 1010, namely 10 in decimal notation.
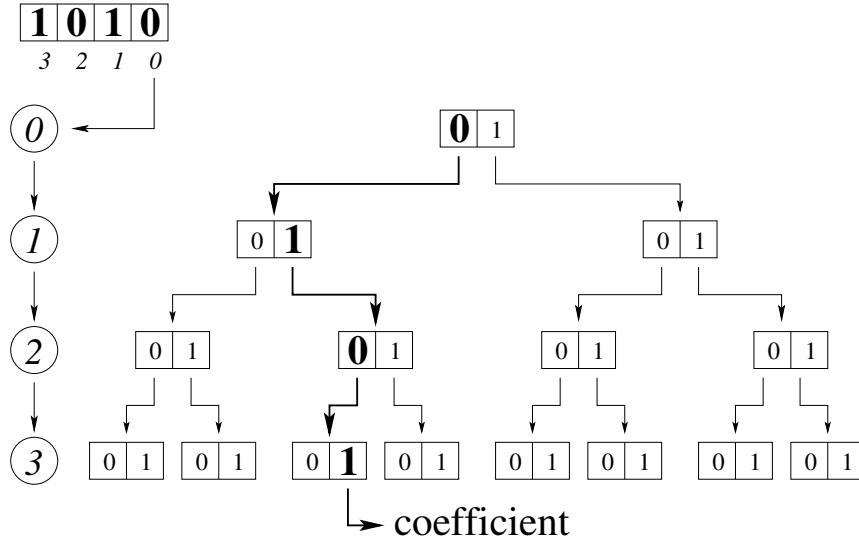
Figure 1.   Illustrating the tree structure for a 4-bit long index (see text).

Let us also illustrate how this structure may be represented in memory, trying to avoid wasting of space. We use two separate arrays, the first one for pointers and the second one for the coefficients, as illustrated in fig. 2. The cells containing pairs of pointers are allocated in the first array, the root of the tree having label zero. The label of a cell is always even: the first element corresponds to the zero bit, the next one (with odd label) to the bit one.

The arrays are initially allocated with appropriate size, and are cleared. A good method is to fill the array of pointers with -1 (denoting an unused pointer) and the coefficients table with zeros. We also keep track of the first unused cell in the array, which initially is set to 2 because the root cell is considered to be in use, and of the first free coefficient, which initially is 0.

We shall use the following notations: $\texttt{cell}(2j)$ is the cell with even label $2j$ in the array; $\texttt{cell}(2j, 0)$ and $\texttt{cell}(2j, 1)$ are the pointers corresponding to a bit 0 or 1 which are stored at locations $2j$ and $2j + 1$, respectively, in the array of pointers; $\texttt{coef}(j)$ is the j-th element of the array of coefficients; $\texttt{cc}$ is the current cell and $\texttt{cb}$ is the current bit (see below for the meaning); $\texttt{fp}$ is the label of the first free (unused) cell of pointers; $\texttt{fc}$ is the label of the first free coefficient; ℓ is the length of the index.

7.2.   Storing the first coefficient

Let us describe how the first coefficient is stored. Suppose we want to store the value x as the coefficient corresponding to a given index. Here is the scheme.

(i) *Initialization:* set $\texttt{cc} = 0$ and $\texttt{cb} = 0$. The values of $\texttt{fp} = 2$ and $\texttt{fc} = 0$ have already been set when during the array allocation.

(ii) *Creating a path:* repeat the following steps until $\texttt{cb}$ equals $\ell - 1$:

    (ii.a) if the bit at position $\texttt{cb}$ in the index is 0, then redefine $\texttt{cell}(\texttt{cc}, 0) = \texttt{fp}$; else redefine $\texttt{cell}(\texttt{cc}, 1) = \texttt{fp}$;
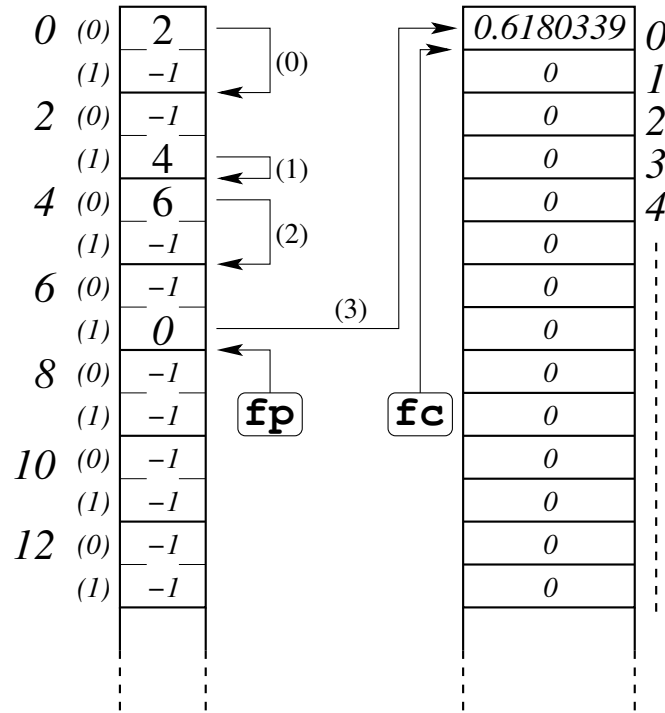
Figure 2.   Illustrating how the tree structure is stored in memory (see text).

(ii.b) set $cc = fp$ and increment fp by 2 (point to the next free cell);

(ii.c) increment cb by 1 (next bit).

(iii) *Store the coefficient:*

(iii.a) if the bit at position cb in the index is 0, then redefine $cell(cc, 0) = fc$; else redefine $cell(cc, 1) = fc$;

(iii.b) set $coef(fc) = x$;

(iii.c) increment fc by 1 (point to the next free coefficient).

Programming this algorithm in a language such as C or FORTRAN requires some 10 to 20 statements.

Let us see in detail what happens if we want to store the coefficient 0.6180339 with index 1010 and $\ell = 4$, as illustrated in fig. 2. Here is the sequence of operations actually made

step (i):	$cc = 0$,	$cb = 0$,	$fp = 2$,	$fc = 0$;	
step (ii):	$cell(0, 0) = 2$,	$cc = 2$,	$fp = 4$,	$cb = 1$,	then ,
	$cell(2, 1) = 4$,	$cc = 4$,	$fp = 6$,	$cb = 2$,	then ,
	$cell(4, 0) = 6$,	$cc = 6$,	$fp = 8$,	$cb = 3$,	end of loop ;
step (iii):	$cell(6, 1) = 0$,	$coef(0) = 0.6180339$,	$fc = 1$,	end of game .	

After this, the contents of the arrays are as represented in fig. 2.

7.3. Retrieving a coefficient

The second main operation is to retrieve a coefficient, which possibly has never been stored. In the latter case, we assume that the wanted coefficient is zero. Here is a scheme.

(i) *Initialization:* set $cc = 0$ and $cb = 0$.

(ii) *Follow a path:* repeat the following steps until cb equals ℓ:

> **(ii.a)** save the current value of cc;
>
> **(ii.b)** if the bit at position cb in the index is 0, then redefine cc as $cell(cc, 0)$; else redefine cc as $cell(cc, 1)$;
>
> **(ii.c)** if $cc = -1$ then the coefficient is undefined. Return 0 as the value of the coefficient;
>
> **(ii.d)** increment cb by 1 (next bit).

(iii) *Coefficient found:* return the coefficient $coef(cc)$.

Let us give a couple of examples in order to better illustrate the algorithm. Suppose that we are looking for the coefficient corresponding to the binary index 1010. By following the algorithm step by step, and recalling that in our example the length of the index is 4, the reader should be able to check that the sequence of operations is the following:

$$
\begin{array}{llll}
\text{step (i):} & cc = 0\,, & cb = 0\,; & \\
\text{step (ii):} & cc = 2\,, & cb = 1\,, & \text{then}\,, \\
 & cc = 4\,, & cb = 2\,, & \text{then}\,, \\
 & cc = 6\,, & cb = 3\,, & \text{then}\,, \\
 & cc = 0 & cb = 4\,, & \text{end of path}\,; \\
\text{step (iii):} & & & \text{return } 0.6180339\,.
\end{array}
$$

The returned value is that of $coef(0)$, stored in the location 0 of the coefficients array.

Suppose now that we are looking for the coefficient corresponding to the binary index 1110. Here is the actual sequence of operations:

$$
\begin{array}{llll}
\text{step (i):} & cc = 0\,, & cb = 0\,; & \\
\text{step (ii):} & cc = 2\,, & cb = 1\,, & \text{then}\,, \\
 & cc = 4\,, & cb = 2\,, & \text{then}\,, \\
 & cc = -1\,, & cb = 2\,, & \text{return zero}\,.
\end{array}
$$

Here the algorithm stops because a coefficient has not been found.

7.4. Other operations

Having implemented the two operations above, the reader should be able to implement also the following operations:

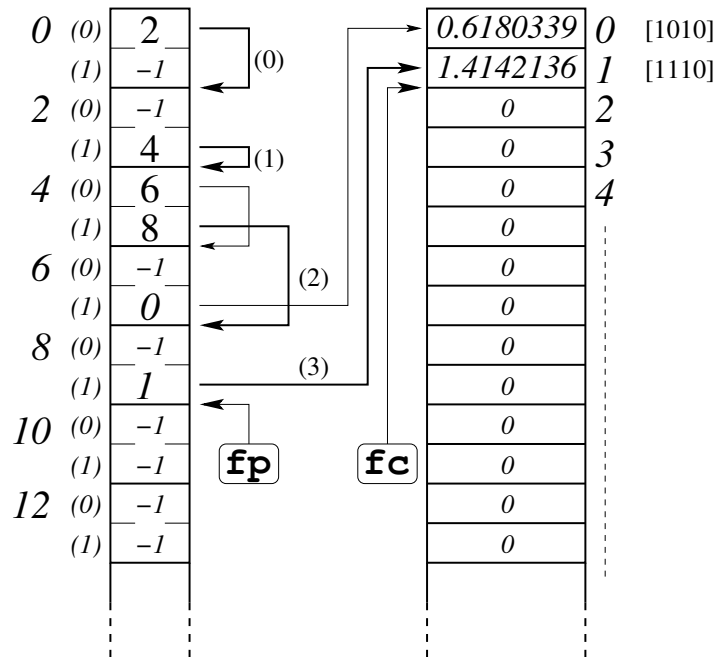**(i)** storing a new coefficient corresponding to a given index;

Figure 3.   Inserting a new coefficient in a tree structure (see text).

(ii) adding something to a given coefficient;

(iii) multiplying a given coefficient by a number.

These are the basic operations that we need in order to perform an elementary computer algebra. Let us add a few hints.

Storing a new coefficient requires perhaps some moment of thinking. Using the index, one should follow the corresponding path in the tree (as in the operation of retrieving) until either happens: the coefficient is found, or the search fails at some point. If the coefficient is found, then it can be overwritten if the new value has to replace the old one. On failure, the path must be completed by appropriately defining the pointers (as in the case of the first coefficient), and then the coefficient can be stored in the appropriate location. As an exercise, suppose that we want to store the coefficient 1.4142136 corresponding to the binary index 1110. After completing the operation the memory should look as in fig. 3.

Adding something to a given coefficient is not very different from the previous operation. Just follow the path. If the coefficient is found, then add the wanted value to it. On failure, just change the "add" operation to a "store" one, and proceed as in the case (i).

Multiplying a coefficient by a constant is even easier. If the coefficient is found, then do the multiplication. On failure, just do nothing.

Further operations can be imagined, but we think that we have described the basic ones. There are just a couple of remarks.

The method illustrated here uses an amount of memory that clearly depends on the number of non zero coefficients of a function. However, this amount is

typically not known in advance. Thus, enough memory should be allocated at the beginning in order to assure that there is enough room. When a function is filled, and we know that it will not be changed, the excess of memory can be freed and reused for other purposes. Every operating system and language provides functions that allow the programmer to allocate memory blocks and resize them on need.

A second remark is that other storing methods can be imagined. E.g., once a function is entirely defined it may be more convenient to represent it as a sequential list of pairs (index, coefficient). This is definitely a very compact representation for a sparse function (although not the best for a crowded one).

8.   Applications

We report here some examples of application of algebraic manipulation that have been obtained by implementing the formal algorithm of sect. 2. We consider three cases, namely the model of Hénon and Heiles, the Lagrangian triangular equilibria for the Sun-Jupiter system and the planetary problem including Sun, Jupiter, Saturn and Uranus (SJSU).

8.1.   The model of Hénon and Heiles

A wide class of canonical system with Hamiltonian of the form

$$H(x, y) = \frac{\omega_1}{2}(y_1^2 + x_1^2) + \frac{\omega_2}{2}(y_2^2 + x_2^2) + x_1^2 x_2 \tag{34}$$

has been studied by Contopoulos, starting at the end of the fifties, for different values of the frequencies. This approximates the motion of a star in a galaxy, at different distances from the center. A wide discussion on the use of these models in galactic dynamics and on the construction of the so called "third integral" can be found in the book of Contopoulos (2002). The third integral is constructed as a power series $\Phi = \Phi_2 + \Phi_3 + \dots$ where Φ_s is a homogeneous polynomial of degree s which is the solution of the equation $\{H, \Phi\} = 0$, where $\{\cdot, \cdot\}$ is the Poisson bracket (see, e.g., Whittaker 1916 or Contopoulos 1960). A different method is based on the construction of the Birkhoff normal form (Birkhoff 1927).

A particular case with two equal frequencies and Hamiltonian

$$H(x, y) = \frac{1}{2}(y_1^2 + x_1^2) + \frac{1}{2}(y_2^2 + x_2^2) + x_1^2 x_2 - \frac{1}{3}x_2^3 \tag{35}$$

has been studied by Hénon and Heiles in 1964. This work has become famous since for the first time the existence of a chaotic behavior in a very simple system has been stressed, showing some figures. It should be remarked that the existence of chaos had been discovered by Poincaré in his memory on the problem of three bodies (Poincare 1889), but it had been essentially forgotten.

A program for the construction of the third integral has been implemented by Contopoulos since 1960. He made several comparisons between the level lines of the integral so found on the surface of constant energy and the figures given by the Poincaré sections of the orbits. A similar calculation for the case of Hénon and Heiles has been made by Gustavson (1966), who used the normal form
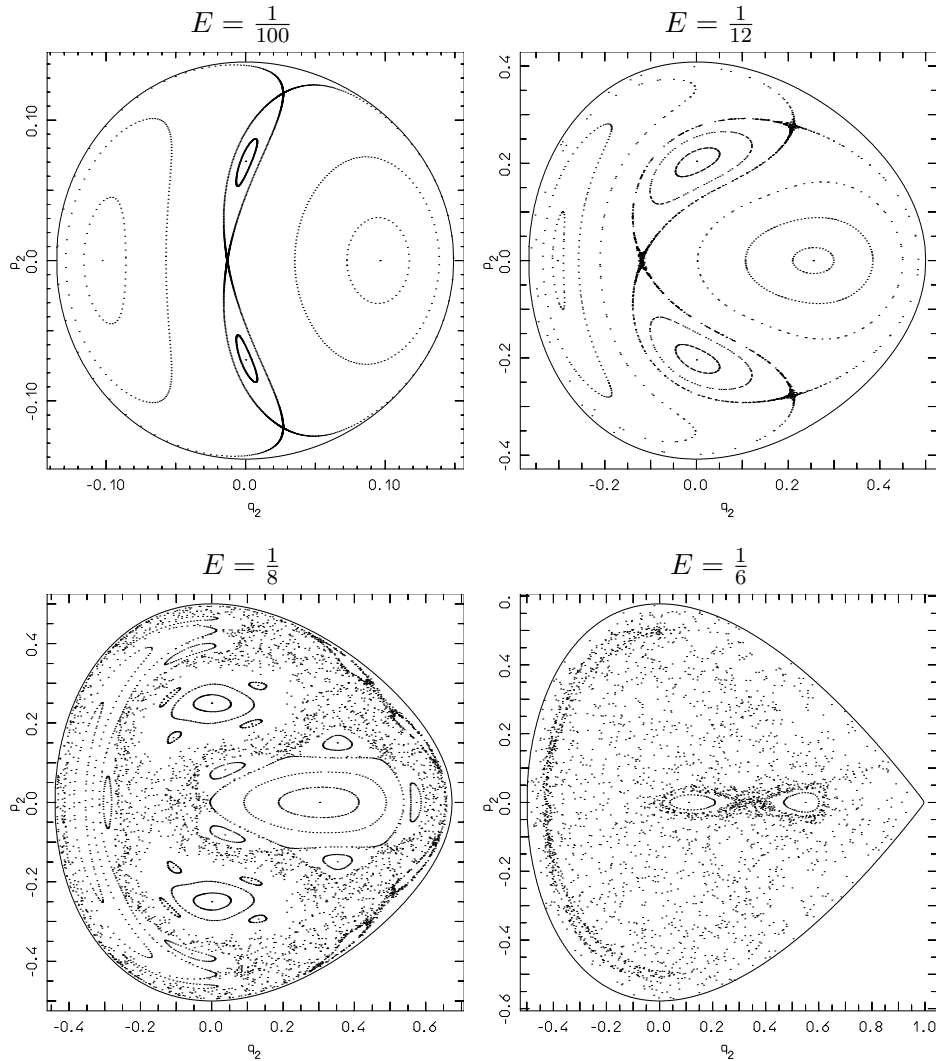
Figure 4.   Poincaré sections for the Hénon and Heiles model. The energies
are as in the original paper.

method. The third integral was expanded up to order 8, which may seem quite
low today, but it was really difficult to do better with the computers available at
that time. Here we reproduce the figures of Gustavson extending the calculation
up to order 58, which is now easily reached even on a PC.

In fig. 4 we show the Poincaré sections for the values of energy used by
Hénon and Heiles in their paper. As stressed by the authors, an essentially
ordered motion is found for $E < \frac{1}{12}$, while the chaotic orbits become predominant
at higher energies.

The comparison with the level lines of the third integral at energy $E = \frac{1}{100}$
is reported in fig. 5. The correspondence with the Poincaré sections is evident
even at order 8, as calculated also by Gustavson. We do not produce the figures
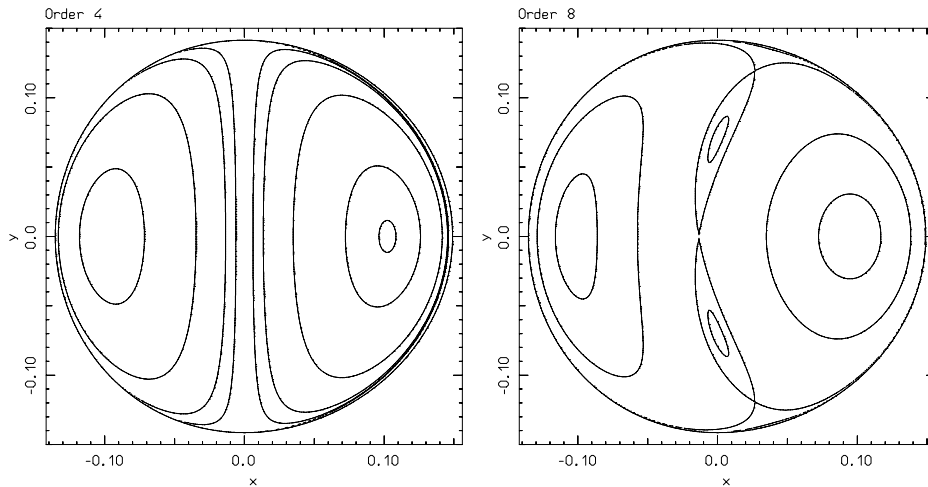
Figure 5. Level lines of the first integral truncated at orders 4 and 8, for energy $E = \frac{1}{100}$. The figure for truncation orders up to 58 are actually the same as for order 8.

for higher orders because they are actually identical with the one for order 8. This may raise the hope that the series for the first integral is a convergent one.

Actually, a theorem of Siegel states that for the Birkhoff normal form divergence is a typical case (Siegel 1941). A detailed numerical study has been made in (Contopolus 2003 and 2004), showing the mechanism of divergence. Moreover, it was understood by Poincaré that perturbations series typically have an asymptotic character (see Poincare 1892, Vol. II). Estimates of this type have been given, e.g., in (Giorgilli 1988 and 1989).

For energy $E = \frac{1}{12}$ (fig. 6) the asymptotic character of the series starts to appear. Indeed already at order 8 we have a good correspondence between the level lines and the Poincaré section, as was shown also Gustavson's paper. If we increase the approximation we see that the correspondence remains good up to order 32, but then the divergence of the series shows up, since at order 43 an unwanted "island" appears on the right side of the figure which has no correspondent in the actual orbits, and at order 58 a bizarre behavior shows up.

The phenomenon is much more evident for energy $E = \frac{1}{8}$ (fig. 7). Here some rough correspondence is found around order 9, but then the bizarre behavior of the previous case definitely appears already at order 27.

The non convergence of the normal form is illustrated in fig. 8. Writing the homogeneous terms of degree s of the third integral as $\Phi_s = \sum_{j,k} \varphi_{j,k} x^j y^k$, we may introduce the norm

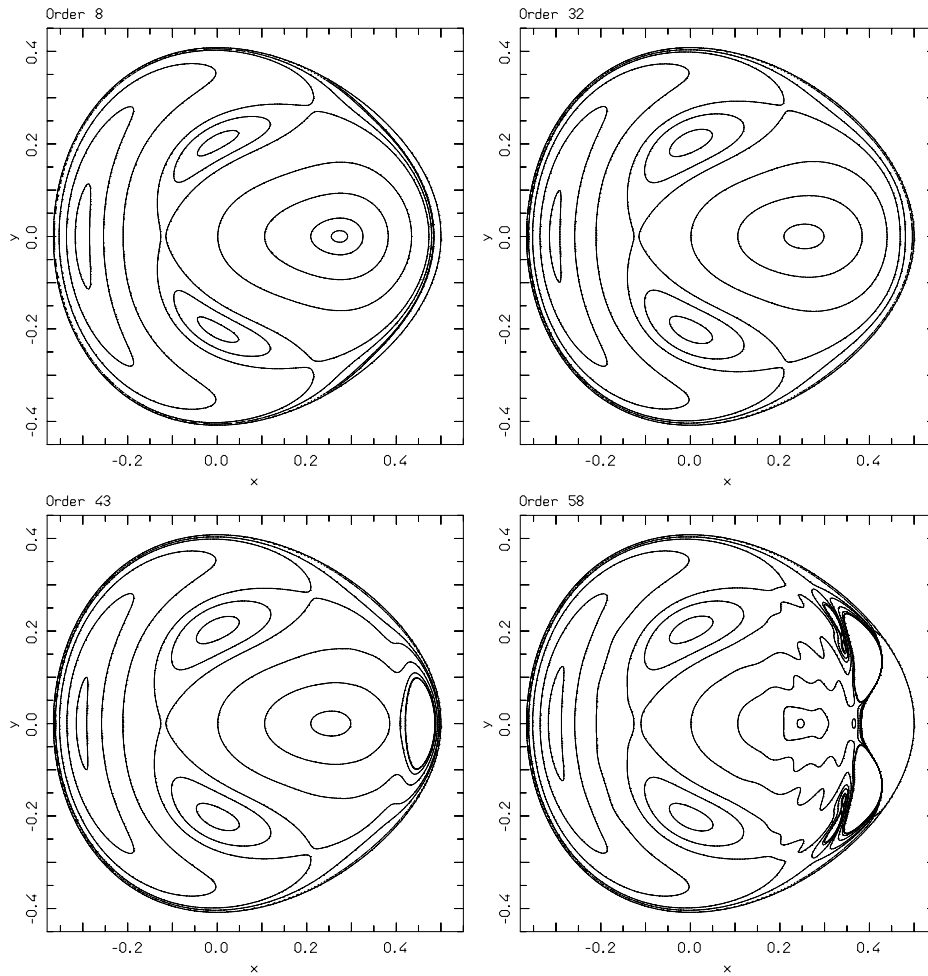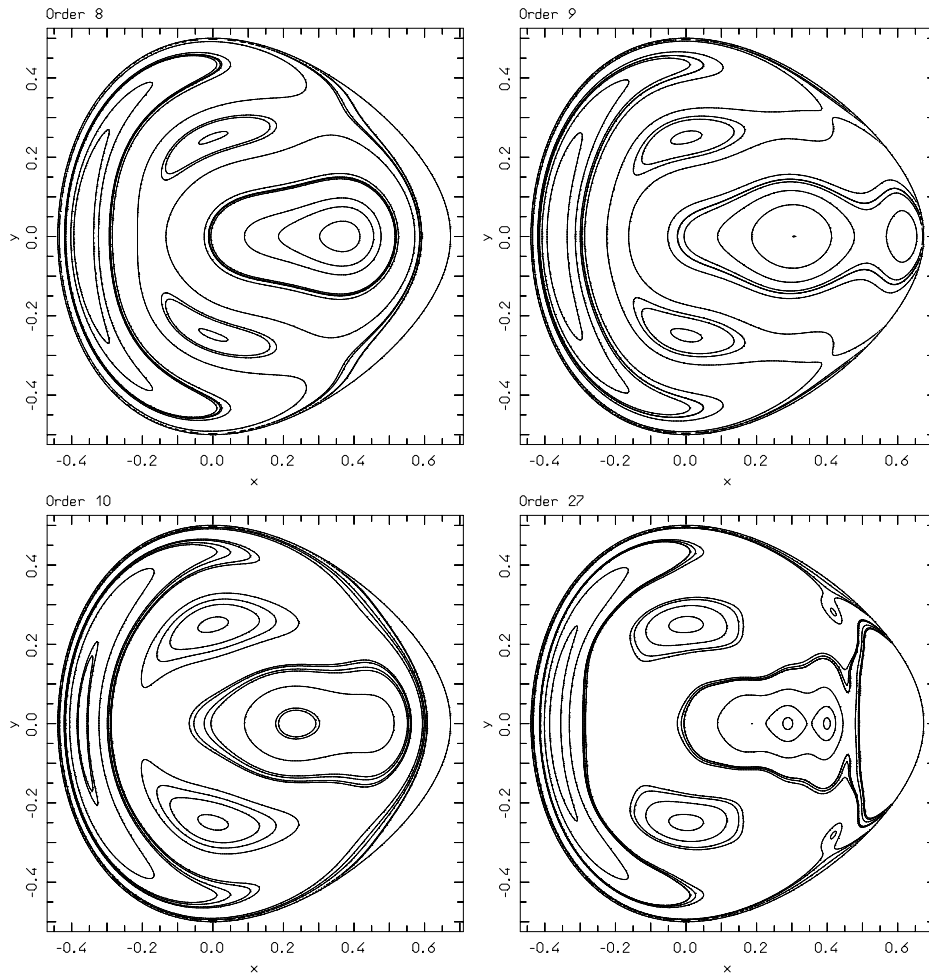$$\left\| \Phi_s \right\| = \sum_{j,k} \left| \varphi_{j,k} \right| .$$

Figure 6.   Level lines of the first integral truncated at orders 8, 32, 43 and 58, for energy $E = \frac{1}{12}$. A good correspondence with the Poincaré sections is found at orders, roughly, 8 to 32. Then the level lines start to disprove, in agreement with the asymptotic character of the series.

Then an indication of the convergence radius may be found by calculating one of the quantities

$$\left\|\Phi_s\right\|^{1/s} \ , \quad \frac{\left\|\Phi_s\right\|}{\left\|\Phi_{s-1}\right\|} \ , \quad \left(\frac{\left\|\Phi_s\right\|}{\left\|\Phi_{s-2}\right\|}\right)^{1/2} \ .$$

The first quantity corresponds to the root criterion for power series. The second one corresponds to the ratio criterion. The third one is similar to the ratio criterion, but in the present case turns out to be more effective because it takes into account the peculiar behavior of the series for odd and even degrees. The values given by the root criterion are plotted in the left panel of fig. 8. The data for the ratio criterion are plotted in the right panel, where open dots and solid

Figure 7.   Level lines of the first integral truncated at orders 8, 9, 10 and 27, for energy $E = \frac{1}{8}$. Some correspondence with the Poincaré sections is found around the order 9. Then the level lines are definitely worse, making even more evident the asymptotic character of the series.

dots refer to the second and third quantities in the formula above, respectively. In all cases it is evident that the values steadily increase, with no tendency to a definite limit. The almost linear increase is consistent with the behavior $\left\lVert \Phi_s \right\rVert \sim s!$ predicted by the theory.

8.2.   The Trojan asteroids

The asymptotic behavior of the series lies at the basis of Nekhoroshev theory on exponential stability. The general result, referring for simplicity to the case above, is that in a ball of radius ϱ and center at the origin one has

$$\left| \Phi(t) - \Phi(0) \right| < O(\varrho^3) \quad \text{for } |t| < O(\exp(1/\varrho^a)) \, ,$$
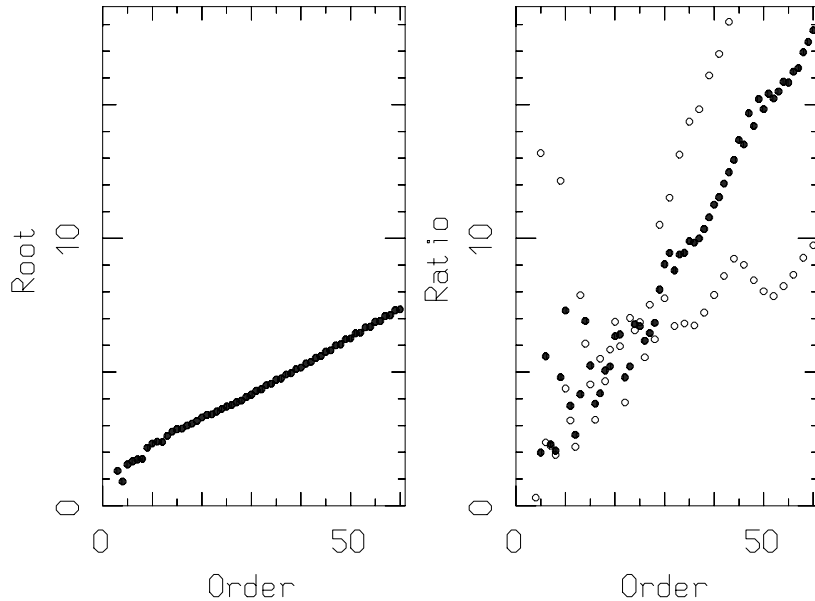
Figure 8. The convergence radius evaluated with the ratio (left) and the root (right) criterion. In both cases the non convergence of the series is evident.

for some positive $a \leq 1$. This is indeed the result given by the theory (see, e.g., Giorgilli 1988). In rough terms the idea is the following. Due to the estimate $\|\Phi_s\| \sim s!$ and remarking that $\dot{\Phi} = \{H, \Phi\}$ starts with terms of degree $s + 1$, one gets $|\dot{\Phi}| = O(s!\varrho^{s+1})$. Then one looks for an optimal degree s which minimizes the time derivative, i.e., $s \sim 1/\varrho$. By truncating the integrals at the optimal order one finds the exponential estimate.

However, the theoretical estimates usually give a value of ϱ which is useless in practical applications, being definitely too small. Realistic results may be obtained instead if the construction of first integrals for a given system if performed by computer algebra. That is, one constructs the expansion of the first integral up to an high order, compatibly with the computer resources available, and then looks for the optimal truncation order by numerical evaluation of the norms.

The numerical optimization has been performed for the expansion of the Hamiltonian in a neighborhood of the Lagrangian point L_4, in the framework of the planar circular restricted problem of three bodies in the Sun-Jupiter case. This has a direct application to the dynamics of the Trojan asteroids (see Giorgilli 1997).

The two first integrals which are perturbations of the harmonic actions have been constructed up to order 34 (close to the best possible with the computers at that time). The estimate of the time of stability is reported in fig. 9. The lower panel gives the optimal truncation order vs. $\log_{10} \varrho$. In the upper panel we calculate the stability time as follows: for an initial datum inside a ball of radius ϱ_0 we determine the minimal time required for the distance to increase up to $2\varrho_0$. Remark that the vertical scale is logarithmic. The units are chosen
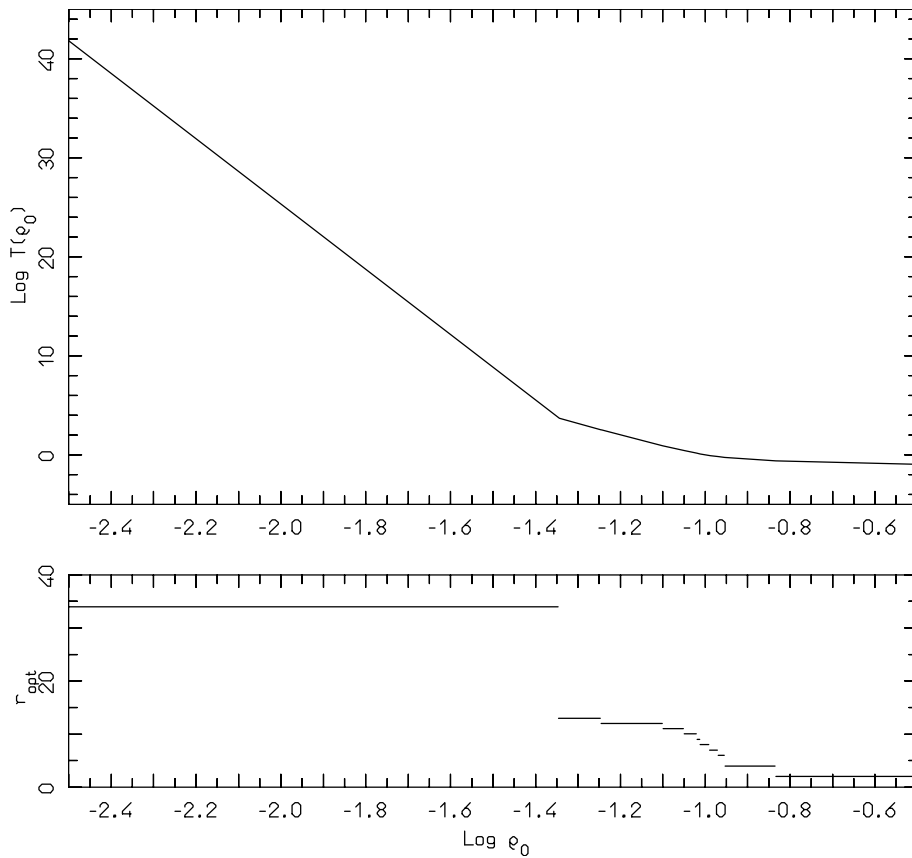
Figure 9.   The estimated stability time and the optimal truncation order
for the L_4 point of the Sun-Jupiter system.

so that $\varrho = 1$ is the distance of Jupiter from the Sun, and $t = 2\pi$ is the period
of Jupiter. With this time unit the estimated age of the universe is about 10^9.
The figure shows that the obtained data are already realistic, although, due to
the unavoidable approximations, only four of the asteroids close to L_4 known at
the time of that work did fall inside the region of stability for a time as long as
the age of the Universe.

8.3.   The SJSU system

As a third application we consider the problem of stability for the planar secular
planetary model including the Sun and three planets, namely Jupiter, Saturn and
Uranus. The aim is evaluate how long the semi-major axes and the eccentricities
of the orbits remain close to the current value (see, Sansottera 2010).

The problem here is much more difficult than in the previous cases. The
Hamiltonian must be expanded in Poincaré variables, and is expressed in action-
angle variables for the fast motions and in Cartesian variables for the slow mo-
tions, for a total of 9 polynomial and 3 trigonometric variables. The expansion of
the Hamiltonian in these variables clearly is a major task, that has been handled
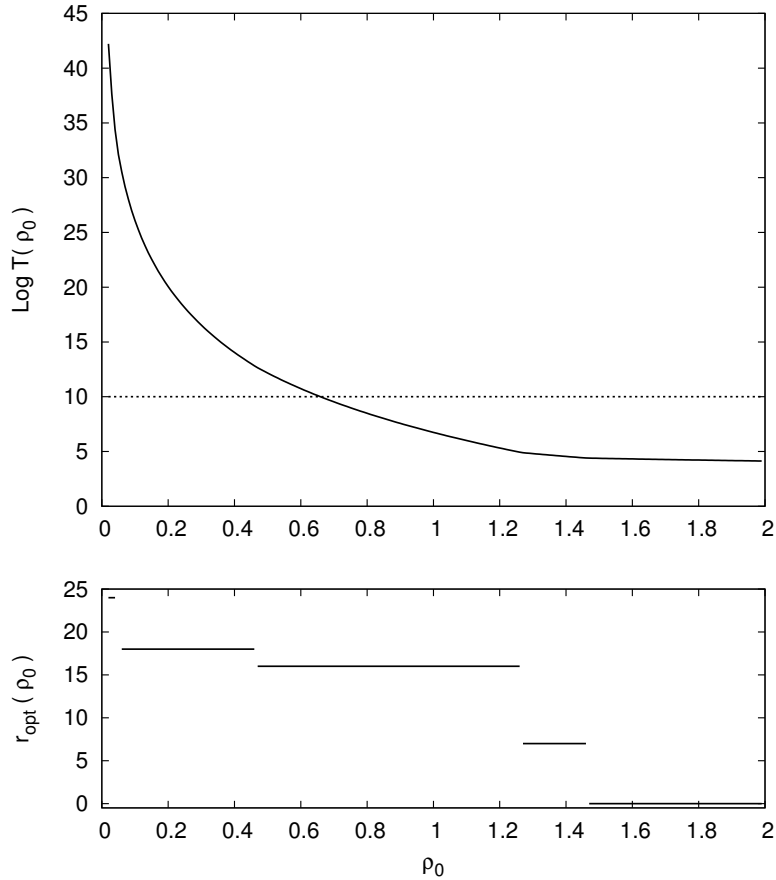via computer algebra.

Figure 10.   The estimated stability time and the optimal truncation order for the SJSU planar system. The dashed line corresponds to the estimated age of the Universe.

The reduction to the secular problem actually removes the fast motions, so that we get an equilibrium corresponding to an orbit of eccentricity zero close to a circular Keplerian one, and a Hamiltonian expanded in the neighborhood of the equilibrium, which is still represented as a system of perturbed harmonic oscillators, as in the cases above. Thus, after a long preparatory work, we find a problem similar to the previous one, that can be handled with the same methods.

The results are represented in fig. 10, where we report again the optimal truncation order and the estimated stability time, in the same sense as above. The time unit here is the year, and the distance is chosen so that $\varrho_0 = 1$ corresponds to the actual eccentricity of the three planets. The result is still realistic, although a stability for a time of the order of the age of Universe holds only inside a radius corresponding roughly to 70% of the real one.

References

Birkhoff, G. D. 1927, *Dynamical systems*, New York

Biscani, F. 2009, arXiv:0907.2076

Broucke, R. & Garthwaite, K. 1969, Cel. Mech., 1, 271

Broucke, R. 1989, Cel. Mec., 45, 255

Contopoulos, G. 1960, 49, 273

Contopoulos, G., 2002 *Order and Chaos in Dynamical Astronomy*, Springer

Contopoulos, G., Efthymiopoulos, C. & Giorgilli, A. 2003, J. Phys. A: Math. Gen., 36, 8639

Contopoulos, G., Efthymiopoulos, C. & Giorgilli, A. 2004, J. Phys. A: Math. Gen., 37, 10831

Giorgilli, A. & Galgani, L. 1978, Cel. Mech., 17, 267

Giorgilli, A., 1979, Comp. Phys. Comm., 16, 331

Giorgilli, A. 1988, Ann. Ist. H. Poincaré, 48, 423

Giorgilli, A., Delshams, A., Fontich, E., Galgani, L. & Simó, C. 1989, J. Diff. Eqs., 20

Giorgilli, A. & Skokos, Ch. 1997, Astron. Astroph., 317, 254

Gustavson, F. G., 1966, Astron. J., 71, 670

Hénon, M. and Heiles, C., 1964, Astron. J., 69, 73

Henrard, J., 1974, Cel. Mech., 10, 497

Henrard, J. 1986, : *Algebraic manipulations on computers for lunar and planetary theories*, Relativity in celestial mechanics and astrometry, 59

Henrard, J., 1989, Cel. Mech., 45, 245

Jorba, J. 1999, Experimental Mathematics, 8, 155

Knuth, D. 1968, *The Art of Computer Programming*, Addison-Wesley

Laskar, J, 1989, D. Benest & C. Froeschlé (eds.), *Les méthodes modernes de ma méchanique céleste (Goutelas, 1989)*, Ed. Frontières, 89

Poincaré, H. 1890, Acta Mathematica

Poincaré, H. 1892, *Les méthodes nouvelles de la mécanique céleste*, Gauthier–Villars, Paris

Rom, A. 1970, Cel. Mech., 1, 301

Sansottera, M., Locatelli U., & Giorgilli, A. 2010, Math. Comput. Simul., doi:10.1016/j.matcom.2010.11.018.

Siegel, C. L. 1941, Ann. Math., 42, 806

Whittaker, E. T. 1916, Proc. Roy Soc. Edinburgh, Sect. A, 37, 95

184

LECTURE

Diffusion in multidimensional Hamiltonian Systems Application to Galactic Dynamics

P.M. Cincotta [1,2], C.M. Giordano[1,2]

(1) Instituto de Astrofísica de La Plata (CONICET)
(2) Grupo de Caos en Sistemas Hamiltonianos
Facultad de Ciencias Astronómicas y Geofísicas - UNLP
Paseo del Bosque, B1900FWA La Plata, Argentina

Abstract. In the present lectures we provide results and discussions concerning the processes that lead to local and global chaotic diffusion in phase space of multidimensional conservative systems. We investigate and provide a measure of the extent of the domain over which diffusion may occur. All these issues are thoroughly discussed by dealing with a toy model such as the perturbed quartic oscillator as well as with a multidimensional conservative map that would be representative of the dynamics of a resonance interaction, which is an important mechanism in several astronomical systems, like the Solar System or galaxies.
Most of the contents of the present lectures are based on previous works of the authors already published in several journals.

1. Introduction

An open issue in multidimensional Hamiltonian systems is the so–called *chaotic diffusion*, that is, those mechanisms under which either global or local actions of an integrable Hamiltonian change over phase space (or action space) under the effect of a non–integrable perturbation. This kind of phenomena arises in Solar System and planetary dynamics, stellar dynamics such as star clusters and galaxies as well as in many other dynamical problems, like chemical reactions or particle acellerators.

At the present any complete theory that could describe global instabilities in phase space of multidimensional systems is still lacking. One could acquire accurate values of several indicators of the stability of the motion, but they would only provide local information for a neighborhood of a some point of phase space. A given orbit in a chaotic component in, for instance a 3D potential could have, positive and large values of two of its Lyapunov exponents, which does not necessarily mean that the unperturbed integrals will vary over vast domains.

Even though near–integrable Hamiltonian systems have been largely investigated –perhaps starting with the study conducted by Poincaré in the late nineteenth century (Poincaré 1893)–, the problem has not been completely elucidated yet. In fact, questions such as the stability of the motion of high–dimensional systems are far from being thoroughly understood. Some progress on transport phenomena has been made during the last two decades, but almost all the at-

tained results concern only low-dimensional symplectic maps or in other cases, non conservative one-dimensional maps, and it appears rather difficult to extent these approaches to be applied to many-dimensional Hamiltonians or conservative maps. Moreover, it is unclear how these approaches provide information about the role played by the torus structure and their invariant manifolds in spreading or preventing diffusion in phase space (see for instance, Meiss 1992, Venegerolies 2008, Korabel & Klags 2004).

Developing further in this direction Chirikov (1979) pioneered a complete survey of this matter in a somewhat heuristic way by using a standard mathematical language. Following this improvement, Cincotta (2002) revisited Chirikov's arguments, particularly that related to the so–called Arnol'd diffusion.

Regarded as a global instability, Arnol'd diffusion appears to be closer to a theoretical conjecture rather than to a real physical process (see Arnol'd 1964; Chirikov & Vecheslavov 1989; Chirikov & Vecheslavov 1993; Chirikov *et al.* 1985; Cincotta 2002). In fact, there remain so many unsolved mathematical details that Arnol'd diffusion, as a global instability, results a rather controversial question (see Lochak 1999). Further, the more recent numerical evidence reveals that Arnol'd diffusion might operate in certain (somewhat artificial) dynamical systems (Guzzo *et al.* 2005, 2006; Lega *et al.* 2010). In fact, in a previous work Lega *et al.* (2008), investigated diffusion in near–integrable Hamiltonian Systems (or symplectic maps) and discuss and compute local and global Arnol'd diffusion in two different scenarios, the Froeschlé four-dimensional map (see Froeschlé 1971, 1972) and a 3D Hamiltonian model previously investigated by for instance Guzzo *et al.* (2002), Froeschlé *et al.* (2005). Also in the lecture by Efthymiopoulos (2012) in this proceedings, a thoroughly discussion about Arnol'd diffusion (or diffusion along a singe and/or double resonance) is presented, using the same 3D Hamiltonian system mentioned above, including rather instructive pictures of how diffusion proceeds. Besides an attempt to relate the size of the remainder of the optimal normal form construction and Chirikov's Arnol'd diffusion coefficient estimate is discussed. We strongly recommend the interested reader to explore in the exhaustive list of references on the subject in this mentioned lecture.

Nonetheless, it has been shown that Arnol'd diffusion–like processes may play a significant role in global chaotic diffusion in phase space, (Giordano & Cincotta 2004); (Cachucho *et al.* 2010), though the mechanisms that drive the diffusion remain still unknown. By *Arnol'd diffusion–like* we mean, chaotic transport along resonance (or relics of resonances), that geometrically resembles Arnol'd diffusion, but in those scenarios most of the phase space is chaotic, mainly due to strong overlap of the main resonances so the systems is quite far from the regime under which Arnol'd diffusion is usually investigated. Roughly speaking, the perturbation is not small enough such that most of the resonance structure is preserved (see the discussions given in the present lecture).

In (Cachucho *et al.* 2010) Chirikov's diffusion approach is applied to the (5,-2,-2) three body mean motion resonance for the (490) Verita's family. It is shown that the theoretical arguments used by Chirikov to describe Arnol'd diffusion could also apply in this realistic problem in which the so–called guiding resonance domain is completely chaotic. However, the scenario of modulational

diffusion (Chirikov *at al.* 1985) could perhaps be more suitable to describe the large diffusion observed in the eccentricity of asteroids.

In the present lecture we discuss a preliminary study of diffusion on phase space for a rather simple dynamical system in order to elucidate its efficiency to connect different chaotic components, already given in Giordano & Cincotta (2004) and Giordano & Cincotta (2008). Despite the simplicity of the adopted model, several results concerning its dynamics apply to any 3D Hamiltonian system exhibiting a divided phase space, such as galactic or planetary systems. In Section 4., two values for the perturbation parameter have been selected so that the dynamics of the toy model Hamiltonian resembles that of a galactic system (for a moderate perturbation, in which case both components are comparable) and asteroidal dynamics (for a large perturbation for which the chaotic component prevails).

Observational evidence, theoretical arguments and N-body simulations suggest that a model resembling an isolated elliptical galaxy should exhibit a divided phase space and therefore the perturbation should be moderate (see, for instance Merritt & Friedman 1996; Merritt & Valluri 1996, Papaphilippou & Laskar 1998; Gerhard & Binney 1985; Poon & Merritt 2002; Muzzio *et al.* 2005).

Herein in Section 5.2. we also show, discuss and measure chaotic diffusion in action space by means of a conservative multidimensional map, the Coupled Rational Shifted Standard Map (CRSSM), whose global dynamical properties have been previously studied in Cincotta *et al.* (2003) –CGS03 hereafter–, for different sets of parameters, by recourse of the Mean Exponential Growth Factor of Nearby Orbits (MEGNO). First introduced by Cincotta & Simó (2000) and latter generalized in CGS03, the MEGNO is a rather efficient indicator of the dynamics belonging to the class of the so–called fast indicators, which has already become of widespread use. Many applications of this tool to Solar System dynamics, exoplanets models and as well as to many other dynamical systems could be found throughout the literature. The obtained results are taken from Cincotta & Giordano (2012).

Furthermore, we provide a simple tool which turns to be suitable for measuring the variation of the unperturbed integrals. We face the difficulty to compute a meaningful diffusion coefficient, due to the fact that, as far as we know, it still remains unclear which would be the appropriate approach to be considered, in particular, how the variance of the variables scales with time. In fact, in case of normal diffusion the variance scales linearly with time, but for what is called abnormal diffusion, the scaling runs like t^b where the parameter b is in general unknown and further, it strongly depends on the local dynamical structure of the deemed region of phase space (see for instance, Cordeiro & Mendes de Souza 2005, Cordeiro 2006, Mestre 2012). The exhaustive numerical exploration performed in Cincotta & Giordano (2012) and the results of their application are discussed in Section 5.5.

2.   A real galaxy

A galaxy is a very complex physical system since it is composed of $\sim 10^{11} - 10^{15}$ stars whose space distribution (or density) generates its own gravitational field. Some types of galaxies present a rotating pattern, spiral arms and bars. Gas

and dust are generally present, star formation processes occur, supernova explosions release a large amount of energy and the chemical evolution certainly should affect the hydrodynamics. Besides, galaxies are not isolated: gravitational interaction with other galaxies and stellar clusters is always present. It is claimed that a super–massive black hole lies in the center of any galaxy. Thus, all these processes and interactions should be considered when modeling a rough real galaxy, which in fact seems to be a rather difficult problem to cope with. Further aspects are still open, like for instance if they are in a steady state or how to cope with the unknown dark matter, if actually dark matter does exist (see the lecture by Carati & Galgani 2012 in the present volume).

3. A galaxy as an idealization of a N–particle Hamiltonian system

Following Binney & Tremaine (1987), let $f^{(N)}(\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N, \boldsymbol{p}_1, \cdots, \boldsymbol{p}_N, t)$ be the N–particle probability density or distribution function (DF) on 2N–D phase space Γ, so it satisfies the Liouville theorem:

$$\frac{\mathrm{d}f^{(N)}}{\mathrm{d}t} = 0.$$

Let us define the 1 – particle DF:

$$f^{(1)}(\boldsymbol{x}_1, \boldsymbol{p}_1, t) = \int f^{(N)} \mathrm{d}^3 x_2 \cdots \mathrm{d}^3 x_N \mathrm{d}^3 p_2 \cdots \mathrm{d}^3 p_N,$$

Assuming that:

- $|(\boldsymbol{x}_i, \boldsymbol{p}_i)|^k f^{(N)} \to 0$ as $|(\boldsymbol{x}_i, \boldsymbol{p}_i)| \to \infty \quad \forall k, \forall i = 1, \cdots, N$,

- $f^{(N)}$ is symmetric in $\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N; \boldsymbol{p}_1, \cdots, \boldsymbol{p}_N$,

- the 2–particle DF could be written in terms of the 2–particle correlation function:

$$f^{(2)}(\boldsymbol{x}_1, \boldsymbol{p}_1, \boldsymbol{x}_2, \boldsymbol{p}_2, t) = f^{(1)}(\boldsymbol{x}_1, \boldsymbol{p}_1, t) \, f^{(1)}(\boldsymbol{x}_2, \boldsymbol{p}_2, t) + g(\boldsymbol{x}_1, \boldsymbol{p}_1, \boldsymbol{x}_2, \boldsymbol{p}_2, t),$$

- the 2 – particle correlation function $g \approx 0$ (Two–body relaxation time: $T_R \sim 10^{15}$ yrs.),

- $N \gg 1$,

- and $f(\boldsymbol{x}, \boldsymbol{v}, t) \equiv \mathrm{N} f^{(1)}(\boldsymbol{x}_1, \boldsymbol{p}_1/m, t)$

- $(\boldsymbol{x}, \boldsymbol{v}) \equiv (\boldsymbol{x}_1, \boldsymbol{p}_1/m)$

we arrive to the so-called collisionless Boltzmann (or Vlasov) equations:

$$\frac{\partial f}{\partial t} + [f, H] = 0, \quad H(\boldsymbol{p}, \boldsymbol{x}) = \frac{\boldsymbol{p}^2}{2} + \Phi(\boldsymbol{x}, t), \qquad \boldsymbol{p} \equiv \boldsymbol{v},$$

being $\Phi(\boldsymbol{x}, t)$ a smooth potential generated by the star distribution. This is the Liouville theorem in the 6–D phase space, μ. Thus, it is necessary to solve:

$$\frac{\partial f}{\partial t} + \nabla_{\boldsymbol{x}} f \cdot \boldsymbol{v} - \nabla_{\boldsymbol{x}} \Phi \cdot \nabla_{\boldsymbol{v}} f = 0,$$

and the Poisson equation:

$$\nabla^2 \Phi = 4\pi G m \int f(\boldsymbol{x}, \boldsymbol{v}, t) \mathrm{d}^3 v.$$

3.1. Steady state solutions

A steady state solution for the collisionless Boltzmann and Poisson equations is given by the so–called strong Jeans Theorem (Lynden–Bell 1962, Binney & Tremaine 1987):

The DF of a steady–state galaxy in which almost all orbits are regular with incommensurable frequencies may presumed to be a function only of three independent isolating integrals.

An *isolating integral* $I(\boldsymbol{x}(t), \boldsymbol{v}(t)) = c$ defines a manifold in μ of lower dimension than $\dim(\mu)$. In Jeans Theorem, the three isolating integrals are, for instance, the three global actions or integrals of $H(\boldsymbol{p}, \boldsymbol{x})$.

An alternative version of the Jeans Theorem could be stated as follows:

If the Hamiltonian of a collisionless stellar system in steady-state equilibrium is Arnol'd–Liouville integrable, the DF has a constant value at every point in an invariant torus of the system (Efthymiopoulos et al. 2008).

But galaxies should present a divided phase space. Therefrom the implementation of Jeans theorem in more realistic stellar systems is rather difficult, for instance:

- It is unclear how the approximate integrals or actions should be included as arguments of the DF when the system is near–integrable;

- Resonances and resonance intersections are dense in phase space;

- One or two integrals do not exist for the chaotic domains of phase pace.

A generalization for non–integrable potentials is given by Merritt (1999):

The phase space density of a stationary stellar system must be constant within every well-connected region.

But:

- An invariant non–resonant torus is a well–connected region when $t \to \infty$,

- A resonant elliptic torus also is a well–connected region of lower dimension (2D) and when $t \to \infty$,

- An hyperbolic tours, also has a lower dimension, and the dynamics is much more complicate when we add a non–in tegrable perturbation to an integrable Hamiltonian,

- Are the chaotic domains in phase space well–connected regions?

Therefore it is rather important to investigate chaotic diffusion in phase space to understand, for instance, if a single chaotic component does exist. To deal with this issue we can use a non–integrable Hamiltonian (toy) model or instead a multidimensional map. Thus, we will be considering a perturbed 3D quartic oscillator and a 4D symplectic map.

4.   Chaotic diffusion of orbits in systems with a divided phase space for a 3D Hamiltonian system

In order to build up an equilibrium model for a galaxy which is assumed to be represented by a smooth gravitational field it is necessary to know beforehand its global dynamics. It seems likely that any realistic model should exhibit a divided phase space, that is, the motion would take place either in a stable, regular component or in one or more unstable, chaotic components. In 3D systems displaying such a dynamics, the existence of three, in general local, invariants, allows the presence of invariant tori where regular, quasi-periodic or resonant motion takes place. The disruption of these local invariants, mainly due to resonance interactions, leads to the appearance of a chaotic component.

However, the first attempts to investigate these matters in the astrophysical literature assumed that for dynamical systems with more than two degrees of freedom, the chaotic component is fully connected. In fairly recent studies, such as Merritt & Valluri (1996) and Merritt & Fridman (1996), the authors study whether this full connection of the chaotic component may occur in realistic physical times. If such were the case, it would imply that the orbit of a star within the chaotic domain would explore the whole region, which, in general, would comprise a large fraction of the energy surface. As a consequence the distribution function on the chaotic component would depend only on the energy (see Merritt (1999) for a thorough discussion).

As shown in Giordano & Cincotta (2004) we obtained numerical evidence that for moderate-to-strong chaotic systems diffusion does not occur over the whole chaotic component and only when the latter fills almost all the energy surface may diffusion become significant but in extremely large times. We chose a rather simple Hamiltonian system, the perturbed quartic oscillator in order to have a complete knowledge of the local and global dynamics and the transition to chaos, and particularly to be able to study in detail the motion in the resonance intersections. The forthcoming sections summarize the main results of this work.

Therefore, let us now be concerned with the perturbed uncoupled quartic oscillator:

$$\tilde{H}(\boldsymbol{p}, \boldsymbol{q}) = \frac{\boldsymbol{p}^2}{2} + \frac{1}{4}(x^4 + y^4 + z^4) + \varepsilon x^2(y + z). \tag{1}$$

whose full dynamics has been investigated by Cincotta & Giordano (2002) and Cincotta *et al.* (2003) by means of the *Mean Exponential Growth factor of*

Nearby Orbits (MEGNO) (a detailed formulation of this technique can be found in Cincotta *et al.*, 2003 and Cincotta & Simó, 2000.)

The advantage of this 3D toy model is that it can be easily written in terms of action variables and that the coordinates admit of a simple Fourier expansion. Indeed, in terms of the unperturbed action-angle variables, $(I_1, I_2, I_3; \theta_1, \theta_2, \theta_3)$, the Hamiltonian (1) can be recast as

$$H(\boldsymbol{I}, \boldsymbol{\theta}) = H_0(\boldsymbol{I}) + \varepsilon V(\boldsymbol{I}, \boldsymbol{\theta}), \tag{2}$$

where

$$H_0(\boldsymbol{I}) = A\left(I_1^{4/3} + I_2^{4/3} + I_3^{4/3}\right), \tag{3}$$

with $A = (3\beta/2\sqrt{2})^{4/3}$, $\beta = \pi/2K(1/\sqrt{2})$, $K(k)$ denoting the complete elliptic integral. The perturbation, which can be assumed to be small as long as $\varepsilon \ll 1$, admits of the Fourier expansion

$$
\begin{aligned}
V(\boldsymbol{I}, \boldsymbol{\theta}) \;=\; & \hat{V}_{12} \sum_{n,m,k=1}^{\infty} \alpha_{nmk}\Big(\cos\big(2(n{+}m{-}1)\theta_1 \pm (2k{-}1)\theta_2\big) \\
& +\; \cos\big(2(n{-}m)\theta_1 \pm (2k{-}1)\theta_2\big)\Big) \\
& +\; \hat{V}_{13} \sum_{n,m,k=1}^{\infty} \alpha_{nmk}\Big(\cos\big(2(n{+}m{-}1)\theta_1 \pm (2k{-}1)\theta_3\big) \\
& +\; \cos\big(2(n{-}m)\theta_1 \pm (2k{-}1)\theta_3\big)\Big),
\end{aligned}
\tag{4}
$$

the functions $\hat{V}_{ij}$ and the coefficients α_{nmk} being

$$
\begin{aligned}
\hat{V}_{ij} \;&=\; 2^{5/2}3\beta^4 I_i^{2/3} I_j^{1/3}, \qquad \alpha_{nmk} = \alpha_n \alpha_m \alpha_k, \\
\alpha_s \;&=\; \frac{1}{\cosh\big((s-1/2)\pi\big)}, \qquad \frac{\alpha_{s+1}}{\alpha_s} \approx \frac{1}{23},
\end{aligned}
$$

and can be split in two, namely V_{xy} and V_{xz}, on introducing the integer vectors $\boldsymbol{l} = (l_1, l_2, 0)$ and $\boldsymbol{k} = (k_1, 0, k_3)$, and the new coefficients $\hat{\alpha}_{l_1 l_2}$ and $\hat{\alpha}_{k_1 k_3}$,

$$
\begin{aligned}
V_{xy}(I_1, I_2; \theta_1, \theta_2) \;&=\; \hat{V}_{12} \sum_{l_1, l_2} \hat{\alpha}_{l_1 l_2} \cos(l_1\theta_1 + l_2\theta_2), \\
V_{xz}(I_1, I_3; \theta_1, \theta_3) \;&=\; \hat{V}_{13} \sum_{k_1, k_3} \hat{\alpha}_{k_1 k_3} \cos(k_1\theta_1 + k_3\theta_3).
\end{aligned}
$$

The concomitant unperturbed frequency vector is given by

$$\boldsymbol{\omega}(\boldsymbol{I}) = \frac{\partial H_0}{\partial \boldsymbol{I}} = \frac{4}{3}A\big(I_1^{1/3}, I_2^{1/3}, I_3^{1/3}\big). \tag{5}$$

The knowledge of the resonance structure of the unperturbed Hamiltonian H_0, is of actual relevance. Here the resonance structure on the energy surface can be easily visualized by introducing a change of coordinates such that

the unperturbed energies in each degree of freedom, h_1, h_2, h_3, become the new action–like variables. We obtain

$$H_0(h_1, h_2, h_3) \quad = h_1 + h_2 + h_3,$$

$$\boldsymbol{\omega}(h_1, h_2, h_3) \quad = \sqrt{2}\beta\big(h_1^{1/4}, h_2^{1/4}, h_3^{1/4}\big),$$

and, in terms of (h_1, h_2), the resonance condition, $\boldsymbol{m} \cdot \boldsymbol{\omega} = 0$, $\boldsymbol{m} \in \boldsymbol{Z^3}/\{\boldsymbol{0}\}$, for $H_0 = h$ can be recast in the form

$$\begin{aligned}
(m_1^4 + m_3^4)\xi^4 + 4m_1^3 m_2 \xi^3 \eta + 6m_1^2 m_2^2 \xi^2 \eta^2 \\
+ 4m_1 m_2^3 \xi \eta^3 + (m_2^4 + m_3^4)\eta^4 - m_3^4 = 0,
\end{aligned} \tag{6}$$

where $\xi = (h_1/h)^{1/4}$, $\eta = (h_2/h)^{1/4}$.

Note that for those harmonics in which one of the m_i is zero, the resonant polynomial (6) can be easily solved to yield

$$\hat{h}_2 = \frac{m_1^4}{m_2^4}\hat{h}_1, \quad m_1 m_2 < 0, \quad m_3 = 0, \tag{7a}$$

$$\hat{h}_2 = 1 - \left(\frac{m_1^4 + m_3^4}{m_3^4}\right)\hat{h}_1, \quad m_1 m_3 < 0, \quad m_2 = 0, \tag{7b}$$

$$\hat{h}_2 = \left(\frac{m_3^4}{m_2^4 + m_3^4}\right)(1 - \hat{h}_1), \quad m_2 m_3 < 0, \quad m_1 = 0, \tag{7c}$$

with $\hat{h}_i = h_i/h$, showing that those resonances associated to resonant vectors with at least one null m_i appear as straight lines on the energy surface $\hat{h}_1 + \hat{h}_2 + \hat{h}_3 = 1$.

The width of any of those resonances can be computed by means of a simple pendulum approximation, which is a suitable description in the case in which we assume each resonance is isolated from the rest.

Let us notice that the map $\mathbf{I} \mapsto \mathbf{h}$ transforms the unperturbed energy surface into a plane, so that we can perform a second (global) change of coordinates, $(h_1, h_2, h_3) \mapsto (e_1, e_2, e_3)$, in such a way that the e_3–axis is normal to the energy plane,

$$e_1 = \frac{1}{\sqrt{6}}(h_1 - 2h_2 + h_3), \tag{8a}$$

$$e_2 = \frac{1}{\sqrt{2}}(h_1 - h_3), \tag{8b}$$

$$e_3 = \frac{1}{\sqrt{3}}(h_1 + h_2 + h_3), \tag{8c}$$

with

$$-\sqrt{\frac{2}{3}} \le \frac{e_1}{h} \le \frac{1}{\sqrt{6}}, \quad -\frac{1}{\sqrt{2}} \le \frac{e_2}{h} \le \frac{1}{\sqrt{2}}, \quad \frac{e_3}{h} = \frac{1}{\sqrt{3}}. \tag{9}$$

The resonance structure of the unperturbed system on the energy surface and the theoretical widths of the principal resonances appearing in the perturbation are presented in Figs. 1[a] and [b], respectively.

On applying the MEGNO to the study of the global dynamics of Eq. (1), one obtains the resonance structure presented in Fig. 1[c]. There, the details of the phase space structure at a low-to-moderate value of the perturbation ($\varepsilon = 5 \times 10^{-3}$) are displayed, depicting the obtained values for the MEGNO in a contour–like plot where resonances can be clearly distinguished.
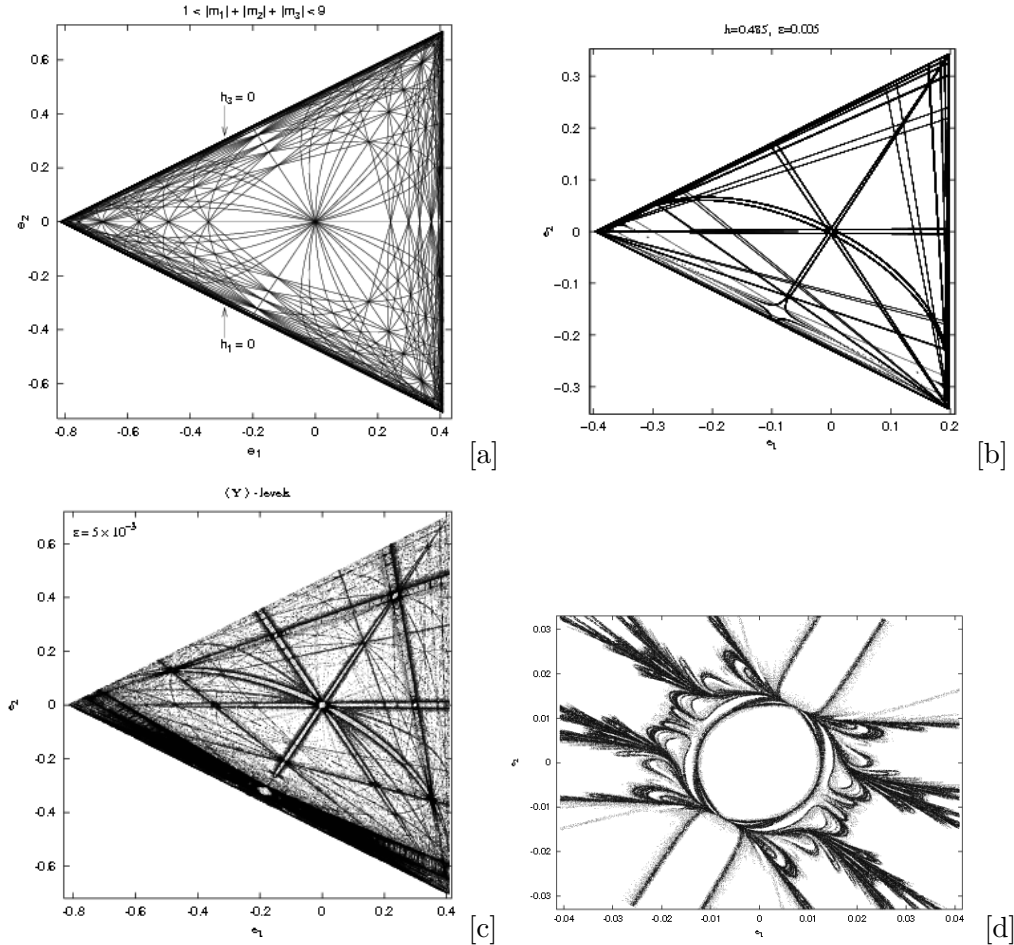
Figure 1.   [a] Resonances of the unperturbed Hamiltonian (3) for $1 < |m_1| + |m_2| + |m_3| < 9$ for the rescaled energy $h = 1$ yielding the theoretical Arnold web on the energy surface. [b] Strongest resonances and their theoretical widths for $\varepsilon = 0.005$. [c] Actual dynamics of the system revealed by the MEGNO for the same value of ε. [d] Blow up around the origin. These figures were taken from Cincotta *et. al*, 2003 and Mestre *et. al*, 2009.

At this stage, a brief reference to MEGNO's behavior is required in order to grasp the dynamical information comprised in such a figure. Let us then recall that in the case of regular motion, the MEGNO ($\overline{Y}$) tends asymptotically to a fixed value independent of the orbit, namely $\overline{Y} \to 2$. Small departures

from this value indicate the proximity of some periodic orbit, where $\overline{Y} \lesssim 2$ and $\overline{Y} \gtrsim 2$ for stable and unstable periodic orbits, respectively. On the other hand, for irregular, stochastic motion $\overline{Y}$ grows linearly with time as $t \to \infty$, at a rate equal to $\sigma/2$, σ denoting the largest Lyapunov Characteristic Number (LCN) of the orbit.

Let us devote the following paragraphs to describe the procedure giving rise to Fig. 1[c], and Figs. 2, which correspond to higher values of the perturbation.
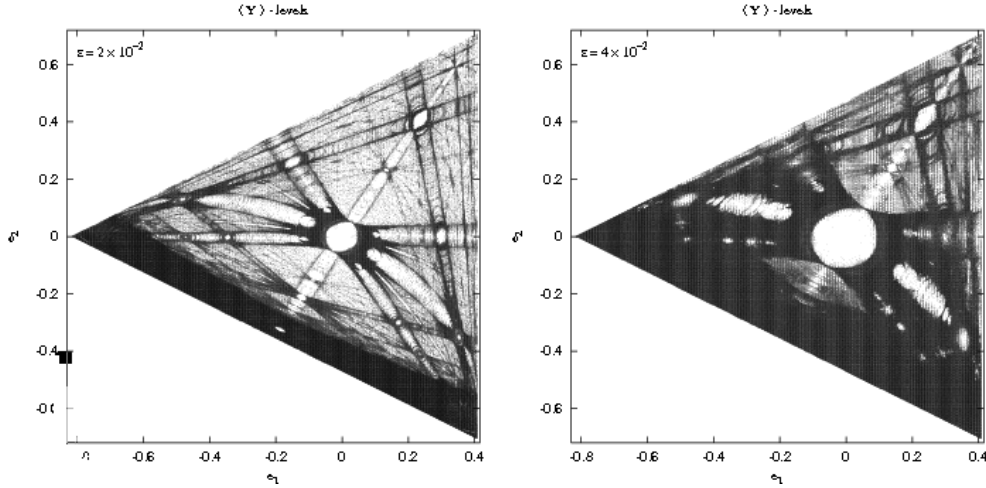
Figure 2. $\overline{Y}(t_f)$–levels on the energy surface for $\varepsilon = 0.02$ and 0.04.

For each adopted value of ε, we take values of h_1 and h_2 with $0 \leq h_1, h_2 \leq h$, $h_3 = h - h_1 - h_2$, where h_1 and h_2 are of the form $jh/1,000$; $j = 0, \ldots, 1,000$. This leads to 501,501 initial conditions for which we take $(x, y, z) = (0, 0, 0)$. The equations of motion are integrated together with their first variationals over a total motion time $t_f = 3,500T$. For the tangent vector, we adopt the initial values $\delta_x = \delta_y = \delta_z = 0$ and δ_{p_i} chosen at random in the interval $(-1, 1)$ and then normalized to 1. For each orbit we compute both $\overline{Y}(t_f)$ and the rate at which the MEGNO grows with time. The actual energies h_1, h_2, h_3, are scaled to the interval $[0, 1]$ (by division through h) and then transformed to the energy plane (e_1, e_2) by means of Eqs. (8).

In Fig. 1[c], corresponding to $\varepsilon = 5 \times 10^{-3}$, we display the values of $\overline{Y}(t_f)$ binned in five intervals, two of them being very narrow and close to 2 (see figure caption for details).

In this picture most of the main resonances can be clearly distinguished as light gray channels surrounded by dark boundaries. Four main resonances are seen to intersect at the origin: the three lines corresponding to the $(1, -1, 0)$, $(1, 0, -1)$, $(0, 1, -1)$ resonances and the curve associated to the $(-2, 1, 1)$ resonance.

The actual width of the resonances as well as the narrow stochastic layers at their edges can be clearly visualized. The center of any resonance "channel" corresponds to a sequence of 2D elliptic tori while its borders ("stochastic layer") correspond to a sequence of 2D hyperbolic tori. We observe a strip of chaotic
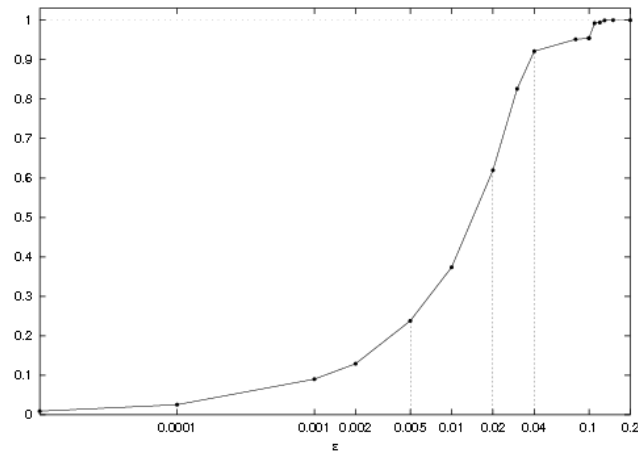
Figure 3. Fraction of chaotic motion vs. ε (in logarithmic scale). The vertical lines refer to the values of ε adopted for Fig. 1[c] and Fig. 2.

motion close to $h_1 = 0$. The presence of this region is easily understood from Fig. 1a as the overlap of the strongest resonances, e.g. $(2, -1, 0)$ and $(2, 0, -1)$, as well as of many others (not shown in Fig. 1[a]).

The dynamics at the intersection of resonances can be quite intricate. To illustrate such a feature, let us note the complexity of the picture reproduced in Fig. 1[d], where we present a zoom around the intersection of resonances at the origin in Fig. 1[c]. That contour plot was obtained with a higher resolution in h_1 and h_2 and for a total motion time $t_f = 350T$. There the MEGNO reveals the existence of several stability zones, which should be responsible for restraining the spread of chaotic motion, acting as barriers to diffusion. They are the well-known sticky tori surrounding the periodic orbit located at the center of the resonance. This plot is also very illustrative to see how the manifolds of lower dimensional tori bend in a complex fashion, giving rise to the many tight loops seen in the picture. These manifolds are important because they are the objects able to carry the motion arriving along one of the resonances either to the "other part" of the resonance or to a different resonance.

As the perturbation is increased, resonances become wider. Fig. 2 displays the actual structure of action space ($\boldsymbol{e}$–space) for two different (rather large) values of the perturbation parameter, namely $\varepsilon = 0.02$ and $\varepsilon = 0.04$. There, some details of the dynamics at moderate-to-high-level perturbations are shown in a plot similar to that in Fig. 1[c]. The character of the motion (resonant, quasi-periodic and stochastic) is represented in gray scale, from white to black, the different $\overline{Y}(t_f)$ intervals have been selected so as to sharpen the details of the phase space structure in each case.

Notice that for $\varepsilon = 0.02$ a significant part of the energy surface still looks regular, with wide resonance domains and a broad chaotic strip, while it looks rather chaotic for a somewhat slightly larger perturbation. Note however that within the weaker chaotic region, the MEGNO is still capable of unveiling the relics of resonance structures.

In Fig. 3 we present the fraction of chaotic motion as ε is increased, fixing a threshold of $\overline{Y}(t_f) = 2$ below which orbits are regarded as regular. The global amount of stochasticity as a function of ε is measured by counting how many pixels have a value of $\overline{Y}$ that exceeds the adopted threshold. The fraction of chaotic motion for the values of ε considered above, namely 0.02 and 0.04, are 38% and 91.7%, respectively. Further, for values of $\varepsilon \geq 0.1$ the system exhibits completely chaotic dynamics.

4.1. Diffusion on the energy surface at moderate–to–high perturbation

Let us now analyze the diffusion on energy surface at moderate–to–high perturbations, following several orbits with initial conditions in different stochastic domains.

Table 1.   Initial conditions on the energy surface for seven orbits with high values of the MEGNO. The energies h_1, h_2 are given in units of $h/250$, where $h = 1/(4\beta^4)$ is the total energy. The exact values of e_1 and e_2 should be obtained by means of Eqs. (8); their approximated values have been included so that they can be easily visualized on the energy surfaces.

i.c.	h_1	h_2	e_1	e_2
i	20	100	−0.0816	−0.3112
ii	10	125	−0.2041	−0.2969
iii	85	69	0.1143	−0.0565
iv	58	21	0.3053	−0.3116
v	70	134	−0.2482	0.0678
vii	104	140	−0.2776	0.2771
vii	191	15	0.3347	0.4157

For seven different initial conditions listed in Table 1 we have followed the wandering of the unperturbed integrals over the (e_1, e_2)-plane. We have picked two initial conditions in the chaotic strip close to $h_1 = 0$, (i) and (ii) –see Fig. 2, another in the region surrounding the more regular central zone at the origin (iii) and four other initial conditions located near the crossing of resonances (iv) − (vii); all of them have very high values of the MEGNO. The origin has also been considered; it actually corresponds to a regular orbit.

The equations of motion have been integrated by means of a Runge–Kutta 7/8th order integrator (the so-called Dopri8 routine, see Hairer *et al*, 1987).

The results corresponding to $\varepsilon = 0.02$ for a total motion time of 3×10^6 characteristic periods T of the system are presented in Fig. 4[a] (we refer to the corresponding Fig. 2 for comparison). We note that even in the case of moderate perturbation, diffusion is completely irrelevant over such a timescale. Only when the resonances labelled by the harmonics $(2, −1, 0)$ and $(2, 0, −1)$ overlap does fast diffusion occur along such resonances, but it is restricted to a relatively small region of the energy surface. For the remaining initial conditions considered the unperturbed integrals remain confined to rather small domains. In particular,
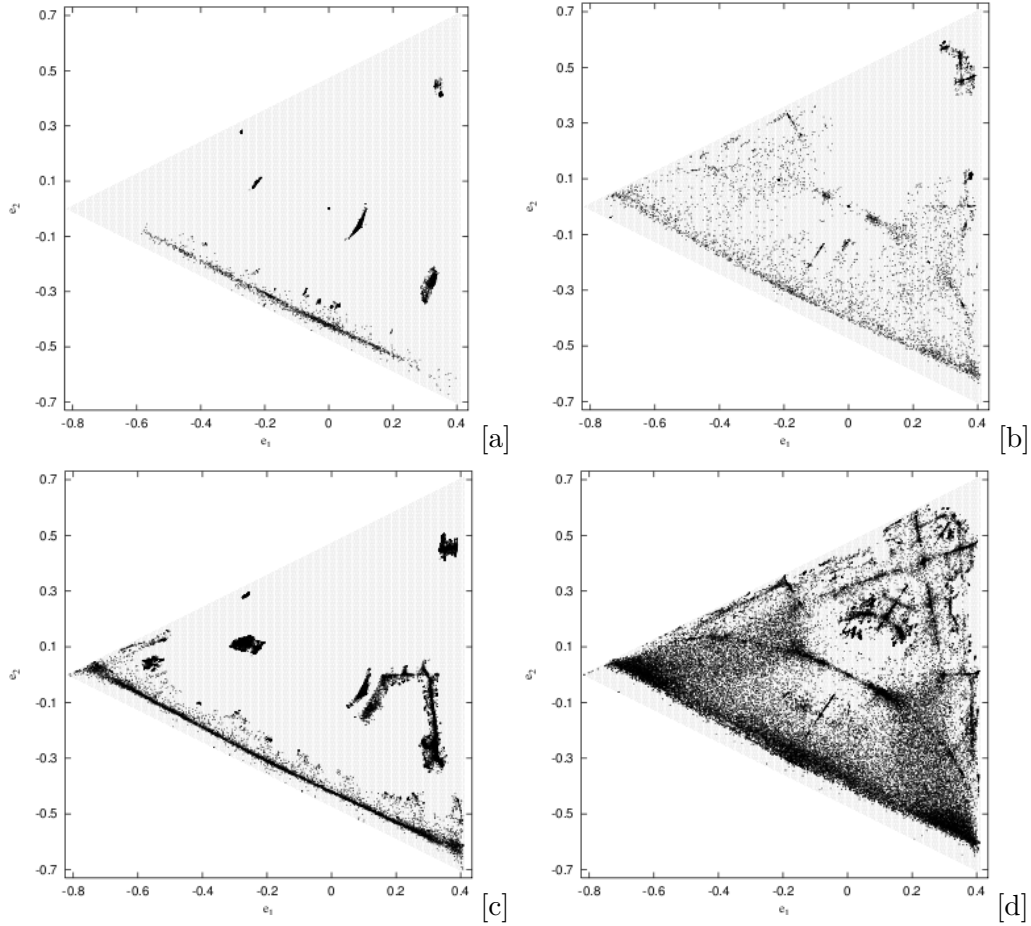
Figure 4. Diffusion on the energy surface at moderate–to–high perturbations after 3×10^6 characteristic periods T of the system for: [a] $\varepsilon = 0.02$, and [b] $\varepsilon = 0.04$; and after $3 \times 10^8 T$ for: [c] $\varepsilon = 0.02$, and [d] $\varepsilon = 0.04$.

for the initial condition (vi) selected near the boundary $h_3 = 0$, the variation of the unperturbed integrals proves to be small, differing slightly from the expected behavior in case of stability. For the regular orbit at the origin, considered just for illustrative purposes, the wandering is restrained to a point, as expected.

The plot in Fig. 4[b] corresponds to $\varepsilon = 0.04$ and should be compared with the contour plot in Fig. 2, where the chaotic regime prevails. Notice must be taken however that in the central region in the latter figure, for which $t_f = 3,500T$, there lie several chaotic orbits; the light gray points and even some of the white points within it have values of the MEGNO larger than 2 (as indicated in the caption). Recall that for constructing Fig. 2 the values of the MEGNO are binned in intervals chosen so as to highlight the dynamical structure of phase space, rather than discriminate stable quasiperiodic motion from chaotic motion. Thus, those chaotic orbits depicted in white in the contour plot having MEGNO values smaller than 2.1 are likely to behave in a regular fashion for rather short timescales, such as $3,500T$. Furthermore, Fig. 3 reveals that for $\varepsilon = 0.04$,

the system shows up as globally chaotic, with just a small fraction of phase space occupied by stable motion. This explains why, for the larger total motion time now considered, 3×10^6 characteristic periods T of the system, diffusion manages to trespass the central zone, the points being mainly concentrated along a fairly well–defined strip corresponding to the curve associated to the $(-2, 1, 1)$ resonance (see below). Note that the upper part of the central zone remains unvisited at this timescale, except for a rather strong but confined concentration of points near the corner.

On considering an even larger timescale, 3×10^8 characteristic periods T of the system, for $\varepsilon = 0.02$ (Fig. 4[c]) the fast diffusion occurring at the overlap of the resonances $(2, -1, 0)$ and $(2, 0, -1)$ after 3×10^6 periods now spreads and continues upwards along the resonances near the borders $h_3 = 0$ and $h_2 = 0$. For the initial condition (vi) chosen at a resonance crossing near $h_2 = 0$, the variation of the unperturbed integrals that was restrained to a small domain in Fig. 4[a] for the lower timescale, now moves upwards, then proceeds to the right and finally goes downwards, the path being drawn by three different resonances. Notice, however, that on this timescale the wandering does not reach the banana-shaped domain on the right corresponding to the initial condition (iii), probably due to the complexity of the dynamics at resonance intersections (recall how intricate such crossings may be). Summing up, on this timescale and for $\varepsilon = 0.02$ the unperturbed integrals still roam over unconnected restraint zones of the energy surface. There remain very localized diffusion domains and the chaotic component seems far from being fully connected.

Instead, for $\varepsilon = 0.04$, after 3×10^8 characteristic periods T of the system the chaotic component is seen to be almost fully connected through the relics of the resonance structure (Fig. 4[d]). The upper right part of the energy surface, formerly empty, now appears densely populated by concentrations mainly along some still distinguishable remaining resonances.
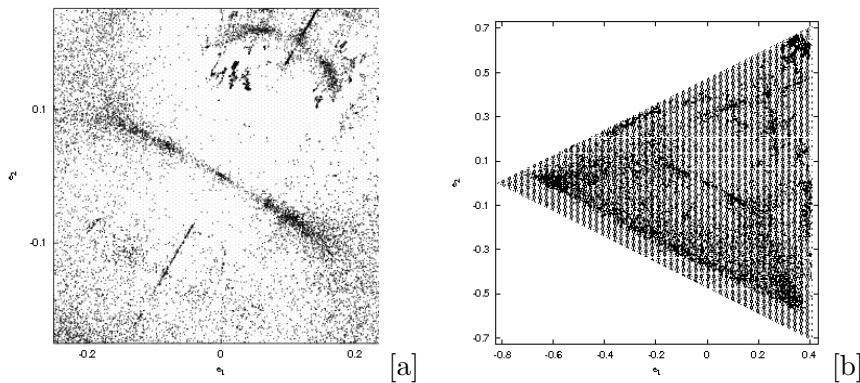

Figure 5.   [a] Enlargement of Fig. 4[d] around $(e_1, e_2) = (0, 0)$. [b] Diffusion of a single orbit over the energy surface for $\varepsilon = 0.04$ and $t_f = 6 \times 10^8 \ T$ (see text for details).

Fig. 5[a] displays an enlargement of the central region in Fig. 4[d], to see whether any unvisited domain still remains in that part of the energy surface for such a large timescale. A strong concentration of points persists along the strip

already present in Fig. 4[b] for 3×10^6 periods, namely, the curve associated to the $(-2, 1, 1)$ resonance.

In view of the results obtained for $\varepsilon = 0.04$ on the larger timescale, we integrate a single orbit, the initial condition (i) taken in the chaotic strip near the border $h_1 = 0$, for an even larger t_f, 6×10^8 characteristic periods T, and trace the path of its unperturbed integrals over the energy surface, with the aim of verifying whether this path covers the whole chaotic component. The results are displayed in Fig. 5[b], and should be compared with the plot in Fig. 4[d] for seven different initial conditions in Table 1 corresponding to chaotic orbits. As seen in the picture, for this single orbit the unperturbed integrals trail over the chaotic component, the relics of resonances serving as routes. It should be mentioned that, for half the total motion time considered, the upper part of the energy surface remained yet unreached, and it took a much longer time to visit the whole chaotic component.

5. Diffusion in multidimensional maps

In this section we present the results concerning the investigation of chaotic diffusion in discrete dynamical systems, most of which were already published in Cincotta & Giordano (2012).

5.1. The Mean Exponential Growth Factor of Nearby Orbits (MEGNO) for maps

For our exploring the action space of a multidimensional map, we wield the generalized MEGNO, for discrete dynamical systems.

The generalized version of the MEGNO, along with its implementation to discrete dynamical systems, which is in order for the current applications, are given also in CGS03. Though, just for the sake of completeness permit us to provide a brief description of how this tool should be applied in the latter case.

For a given map P, the initial point Q_0 is iterated to yield the points $Q_k = P^k(Q_0)$, while the differential map DP transports an initial "random" tangent vector, v_0, $\|v_0\| = 1$, providing vectors $v_k = DP^k(Q_0)v_0$. Then, after N iterates, the generalized MEGNO is computed by means of

$$Y_{m,n}(N) = (m+1) N^n \sum_{k=1}^{N} \ln\left(\frac{\|v_k\|}{\|v_{k-1}\|}\right) k^m, \tag{10}$$

and

$$\overline{Y}_{m,n}(N) = \frac{1}{N^{m+n+1}} \sum_{k=1}^{N} Y_{m,n}(k). \tag{11}$$

Further, a slight additional modification results profitable for the choice $(2, 0)$ of the exponents (m, n), which leads as to the quantity

$$\hat{Y}_{2,0}(N) = \left[4\overline{Y}_{2,0}(N) - 2\right]/N, \tag{12}$$

which tends to 0^- in the regular case and to σ_i, which is the LCN of the orbit, in the case of an irregular trajectory. Negative values of $\hat{Y}_{2,0}(N)$ appear for regular orbits –provided N is taken not too small–, while small positive values identify mild chaos.

5.2. The Coupled Rational Shifted Standard Map (CRSSM)

Let us introduce a 4D conservative map, namely, the Coupled Rational Shifted Standard Map (CRSSM) which provides a fairly good representation of the interaction of two perturbed resonances. Therefore, its dynamics would well serve to model many dynamical scenarios in astronomy and astrophysics.

The CRSSM is defined by:

$$
\begin{aligned}
y_1' &= y_1 + \varepsilon_1\, f_1(x_1) + \gamma_+\, f_3(x_1 + x_2) + \gamma_-\, f_3(x_1 - x_2), \\
y_2' &= y_2 + \varepsilon_2\, f_2(x_2) + \gamma_+\, f_3(x_1 + x_2) - \gamma_-\, f_3(x_1 - x_2), \\
x_1' &= x_1 + \varepsilon_1\, y_1', \\
x_2' &= x_2 + \varepsilon_2\, y_2',
\end{aligned}
\tag{13}
$$

with $x_i \in [0, 2\pi)$, $y_i \in [0, 2\pi/\varepsilon_i)$, $i = 1, 2$, and

$$
f_i(x) = \frac{\sin(x + \varphi_i)}{1 - \mu_i \cos x} - \Delta_i, \qquad \Delta_i = \frac{\mu_i \sin \varphi_i}{\sqrt{1 - \mu_i^2} + 1 - \mu_i^2}, \quad i = 1, 2, 3,
\tag{14}
$$

where $\mu_i \in [0, 1)$, and the quantities Δ_i are fixed so that the f_i functions have zero average.

Notice that (13) is a variation of two coupled standard maps so modified that symmetry is lost through the phase φ_i, and the entire character of f_i is broken due to the parameters $\mu_i \in [0, 1)$. The terms in $(x_1 + x_2)$ and $(x_1 - x_2)$ are coupled by the small terms including γ_+ and γ_- respectively. Thus defined, the map is a 4D–torus.

Though the CRSSM has too many free parameters most of them are kept fixed through all our study, being ε_2 the only one to be varied. We refer to CGS03 for the results concerning different sets of parameters.

Herein we fix $\mu_1 = 0.5$, $\varphi_1 = 1$, $\mu_2 = 0.4$, $\varphi_2 = 2$, $\mu_3 = 0.6$, $\varphi_3 = 3$, $\varepsilon_1 = 0.1$, $\gamma_+ = 0.1$, $\gamma_- = 0.05$ and take two distinct values of $\varepsilon_2 = 0.2, -0.2$.

5.3. Results provided by the MEGNO

For constructing Fig. 6, the MEGNO was computed for an equi-spaced grid of 1000×1000 pixels in the domain $(y_1 \varepsilon_1/2\pi, y_2 \varepsilon_2/2\pi) \in [0, 1) \times [0, 1)$, the initial values for the remaining variables being $x_1 = 0, x_2 = 0$. Recall that this is a problem of higher dimension, so that the iterates under DP of two "random" initial vectors v_{10}, v_{20}, satisfying $\|v_{i0}\| = 1$, had to be computed -plus orthogonalized and renormalized at each step-, and the maximum of the two resulting values of $\hat{Y}_{2,0,\text{rs}}$ –one associated to each direction– was taken to determine the character of each trajectory. The picture on the right corresponds to $\varepsilon_2 = 0.2$ and the one on the left to $\varepsilon_2 = -0.2$.

The contour–like plots exhibit the obtained values for $\log(\hat{Y}_{2,0,\text{rs}})$ after $N = 10^4$ iterations, so that the details in each figure be highlighted. It is interesting
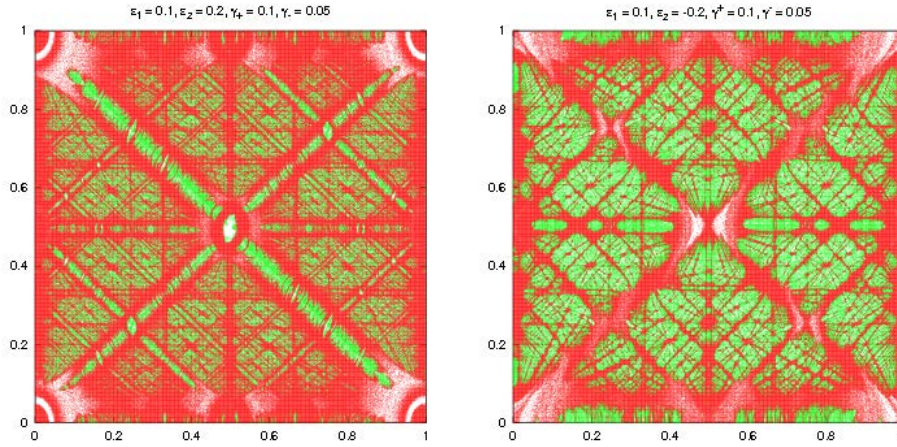
Figure 6. $\hat{Y}_{2,0,\mathrm{rs}}$–levels for the CRSSM on the (y_1, y_2)–plane taking values of ε_2 with opposite signs (0.2 at the left and -0.2 at the right) . Initial conditions of regular behavior are plotted in white, those of mild local instability in green while those exhibiting strong local instability in red.

to remark that in the coupling we face with an indefinite form rather than a positive definite one, which produces a pretty dramatic effect on the resonances. In fact, observe that in the case of positive ε_2 most of the resonances have an elliptical chain of tori at their center while for the negative value of such a parameter several resonances show up as totally hyperbolic. Note that we are dealing with an a priori unstable system (see for instance Lega *et. al*, 2010).

The resonances in the vertical and horizontal directions are present even if $\gamma_\pm = 0$ and their amplitudes depend, essentially, on ε_j, $j = 1, 2$. These resonances appear as white or green channels, despite of the sign of ε_2 and are the same present in the uncoupled Standard Map. Indeed, the sign of ε_i does not modify the uncoupled map, but it certainly affects the coupled one, since rescaling the y-variables, the CRSSM can be recast as

$$
\begin{aligned}
y_1' &= y_1 + \varepsilon_1^2\, f_1(x_1) + \varepsilon_1\gamma_+\, f_3(x_1 + x_2) + \varepsilon_1\gamma_-\, f_3(x_1 - x_2), \\
y_2' &= y_2 + \varepsilon_2^2\, f_2(x_2) + \varepsilon_2\gamma_+\, f_3(x_1 + x_2) - \varepsilon_2\gamma_-\, f_3(x_1 - x_2), \\
x_1' &= x_1 + y_1', \\
x_2' &= x_2 + y_2',
\end{aligned}
\tag{15}
$$

where $(x_i, y_i) \in [0, 2\pi) \times [0, 2\pi)$.

Coupling the degrees of freedom, several new resonances arise, namely, those of the form $\alpha y_1 + \beta y_2 = $ const., with α and β non-vanishing constants. In fact, the map (15) can be derived from the 2.5 degrees of freedom Hamiltonian

$$
\begin{aligned}
H(p_1, p_2, x_1, x_2, t; \varepsilon_1, \varepsilon_2) &= H_0(p_1, p_2, x_1, x_2, t; \varepsilon_1, \varepsilon_2) + \gamma_+ V_p(x_1, x_2, t; \varepsilon_1, \varepsilon_2) \\
&+ \gamma_- V_n(x_1, x_2, t; \varepsilon_1, \varepsilon_2),
\end{aligned}
\tag{16}
$$

where

$$
H_0(p_1, p_2, x_1, x_2, t; \varepsilon_1, \varepsilon_2) = \frac{\left(p_1^2 + p_2^2\right)}{2} + \left(\varepsilon_1^2 U_1(x_1) + \varepsilon_2^2 U_2(x_2)\right) \delta_{2\pi}(t),
$$

$$p_i = \frac{y_i}{2\pi}, \qquad f_i(x_i) = -\frac{1}{2\pi}\frac{dU_i}{dx_i}, \qquad i = 1, 2, \quad (17)$$

being $\delta_{2\pi}(t)$ the 2π–periodic delta defined through its Fourier expansion so that $t \bmod 2\pi$, and the coupling terms can be written as

$$V_p(x_1, x_2, t; \varepsilon_1, \varepsilon_2) = (\varepsilon_1 + \varepsilon_2)U_3(x_1 + x_2)\delta_{2\pi}(t),$$
$$V_n(x_1, x_2, t; \varepsilon_1, \varepsilon_2) = (\varepsilon_1 - \varepsilon_2)U_3(x_1 - x_2)\delta_{2\pi}(t). \qquad (18)$$

The relationship between f_3 and U_3 is similar to that defined above involving f_i and U_i for $i = 1, 2$. Since every function of t, x_1, x_2 id periodic, its Fourier expansion generates all kinds of resonances that, in action space, appear as straight lines as already mentioned. Indeed, the unperturbed frequencies, -that is when $\varepsilon_i = \gamma_i = 0$-, are $(\omega_1, \omega_2) = (y_1, y_2)$.
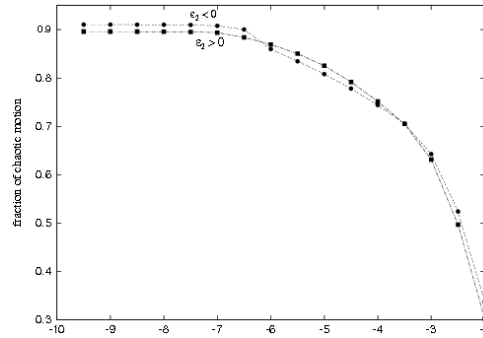
Figure 7.   Fraction of chaotic motion in the maps corresponding to $\varepsilon_2 = 0.2$ and $\varepsilon_2 = -0.2$ against different threshold values of $\log(\hat{Y}_{2,0,\mathrm{rs}})$.

Let us now refer to Fig. 7 which displays the fraction of chaotic motion for the maps corresponding to the two opposite values of ε_2, for different thresholds of $\log(\hat{Y}_{2,0,\mathrm{rs}})$ below which global motion is regarded as regular. From this figure it could be deduced that $\log(\hat{Y}_{2,0,\mathrm{rs}}) \approx -6.5$ is a fairly adequate value to be adopted as a threshold in order to separate regular from chaotic dynamics. Notice that, for both values of ε_2, action space is almost completely chaotic – only about its 10% corresponding to regular motion– so it can be claimed that global chaos has set up in the system.

5.4.   Diffusion on action space

With the aim of gaining some insight on the way diffusion operates, several orbits with initial conditions embedded in different chaotic domains were traced onto action space for both values of ε_2.

The wandering of the unperturbed integrals on the (y_1, y_2)–plane was pursued during 10^7 iterations for the initial conditions listed in Table 2, which, not only are located near the crossing of resonances, but have high values of the MEGNO, except for the initial conditions (iv) for $\varepsilon_2 = 0.2$ and (i) for $\varepsilon_2 = -0.2$, which correspond to stable motion.

As shown in Fig. 8 (corresponding to $\varepsilon_2 = 0.2$), the unperturbed integrals remain confined to rather small domains, so that diffusion turns out to be inefficient for a rather large number of iterates, as we observed in the case of the

Table 2. Initial conditions on action space in units of $2\pi/\varepsilon_j, j = 1, 2$ (the same as in Fig.6) for five different orbits and its concomitant value of $\log(\hat{Y}_{2,0,\text{rs}})$ for $\varepsilon_2 = 0.2$ (fourth column) and $\varepsilon_2 = -0.2$ (last column) after 10^4 iterations.

i.c.	y_1	y_2	$\log(\hat{Y}_{2,0,\text{rs}})$ $\varepsilon_2 = 0.2$	$\log(\hat{Y}_{2,0,\text{rs}})$ $\varepsilon_2 = -0.2$
(i)	0.000	0.110	-2.435	-10.000
(ii)	0.000	0.500	-1.584	-1.436
(iii)	0.331	0.353	-1.712	-1.486
(iv)	0.500	0.500	-10.000	-1.449
(v)	0.074	0.336	-1.824	-1.794
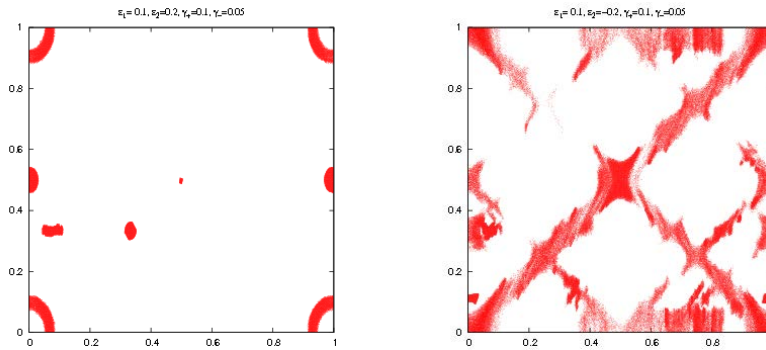
Figure 8. Diffusion on action space after 10^7 iterations of the map for the initial conditions listed in Table 2.

Hamiltonian system for the smaller perturbation parameter. This behaviour is usually referred to as stable chaos. Several papers due to Milani and co-workers (Milani & Nobili, 1992); (Milani, 1993); (Milani & Farinella, 1994); (Milani *et al.*, 1997) also address this topic. Moreover, further studies deal with the same phenomena of stable chaos (Morbidelli & Froeschlé 1995), which arises, for instance, in Solar System dynamics.

On the other hand, the plot in the panel corresponding to $\varepsilon_2 = -0.2$ exhibits the significant efficacy of the diffusive process (except for the initial condition (i) defining a stable orbit), and evinces that the relics of the unperturbed resonances serve as paths for diffusion. This mechanism is termed Arnol'd diffusion–like process and should not be confused with Arnol'd diffusion. In fact, this is what we have already observed in the case of the perturbed quartic oscillator for the largest value of the perturbation parameter. A detailed discussion about this issue and its departure from actual Arnol'd diffusion could be found in Giordano & Cincotta (2004); Cachucho *et al.* (2010) as well as in CGS03. Nonetheless, let us stress that the observed diffusion presents a geometrical resemblance with the Arnol'd theoretical conjecture according to which diffusion proceeds on phase space through the chaotic layers of the resonance web.
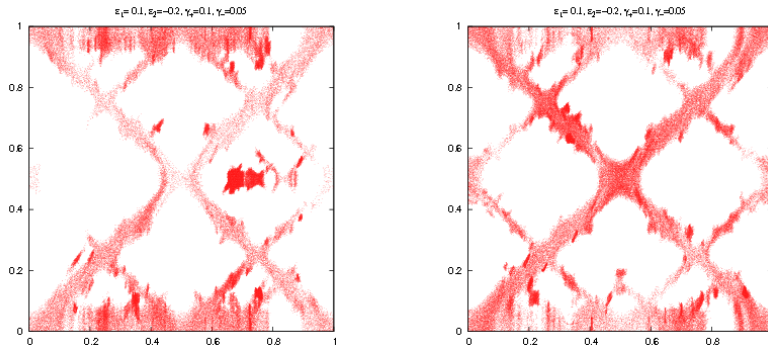
Figure 9.   Diffusion on action space after 2×10^8 iterations of the map corresponding to $\varepsilon_2 = -0.2$ for i.c. (ii) in [a] and (iv) in [b].

Moreover, Fig. 9 illustrates that the orbits with initial conditions (ii) and (iv) for the case $\varepsilon_2 = -0.2$ sweep a rather large fraction of action space after 2×10^8 iterations. However, notice must be taken that the chaotic component of phase space is not fully connected for this timescale, since some chaotic domains remain unreached.

Fig. 10 depicts the time evolution of the initial condition (iv) after different iterations of the map in the case $\varepsilon = -0.2$. For 5×10^6 iterates the orbit remains confined to the central region while it spreads along some resonances for 10^8 iterates of the map. During the last 10^9 iterations the orbit explores several additional resonance remnants, though it is still far from visiting the entire chaotic component (compare with Fig. 6). Furthermore, for $N = 2 \times 10^9$, the trajectory seems to remain trapped in the horizontal resonance at $y_2 = 0.5$ (which has an elliptic structure) and no significant difference between the domains swept in the two last intervals of time is observed. These plots in fact could be considered as time-contour diagrams, since the more crowded regions should be associated to those domains of action space where the orbit spends more time.

5.5.   A measure of diffusion

The classical approach for studying diffusive processes associated with the variation of any observable involves the analysis of the evolution of its mean square displacement. Certainly, the type of diffusion most studied is the normal one, mainly characterized by the linear scaling of the mean square displacement with time. Though, deviations from normal diffusion are frequently observed in many dynamical systems (see for instance Zhou *et al.* 2002; Cordeiro & Mendes de Souza 2005; Cordeiro 2006. This phenomena, termed anomalous diffusion in Metzler & Klafter (2000), affords also a characterization through the scaling of the variance (or the mean square displacement) with time but of a more general form.

Herein, we undergo the computation of a diffusion coefficient under the framework of normal diffusion in a rather simple fashion -following Chirikov (1979)-.
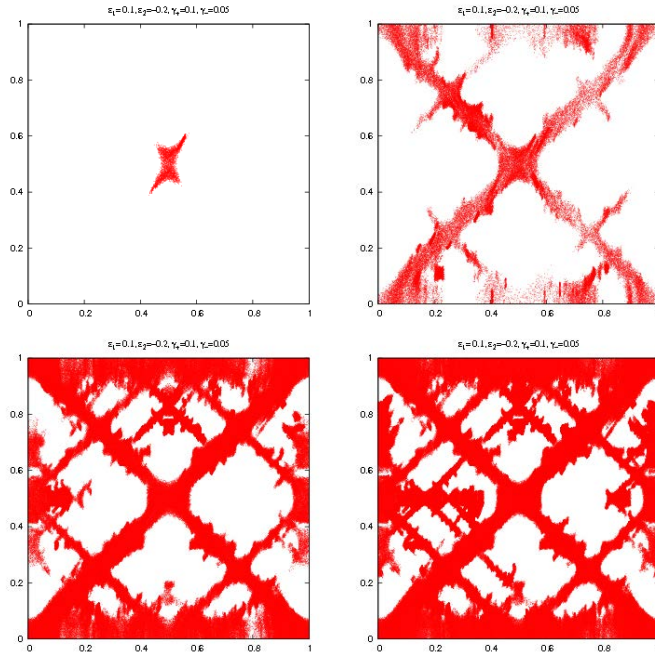
Figure 10.   Diffusion after 5×10^6, 10^8, 10^9, 2×10^9 for a single orbit starting at the center for $\varepsilon = -0.2$.

Instead of tracing the evolution of a single orbit of the map, a diffusion coefficient for an ensemble of nearby orbits could be computed. To this end, and following the standard procedure in the case of random walk processes, let us define the finite time diffusion coefficient as the mean-square spread in y_i as (see Meiss, 1992)

$$\hat{D}_i = \frac{\text{Var}[y_i]}{N}, \tag{19}$$

where $\text{Var}[y_i]$ stands for the variance of the action and N for the entire time or number of iterates, so that the mean square span in each unperturbed action is given by

$$\sigma_i[y_i] = \sqrt{N \; \hat{D}_i}. \tag{20}$$

Finally, we introduce the quantity

$$D = \frac{\hat{D}_1 + \hat{D}_2}{2} \tag{21}$$

as a measure of the normal diffusion on the unperturbed actions–plane.

Recall that the map (13) is defined on a torus so, in order to avoid border effects that would artificially enlarge the value of the y_i–variance, the new variables

$$z_i = \cos(2\pi y_i), \qquad i = 1, 2. \tag{22}$$

are introduced, and the diffusion measure D is computed on the (z_1, z_2)–space instead of on the actual action space (y_1, y_2).

The roam of the unperturbed integrals onto the (y_1, y_2)–plane for five bundles around the initial conditions in Table 3 is depicted in Fig. 11, but, as already stated, it is the wandering onto the (z_1, z_2)–plane, shown in Fig. 12, the one to be used to compute the coefficient D defined by (21). For each one of these figures, an ensemble of 100 initial conditions chosen at random in a neighborhood of size 10^{-7} of conditions (i) to (v) were followed during 10^7 iterations of the map for both values of ε_2.

Furthermore, the averaged value of the diffusion for each ensemble was computed after 10^7 iterations. The concomitant out-coming values of D are displayed in Table 3.

Table 3.   Initial conditions on action space in units of $2\pi/\varepsilon_j, j = 1, 2$ (the same as in Fig.6) for five different bundles of orbits and its concomitant value of $\log(D)$ for $\varepsilon_2 = 0.2$ (fourth column) and $\varepsilon_2 = -0.2$ (last column).

i.c.	y_1	y_2	$\log(D)$	$\log(D)$
			$\varepsilon_2 = 0.2$	$\varepsilon_2 = -0.2$
(i)	0.000	0.110	-8.666	-10.145
(ii)	0.000	0.500	-11.093	-7.438
(iii)	0.331	0.353	-9.027	-7.050
(iv)	0.500	0.500	-13.033	-7.993
(v)	0.074	0.336	-9.512	-8.034

Let us notice that the averaged values obtained for D fairly succeed in yielding a measure of the domain covered by the wandering of each bundle of initial conditions, displayed in Figs. 12.
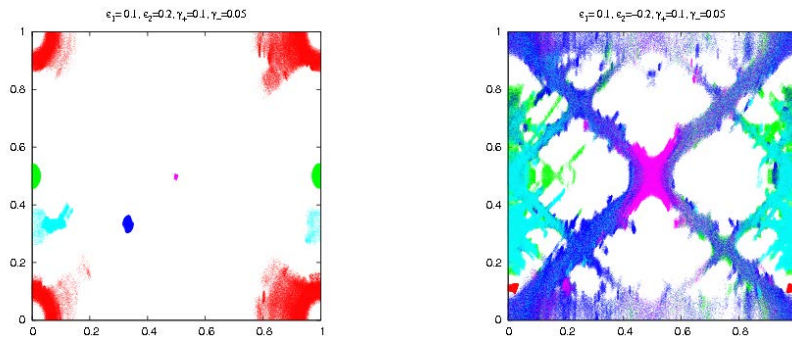

Figure 11.   Diffusion on action space after 10^7 iterations of the map for the bundles of 100 initial conditions chosen at random around initial conditions listed in Table 3. Initial condition (*i*) in red, (*ii*) in green, (*iii*) in blue, (*iv*) in purple and (*v*) in light–blue.

In fact, for $\varepsilon = 0.2$ the lowest value of D corresponds to regular orbits at (iv), whose roam is practically restrained to a point being the motion essentially stable. For the remaining initial conditions, the out-coming value of D has the
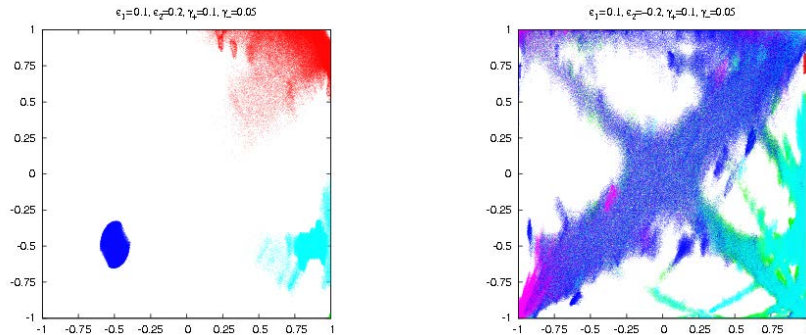
Figure 12.   Diffusion on (z_1, z_2)–space after 10^7 iterations of the map for the bundles of 100 initial conditions chosen at random around initial conditions listed in Table 3, using the same color scheme.

same order of magnitude, the unperturbed integrals remaining confined to rather small domains of comparable extent for the number of iterates considered.

This fact can be confirmed when plotting the values of D for every orbit in each bundle as in Fig. 13, in which the last point in each curve corresponds to the averaged value for the ensemble.
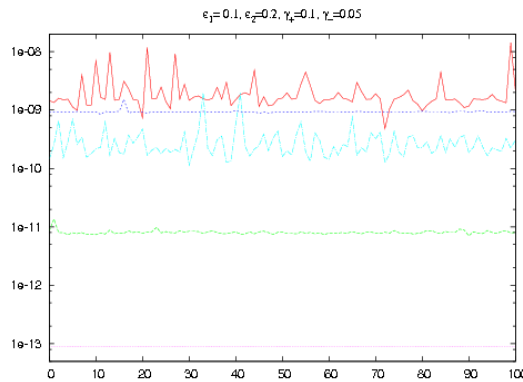
Figure 13.   Diffusion measure on action space after 10^7 iterations of the map for $\varepsilon_2 = 0.2$ for a random selection of 100 initial conditions around (i) in red, (ii) in green, (iii) in blue, (iv) in purple, and (v) in light-blue.

The same procedure leads to Fig. 14 corresponding to $\varepsilon_2 = -0.2$. In particular this plot reveals the large dispersion in the values of D for each bundle, which gives account of the rather extended area swept by every bundle onto the (z_1, z_2)–plane.

Certainly, we could not invoke normal diffusion when dealing with systems of divided phase space. In fact, in such cases we ignore how the variance of the actions scale with time. Furthermore, though theories on abnormal diffusion and on sub–diffusion or super–diffusion have been developed, they seem to apply only for low dimensional systems.
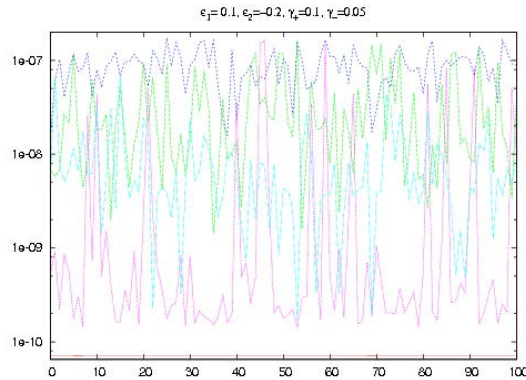
Figure 14.   Diffusion measure on action space after 10^7 iterations of the map for $\varepsilon_2 = -0.2$ for a random selection of 100 initial conditions around (i) in red, (ii) in green, (iii) in blue, (iv) in purple, and (v) in light-blue.

Thus, the computation of a diffusion coefficient reveals itself as a cumbersome task. For instance, in Fig. 14 we observe fluctuations of three orders of magnitude in the 100 values of D within a domain of 10^{-7} around some of the chosen initial conditions.

Therefrom, in the forthcoming section we go beyond in the search of an alternative tool to measure diffusion bringing into play the finite time Shannon or Arnol'd Entropy.

5.6.   The Shannon Entropy

Let us consider the time evolution of both normalized unperturbed actions $(y_1\varepsilon_1/2\pi, y_2\varepsilon_2/2\pi)$ upon the 2–torus T, or the unit square with opposite sides identified, as time goes by.

The evolution of the unperturbed integrals onto T is expected to be constrained to a small domain whenever a low diffusion takes place, while a large roam should be in order in case of fast diffusion.

An adequate tool to measure diffusion is the finite time *Shannon Entropy*. A theoretical background on Shannon Entropy can be found in e.g Shannon & Weaver (1949); Katz (1967); Arnol'd & Avez (1989); Wehrl (1978).

Thus, the main idea of the present approach is to make use of the Shannon Entropy -SE hereafter- in order to measure the wandering of the unperturbed action variables when viewed on the unit square.

Following Arnol'd & Avez (1989), let us recall the definition of the function Z over $[0, 1]$:

$$
\begin{aligned}
Z(x) &= -x\ln x,\ x \in (0, 1] \\
Z(0) &= 0,
\end{aligned}
$$

which is continuous, nonnegative and strictly concave ($Z'' < 0$) and $Z(x) = 0$ corresponds to either $x = 0$ or $x = 1$.

Further, let $\alpha = \{a_i;\ i = 1, \cdots, q\}$ be a partition of T, that is to say, a collection of q boxes that cover the whole unit square. The boxes are assumed

to be both measurable and disjoint, that is to say

$$\mu\left(T - \cup_i a_i\right) = 0, \qquad \mu\left(a_i \cap a_j\right) = 0, \;\; i \neq j,$$

where μ stands for a given measure.

We can define a probability density on T by

$$\varrho(\mathrm{x}) = \frac{1}{N}\sum_{i=1}^{N}\delta(\mathrm{x} - \mathrm{x}_i), \tag{23}$$

where x_i denotes the result of a given iterate on T, and δ is the *delta function*. It can be verified at once that

$$\int_T \varrho(\mathrm{x})\;d\mathrm{x} = 1, \tag{24}$$

and the probability of the element a_i of the partition turns out to be

$$\mu(a_i) = \int_{a_i} \varrho(\mathrm{x})\;d\mathrm{x}. \tag{25}$$

Finally, the entropy of the partition α is defined as

$$S(\alpha) = \sum_{i=1}^{q} Z\left(\mu(a_i)\right) = -\sum_{i=1}^{q}\mu(a_i)\ln(\mu(a_i)). \tag{26}$$

Let us notice that for a given partition α, the entropy is a bounded quantity. In fact, it is $0 \leq S(\alpha) \leq \ln q$. The minimum value is reached when all points fall in the very same element of the partition, say the k–th element, which would correspond to the case of almost full stability (for instance when all the iterates lead to the same action value). Indeed, in such a case, we have a unique nonzero value $\mu(a_k) = 1$, yielding $S = 0$. On the other hand, the maximum value, $S(\alpha) = \ln q$, will be reached whenever the q elements of the partition have equal measure $\mu(a_i) = 1/q$, that corresponds to the situation in which the unperturbed actions wander all over the unit square in a uniform fashion, i.e. in case of ergodicity.

Hereafter, S will denote the normalized value of the entropy (i.e. the entropy divided by $\ln q$) for a given partition α.

Note that if we assume all the non-empty elements of the partition, say q_0, having the same measure, then $S \approx \ln q_0 / \ln q$.

Let us now accomplish the computation of the finite time SE for a given set of initial conditions for the CRSSM (13) with the same set of parameters.

To this end, we take a partition of $q = m \times m$ bidimensional boxes that cover the whole unit square, for an equi–spaced grid of 1000×1000 initial conditions in the domain $(y_1\varepsilon_1/2\pi, y_2\varepsilon_2/2\pi) \in [0,1) \times [0,1)$. The initial values for the remaining variables are $x_1 = 0, x_2 = 0$, and we consider $N = 10^7$.

Let us say that we have defined the partition in regards to the total number of iterates such that in case of a uniform distribution the same number of points
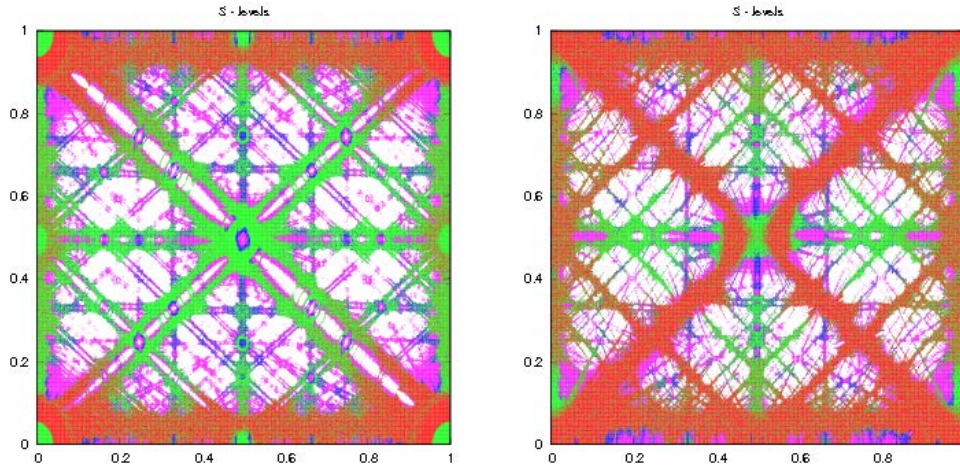
Figure 15.   S–levels for the CRSSM on the (y_1, y_2)–plane after $N = 10^7$ iterates of the map for a grid of 1000×1000 initial conditions of the unperturbed actions for ε_2 with opposite signs. Values of $S < 0.38$ are depicted in white, $0.38 \leq S < 0.45$ in purple, $0.45 \leq S < 0.5$ in blue, $0.5 \leq S < 0.66$ in green and $0.66 \leq S$ in red. In terms of the number of visited elements of the partition, q_0, the intervals correspond to $q_0 < 2 \times 10^2$ in white, $2 \times 10^2 \leq q_0 < 5 \times 10^2$ in purple, $5 \times 10^2 \leq q_0 < 10^3$ in blue, $10^3 \leq q_0 < 10^4$ in green and $10^4 \leq q_0$ in red.

would lie in each individual cell. The results are given in Fig. 15, the plot on the left corresponding to $\varepsilon = 0.2$ and that on the right to $\varepsilon = -0.2$.

Let us remark that the results obtained by means of the SE are also in quite good agreement with those obtained by recourse to the MEGNO, except in those regions of slow or fast diffusion.

Let us note that the resonance structure arising in the contour–plots obtained with the SE resembles the one revealed by the MEGNO. Moreover, the effect on the resonances due to the change of sign of ε_2 remains noticeable. Indeed, as already pointed out, in the case of positive ε_2 most of the resonances display an elliptical chain of tori at their center while for negative ε_2 several resonances show up as totally hyperbolic. However, while the MEGNO just measures the local hyperbolicity of a certain point in phase space, the SE provides information about chaotic diffusion in such a point. This can be clearly seen, for instance in the case for $\varepsilon_2 = -0.2$, by the small values adopted by the SE along the resonance at $y_2 = 0.5$, whose structure seems to act as a barrier to diffusion as we have already shown.

Although the SE depends on the partition, in our experiments no significant differences have been observed when using different α's.

While for $N \to \infty$ the SE (in phase space) should become the metric entropy -which is related to the sum of the positive Lyapunov's exponents- this seems not to be the case for any finite, though very large, number of iterates N.

6.   Theoretical considerations, observational constraints and open problems

In this section we summarize some theoretical considerations, observational data and results from N–body simulations and finally some open questions.

6.1.   Theoretical considerations

- The overlap of resonances (heteroclinic intersections) and Arnol'd diffusion–like processes are well–known mechanisms that lead to the transition from regularity to gross chaos (or gross instabilities).

- We claim that "classical" Arnol'd diffusion does not play any role in galactic dynamics (even in asteroidal dynamics), since in general "the perturbation" is not small enough, and even if it might work, its timescale is quite large for any real system, as shown in our simple experiments. We do not believe that Arnol'd diffusion could describe global instabilities.

- Though one could get accurate values of any indicator of the stability of the motion, they only provide the local rate of exponential divergence and this is not related to any global instability.

- Chaos or global instabilities should be understood as large variations of the unperturbed integrals (diffusion). Unfortunately, as far as we know, it does not yet exist any theory that could describe global diffusion (instability) in phase space.

- However what it is actually relevant is the extent of the domain upon which the unperturbed integrals change and, physically, the timescale in which this diffusion occur.

- Those timescales for galaxies which are less or similar than the Hubble time $T_H \sim 10^3 T_c$, being T_c the characteristic timescale for a given galaxy, seem to be too short in order to any kind of global diffusion be efficient. Perhaps in some cases, the overlap of resonances, whose rate is $\sim$ some power of the perturbation parameter could lead to some diffusion, connecting nearby regions of phase space.

- Arnol'd diffusion, or to be more precise, Arnol'd mechanism, only states that two points of phase space separated by a distance of $\boldsymbol{O}(1)$ could be connected. This result does not imply any global instability and, in fact, just for this case it requires exponentially large times. Thus, we believe that Arnol'd diffusion does not play any role in galactic dynamics.

- Arnol'd diffusion–like processes perhaps also could not operate in galactic dynamics, as we have already shown. It might be a plausible mechanism in asteroidal dynamics though.

6.2.   Observational data and N–body simulations

- Observations with the HST revealed the presence of very high stellar densities at the center of early-type galaxies, which suggest that rhey could

be fitted by a power law of the form $(r^{-\gamma})$. The evidence of large central masses was also reinforced from high-resolution kinematical studies of nuclear stars and gas, which disclosed the presence of compact dark objects with masses in the range of $10^{6.5} - 10^{9.5} M_\odot$, presumably super–massive black holes. These observational results have produced a substantial change in the classic ideas on dynamics in triaxial galaxies.

- Numerical simulations show that the addition of a central mass to an integrable triaxial potential has deep effects on its dynamics, at least for the boxlike orbits which mainly cover the central region of triaxial galaxies.

- Black holes and central density cusps scatter these particular orbits during each close passage giving rise to chaos in the system. The sensitivity of boxlike orbits to deflexions also drives a rounder central distribution of mass. This slow evolution towards axisymmetry suggests that stationary triaxial configurations could not exist for a strong central density cusp.

- For such large values of M_{bh}, the box orbit phase space is almost completely stochastic since overlap of resonances and diffusive processes could take place in rather short timescales.

- This result turned out to be substantially attractive because this critical black hole mass was close to the one observed and also close to the mass which induced a sudden evolution towards the axisymmetry in N–body simulations.

- Merritt & Fridman (1996) arrived to similar conclusions analyzing two triaxial power law models $r^{-\gamma}$: the steep ($\gamma = 2$) and the weak ($\gamma = 1$) cusp. They found, in agreement with Gerhard & Binney (1985), that triaxial galaxies with such a huge concentration of mass would evolve towards a central axisymmetry, as box orbits loose their distinguishability.

- For these models, in which a large fraction of phase space is dominated by a chaotic dynamics, the construction of self-consistent solutions requires the inclusion of stochastic orbits besides the regular ones. A system thus built evolves, mainly close to its center, as stochastic orbits mix through phase space.

- Though it is possible to build this kind of solutions for a weak cusp model, this is not the case for a strongly concentrated model. This would imply that triaxility is not consistent with a high central density.

- How could the correlation between the velocity dispersion and the black holes masses be explained?

6.3.   Open problems

- Nature's ability to build stationary non–axisymetric stellar systems is still en open matter.
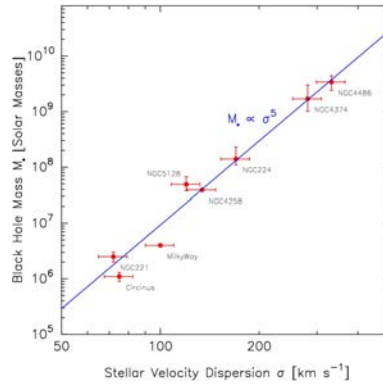
Figure 16.   Correlation between the velocity dispersion and the mass of the (assumed) black hole in different galaxies. (This figure was generously provided by D. Merritt).

- Merritt's generalization of Jeans Theorem rests on a very strong assumption: a completely connected chaotic component must exist in a 3D system with divided phase space. It seems that this could happen only whether the chaotic component has a large measure and "$t \to \infty$", which, from a physical point of view, it would not be possible in galactic systems, where the chaotic component has a small measure and $t \lesssim T_H$.

- An important fact to be stated is that when chaos sets up, the unperturbed global integrals (or actions) have a discontinuous dependence on phase space variables. Indeed, close to resonant tori, despite the existence of three local integrals, the unperturbed orbital structure is not preserved and the topology of phase space changes. Moreover, at least one integral does not exist on the stochastic layer.

- Close to strong non–resonant tori, those satisfying the Diophantine condition, local integrals are just corrections of order ε of the unperturbed global integrals. On the other hand, when the system is close to an elliptic resonant torus, new local integrals appear: the pendulum Hamiltonian H_r and linear combinations, K_2, K_3, of the unperturbed actions at the resonant point.

- It is not possible to assume that the DF has the form $f(H, I_2, I_3)$ in the whole regular component. This could be true only for strong non–resonant tori, but since resonances are dense in phase space, the DF should be locally defined as: $f_n(H, I_2, I_3) + \varepsilon g(H, I_2, I_3)$ in a neighborhood of a non–resonant torus and $f_r(H_r, K_2, K_3)$ in a vicinity of an elliptical resonant torus.

- Is the distribution function of the galaxy best fitted by a two–(local)integral or three-(local)integral model?

- Nothing could be said about f in the chaotic domain since there is no theoretical support to argue that the whole chaotic region is fully connected. Clearly, a notorious discontinuous dependence of f on the integrals is expected.

- Does the introduction of a black hole ("singularity") at the origin change the approach to the problem? Indeed, if we perform a multipolar expansion of the potential for an elliptical galaxy, the introduction of a mass point perturbation would drastically modify the unperturbed part of the Hamiltonian: that corresponding to the monopole term.

- How should diffusion be measured? It seems natural to think that it is related with the variance of the integrals. But in which way?

- How should the existence of barriers and "accelerators" of diffusion be included in the computation of the coefficient?

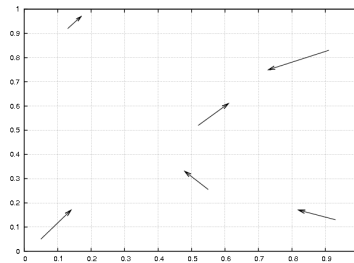- And finally, how could diffusion routes be predicted?

Figure 17.   The length of the arrows would measure the rate of diffusion at that point in phase space and their direction points out its route after some time interval.

References

Arnol'd, V. I., 1964, Sov. Math.-Dokl. 5, p. 581.

Arnol'd, V. & Avez, A., 1989, Ergodic Problems of Classical Mechanics, 2nd. Ed. (Addison-Wesley, New York).

Binney, J. & Tremaine, S., 1987, Galactic Dynamics, (Princeton University Press: New Jersey).

Cachucho, F., Cincotta, P., Ferraz–Mello, S., 2010, CMDA, 108, 35.

Carati, A & Galgani, L., 2012, Proc. of the Third La Plata School on Astronomy and Geophysics, Cincotta, P., Giordano, C., & Efthymiopoulos, C. (eds.), (La Plata), this volume.

Chirikov, B.V., 1979, Physics Reports 52, 263.

Chirikov, B., Lieberman, M., Shepelyansky, D., Vivaldi, F., 1985, Physica D, 14, 289.

Chirikov, B. & Vecheslavov, V., 1989, Preprint INP, 89, Novosibirsk.

Chirikov, B. & Vecheslavov, V., 1993, J. Stat. Phys. 71, 243.

Cincotta, P.M. & Simó, C., 2000, A&A, 147, 205.

Cincotta, P.M., 2002, New Astronomy Reviews, 46, 13.

Cincotta, P.M. & Giordano, C.M., 2002, Proc. Advances in Space Dynamics 2. Applications in Astronomy, Winter O., Prado A.F.B.A, (eds.), (I.N.P.E., São Paulo), 237.

Cincotta, P.M., Giordano, C.M. & Simó, C., 2003, Physica D, 182, 151 (CGS03).

Cincotta, P.M. & Giordano, C.M. 2008, New Nonlinear Phenomena Research, (New York: Nova Science Pub. Inc., Perlidze T. ed.), Ch. 6, 393.

Cincotta, P.M. & Giordano, C.M., 2012, IJBC, in press.

Cordeiro, R. & Mendes de Souza, L., 2005, AJ, 439, 375.

Cordeiro, R., 2006, AJ, 132, 2114.

Efthymiopoulos, C., Voglis, N. & Kalapotharakos. C., 2007, LNP, 729, 297.

Efthymiopoulos, C., 2012, Proc. of the Third La Plata School on Astronomy and Geophysics, Cincotta, P., Giordano, C., & Efthymiopoulos, C. (eds.), (La Plata), this volume.

Froeschlé, C., 1971, Astrophys. Space Sci. 15, 110.

Froeschlé, C., 1972, å, 16, 172.

Froeschlé, C., Guzzo, M. & Lega, E., 2005, CMDA, 92, 243.

Gerhard, O.E. & Binney, J.J., 1985, MNRAS, 216, 467.

Giordano, C. M. & Cincotta, P., 2004, A&A, 423, 745.

Guzzo, M., Lega E. & Froeschlé, C., 2002, Physica D, 163, 1.

Guzzo, M., Lega, E., & Froeschlé, C., 2005, DCDS B 5, 687.

Guzzo, M., Lega, E. & Froesché, C., 2006, Nonlinearity, 19, 1049.

Katz, A., 1967, Principles of Statistical Mechanics, The Information Theory Approach, (San Francisco: W.H. Freeman & Co.).

Hairer, E., Norsett, S. & Wanner, G., 1987, Solving Ordinary Differential Equations. I. Nonstiff Problems, (Springer: New York).

Korabel, N. & Klages, R., 2004, Physica D, 187, 66.

Lega, E., Froeschlé, C. & Guzzo, M., 2008, LNP, 729, 29.

Lega, E., Guzzo, M. & Froeschlé, C., 2010, CMDA, 107, 115.

Lochak, P., 1999, Proc. Hamiltonian Systems with Three or More Degrees of Freedom', Simó C. (ed.), (Kluwer, Dordrecht), 168.

Lynden-Bell, D., 1962, MNRAS, 124, 1.

Mestre, M., 2012. Difusión Caótica de Sistemas Hamiltonianos casi–integrables, Ph.D Thesis. Universidad Nacional de La Plata.

Merritt D. & Friedman T., 1996, ApJ, 460, 136.

Merritt D. & Valluri M., 1996, ApJ, 471, 82.

Merritt, D., 1999, PASP, 111, 129.

Meiss, J., 1992, Rev. Mod. Phys. 64, 795.

Mestre, M, Cincotta, P. & Giordano, C., 2009. IJNLM, 44, 179.

Metzler, R. & Klafter, J., 2000, Physics Reports, 339, 1.

Milani, A. & Nobili, A. M., 1992, Nat, 357, 569.

Milani, A., 1993, CMDA, 57, 59.

Milani, A. & Farinella, P., 1994, Nat, 370, 40.

Milani, A., Nobili, A. M. & Knezevic, Z., 1997, Icarus, 125, 13.

Morbidelli, A. & Froeschlé, C., 1996, CMDA, 63, 227.

Muzzio, J.C., Carpintero, D.D., & Wachlin, F.C., 2005, CMDA, 91, 173.

Papaphilippou Y. & Laskar J., 1998, A&A, 329, 451.

Poincaré, H., 1893, Les Méthodes Nouvelles de la Mécanique Céleste, (Gautier–Villars: Paris).

Poon, M. & Merritt, D., 2002, ApJ, 568, 89.

Shannon, C. & Weaver, W., 1949, The Mathematical Theory of Communication, (Urbana: Illinois U.P.).

Venegeroles, 2008, Phys. Rev. Lett., 101, 54102.

Wehrl, A., 1978, Rev. Mod. Phys., 50, 221.

Zhou, J., Sun, Y. & Zhou L., 2002, CMDA 84, 409.

LECTURE

Dynamical chaos in the Solar system

Ivan I. Shevchenko

Pulkovo Observatory of the Russian Academy of Sciences,
St. Petersburg, Russia

Abstract. Methods and results of a new major part of science on the dynamics of the Solar system bodies, namely, the part devoted to researches of dynamical chaos in the motion of celestial bodies, are described and analyzed. The dynamical chaos (non-determined dynamical behavior in the absence of any random perturbations) in the motion of celestial bodies is caused by interaction of nonlinear resonances, either orbital or spin-orbit. The resonances and chaos in the motion of minor and large bodies of the Solar system, — planetary satellites, asteroids, comets, and planets, are considered. Special attention is given to the problem of observability of the chaotic behavior; in particular, methods of analytical estimating the Lyapunov time, specifying the "predictability horizon" of the motion, are described.

1. Introduction

Within many centuries nothing had seemed to humankind to be farther from chaos and accident than the repeated, measured motion of the Solar system bodies. Integer relations between the orbital periods of some of them, first discovered in the 18th century, seemed to be an obvious confirmation for the reigning order and harmony. However, in sharp contrast with this seeming evidence, these were just such relations (named *commensurabilities* or *resonances*) that appeared to be the cause of *dynamical chaos* — the phenomenon not at all related to any random perturbations on the moving bodies, but concealed in the very nature of the motion.

The idea of harmonious order based on geometrical relations inspired Johannes Kepler in his first researches of the motion of planets. In 1593 he noticed, that the radii of the circle inscribed in an equilateral triangle and the circle described around it have approximately the same ratio as the radii of the orbits of Jupiter and Saturn (Fig. 1). Under impression of this observation he developed a model of the Solar system as a concentric sequence of five principal polyhedra. The Solar system in this model, named "Kepler's goblet" (Fig. 2), obeyed to a certain static geometrical order.

In 1784 Pierre Simon Laplace paid attention to another, not geometrical but dynamical, relation between the orbits of Jupiter and Saturn: the orbital periods of these planets are close to integer commensurability (resonance) 2/5. Taking this relation into account, he explained the observed anomalies in the motion of Jupiter and Saturn and revealed their periodic (instead of monotonous
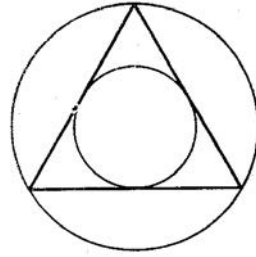
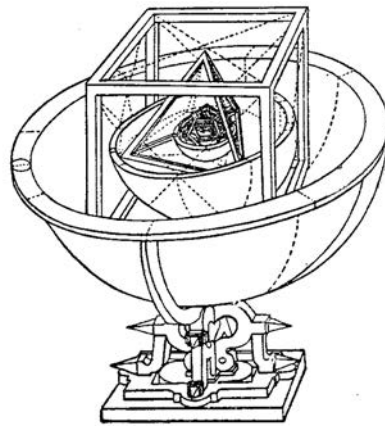Figure 1.   The orbits of Jupiter and Saturn, and Kepler's triangle.

Figure 2.   "Kepler's goblet".

— *secular*) character. This triumph of celestial mechanics, alongside with other outstanding achievements (in particular, with Alexis Claude Clairaut's rather exact prediction of the date of returning of the Halley comet in 1759), allowed Laplace to put forward an idea that the world is completely predetermined: if, at any moment of time, the positions and velocities of all the world-comprising particles are set, then the world's subsequent history is predetermined. As an embodiment of this idea, a mechanical device — an *orrery* — can serve, representing the motions and phases of the Solar system planets and their satellites by means of a clockwork. As A. Pannekoek (1961) wrote, "the Solar system was thought to be a huge mechanism brought and pushed in the motion solely by the force of universal gravitation. It was a fully cognizable and calculable clockwork which kept the motion forever".

Thus, at the end of the 18th century the consideration of a *single resonance* in Laplace's theory resulted in the formulation and general scientific recognition of determinism. The full predefiniteness of the motion of both large and minor bodies of he Solar system was not put into question. It was believed that the more accurate are observations and the more perfect is the theory, then the greater is the time interval for which it is possible to predict the motion

of any celestial body. It is paradoxical, but two centuries later an analysis of *interaction of resonances* in works by Boris Chirikov, Jack Wisdom and other scientists destroyed the deterministic concept. The refusal of it was promoted, in particular, by returning of the Halley comet (the third returning after that predicted by Clairaut) in 1986: now its orbit was considered not as an example of a completely calculable predetermined motion but as an example of manifestation of *dynamical chaos*. Thus the Laplace determinism had existed on the time span of three orbital revolutions of the Halley comet. It is curious that even in the days when the Laplace determinism dominated in the scientific thought absolutely, the educated community as a whole did hardly perceive the cosmic world as an ideally predictable mechanism: e. g. the apparitions of comets were generally perceived as sudden and random events.

2. Pendulum, resonances and chaos

Resonance represents the central concept of nonlinear dynamics. Chirikov (1982) defines it so: "Resonance is understood as such situation when some frequencies of a non-perturbed system are close to each other or to frequencies of an external perturbation". How one can be convinced in the existence of resonance in the motion of any particular celestial bodies? In fact, any observed ratio of frequencies can be approximated by some rational number with any degree of precision; but does resonance actually exist? To solve this problem, a *resonant phase* (named also *resonant* or *critical angle*, or *resonant* or *critical argument*) is defined. This is a linear combination (algebraic sum) of angular variables of the system with integer coefficients, the choice of which defines the resonant relation between frequencies. If the amplitude of variation of the resonant phase is limited, i. e. this angle librates, similarly to librations of a pendulum, then the system is in resonance; if it increases or decreases unlimitedly, i. e. rotates, then resonance is absent. The trajectory at the boundary between libration and rotation carries the name of the *separatrix*. So, the dynamics of a rigid pendulum (shown in Fig. 3) provides a model of resonance. In certain sense this model of resonance is universal (Chirikov 1979, 1982).

In celestial mechanics one deals, as a rule, with *nonlinear resonances*: the frequency of the phase oscillations on resonance depends on the amplitude (energy) of oscillations, as in the pendulum example.[1] In case of linear resonance the frequency does not depend on the amplitude.

A slightest external push of the rigid pendulum, placed near the upper position of equilibrium ($\varphi = \pm\pi = \pm180°$; the angle φ is defined in Fig. 3), is capable to change the motion character considerably (e. g. to replace oscillation with rotation). This phenomenon is nothing but the so-called essential dependence on initial conditions. What would occur, if the pendulum or any other system with a separatrix were subject to periodic perturbation? The system motion near the separatrix in the typical case, i. e. for the majority of the initial conditions and parameter values, becomes most unusual. Now this is well-known, but for the

---

[1]The properties of nonlinear resonance are described in detail in Chirikov's general review (Chirikov 1982), where the fundamental concepts of nonlinear dynamics are explained in most accessible and, at the same time, rigorous way.

first time the confusing and intricate behavior of the trajectories close to the *perturbed separatrix* (in the celestial-mechanical *three-body problem*) was pointed out by Henri Poincaré in 1899 (Poincaré 1899). However, it was not supposed at that time that the character of this intricate motion is in any sense "random".

Figure 3.   The pendulum.

For a long time the studies of the motion near the separatrix did not concern chaotic behavior at all, but were limited to analysis of special cases where the traditional approach was possible. In 1908 Andrew Stephenson published a paper (Stephenson 1908) on the dynamics of the inverted ($\varphi = \pi = 180°$) rigid pendulum with the vibrating point of suspension. He found that the vertical vibration of the suspension point with particular values of frequency and amplitude is capable to stabilize the inverted pendulum. Later on, at the end of the forties of the 20th century, Peter Kapitsa (1951, 1954) demonstrated this effect in real experiments. Modern opportunities of computer techniques enable one to look at this effect in a new aspect: if one constructs a *section of the phase space* for such a pendulum, it becomes evident that the domain of stability represents only a tiny island in an extensive chaotic "sea" formed by the trajectories with obviously irregular behavior. Up to the middle of the 20th century this chaotic motion was not subject to scientific research, neither the chaotic motion of any other dynamical system.

In 1959 Chirikov for the first time described dynamical chaos as a phenomenon arising due to *interaction of resonances*. For a criterion of occurrence of chaos he offered the criterion of *overlapping of resonances* (Chirikov 1959, 1979). Let us give an explaining example. The phase space in case of the non-perturbed rigid pendulum has two dimensions, defined by two variables: the angle φ of the pendulum deviation from the vertical, and the momentum $p = ml\dot{\varphi}$, where m is the mass of the pendulum, l is its length, $\dot{\varphi}$ is the rate of variation of the angle φ. In the well-known *phase portrait* "φ–p" of the non-perturbed pendulum (Fig. 4), a single domain ("cell") of librations, bound by the non-perturbed separatrix, is present. In other words, the pendulum model of resonance in the non-perturbed case describes single resonance. If one "switches on" a periodic perturbation, namely, vibration of the suspension point, then the phase space of this dynamical system is not two-dimensional anymore, and for

comparison with the phase portrait in the non-perturbed case it is necessary to construct the *phase space section*. It is built as follows: one plots the values of the system variables on the graph not continuously but "stroboscopically", i. e. discretely at constant intervals of time equal to the period of perturbation. On the section constructed in this way one discovers not one but three domains of librations — three resonances (Fig. 5). If the perturbation frequency is relatively large, the separation of resonances in the momentum is large and they almost do not interact. On reducing the frequency of perturbation, the resonances approach each other and appreciable chaotic layers emerge in the vicinity of the separatrices, where, as it is well visible in Fig. 5, the motion is irregular; on reducing further the frequency of perturbation, the layers merge into a single chaotic layer — the result of interaction of resonances at their strong mutual approach in the phase space.
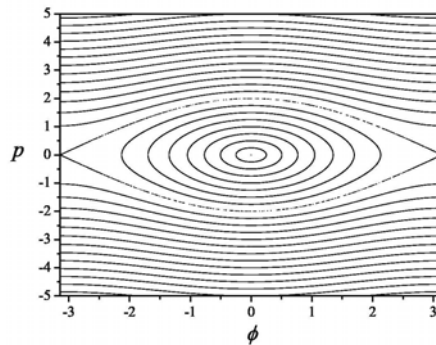

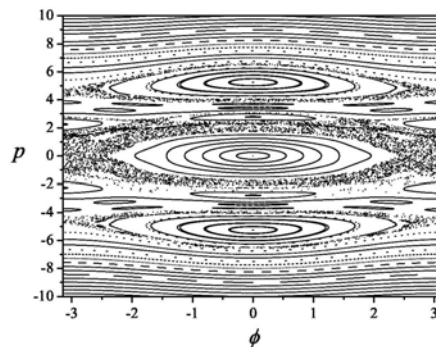Figure 4.   A phase portrait of the non-perturbed pendulum.

Figure 5.   A triplet of interacting resonances at a moderate relative frequency of perturbation.

In Chirikov's saying, "...the physicist first of all tries to find out which resonances play role in this or that system and how do they interact with each other" (Chirikov 1982). It is just the presence of resonances, often regarded to be embodiments of order, leads to the unpredictable, chaotic character of the

motion. In other words, the presence of resonances in the phase space causes the presence of the chaotic component in this space. However, as we have just seen, for chaos to exist, the presence of not one but two or more resonances in the phase space is required, because their interaction is necessary.

3.   Separatrix maps and chaotic layers

The motion in the vicinity of the perturbed separatrices of resonances is chaotic. For weakly perturbed systems the *chaotic layer* concept is convenient to apply for description of the chaotic component. The chaotic layer is a zone in the vicinity of the separatrices in the phase space where the dynamical system moves chaotically. The chaotic layer theory has applications in various areas of physics, mechanics and, in particular, in celestial mechanics (Chirikov 1979; Shevchenko 2007a, 2007b, 2010). The key role in this theory is played by the *separatrix maps*. They represent the motion of a system close to separatrices in a discrete way ("stroboscopically", in the same manner as in the described above construction of the phase space sections): system's state, set by the "time" and "energy" variables, is mapped discretely at the moments of passage of the positions of equilibrium by the pendulum describing the resonance. A rather general example of the separatrix map is

$$
\begin{aligned}
y' &= y + \sin x, \\
x' &= x + \lambda \ln |y'| + c,
\end{aligned}
\tag{1}
$$

where x, y, and x', y' are the "time" and "energy" variables (in the normalized unitless form) at two moments of time following one after another; λ, c are the constants expressed through physical parameters of the original system.

    In deriving the equations of the separatrix map in classical form (Chirikov 1979), the *perturbed pendulum* model considered above, i. e. the dynamical system describing the pendulum with a periodic perturbation, serves for the model of nonlinear resonance. The main parameter of the separatrix map is the so-called *adiabaticity parameter* λ equal to the ratio of the frequency of perturbation to the frequency of small phase oscillations on resonance. The separatrix map is applicable for the description of the motion at any values of this parameter (Shevchenko 2000). An example of the chaotic layer described by the separatrix map is shown in Fig. 6. The separatrix map in case of asymmetric perturbation represents a more complicated algorithm: it contains conditional transfer statements (Shevchenko 1999).

4.   Lyapunov exponents

For dynamical chaos to emerge, three basic conditions must be satisfied (Devaney 1986; Meiss 1992): (1) *Sensitive dependence on initial conditions*, implying that nearby trajectories diverge in time exponentially. (2) The boundedness of the motion, implying that the exponential divergence is not just a smooth infinite expansion. (3) The set of trajectories with such behavior must have non-zero measure.
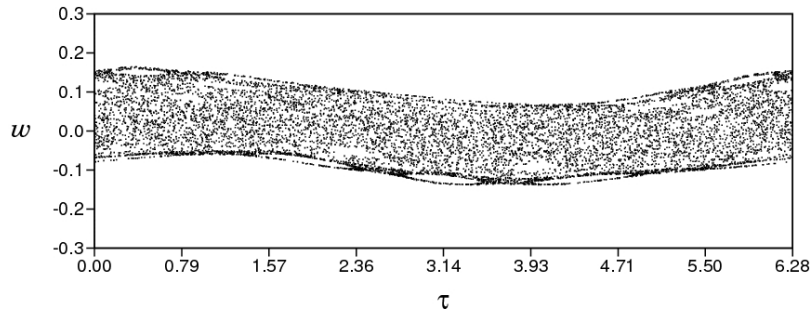
Figure 6. The chaotic layer in the "time–energy" coordinates, an example (Shevchenko 1998b).

The rate of divergence of the close trajectories (in the phase space and in the logarithmic scale of distances) is characterized by the *maximum Lyapunov exponent*. A distinction of the maximum Lyapunov exponent from zero testifies that the motion is chaotic (Chirikov 1979; Lichtenberg & Lieberman 1992). The inverse of this quantity, $T_{\mathrm{L}} \equiv L^{-1}$, is the *Lyapunov time*. It represents the characteristic time of predictable dynamics. The importance of this timescale for celestial mechanics is provided by the fact that any exact theory of the motion of any celestial-mechanical system cannot be constructed on any times much greater than its Lyapunov time.

The art of calculation of the Lyapunov exponents on computers has more than a thirty-year history and during this time it has become an extensive part of applied mathematics; see reviews in (Froeschlé 1984; Lichtenberg & Lieberman 1992). Modern numerical methods provide calculation of the Lyapunov exponents which is effective and precise. On the other hand, methods of analytical estimating the Lyapunov exponents appeared only recently (Holman & Murray 1996; Murray & Holman 1997; Shevchenko 2002a). The method of estimating the maximum Lyapunov exponent (Shevchenko 2002a; Shevchenko 2008b), based on the theory of separatrix maps, allows one to obtain analytical estimates of the Lyapunov exponents (in accord with the numerical-experimental ones) in a number of problems on dynamics of the Solar system bodies; see the review by Shevchenko (2007a).

Generally, estimating the Lyapunov exponents is one of the most important tools in the study of chaotic motion (Lichtenberg & Lieberman 1992), in particular in celestial mechanics.

5. Analytical estimating the Lyapunov exponents

Let us review the problem of analytical estimating the Lyapunov exponents, following (Shevchenko 2007b). In (Shevchenko 2002a), an analytical method for estimating the maximum Lyapunov exponent of the chaotic motion in the vicinity of the separatrices of perturbed nonlinear resonance was derived in the framework of the separatrix map theory. Following this general approach, we represent the maximum Lyapunov exponent L of the motion in the main chaotic layer of the system as the ratio of the maximum Lyapunov exponent L_{sx} of its separatrix map and the average half-period of libration (or, equivalently, the

average period T of rotation) of the resonance phase φ inside the layer. For convenience, we introduce a non-dimensional quantity $T_{\text{sx}} = \Omega T$. Then the general expression for L is

$$L = \Omega \frac{L_{\text{sx}}}{T_{\text{sx}}}. \tag{2}$$

The quantity $T_{\text{L}} \equiv L^{-1}$, by definition, is the Lyapunov time.

5.1. The model of perturbed resonance

For the model of nonlinear resonance we take the Hamiltonian:

$$H = \frac{\mathcal{G}p^2}{2} - \mathcal{F}\cos\varphi + a\cos(\varphi - \tau) + b\cos(\varphi + \tau). \tag{3}$$

Under general conditions (Chirikov 1979; Lichtenberg & Lieberman 1992), a model of nonlinear resonance is provided by the Hamiltonian (3). The first two terms in Eq. (3) represent the Hamiltonian H_0 of the unperturbed pendulum; φ is the pendulum angle (the resonance phase angle), p is the momentum. The periodic perturbations are given by the last two terms; τ is the phase angle of perturbation: $\tau = \Omega t + \tau_0$, where Ω is the perturbation frequency, and τ_0 is the initial phase of the perturbation. The quantities $\mathcal{F}$, $\mathcal{G}$, a, b are constants. The frequency of the pendulum small-amplitude oscillations $\omega_0 \equiv (\mathcal{F}\mathcal{G})^{1/2}$.

An example of the phase space section of the Hamiltonian (3) at $\tau = 0 \bmod 2\pi$ is shown in Fig. 5 ($\Omega = 5$, $\omega_0 = 1$, $a = b$, $\varepsilon \equiv \frac{a}{\mathcal{F}} = 0.5$). This is a chaotic resonance triplet.

The motion near the separatrices of Hamiltonian (3) is described by the separatrix algorithmic map (Shevchenko 1999):

$$\begin{aligned} &\text{if } w_n < 0 \text{ and } W = W^- \text{ then } W = W^+, \\ &\text{if } w_n < 0 \text{ and } W = W^+ \text{ then } W = W^-; \\ &w_{n+1} = w_n - W\sin\tau_n, \\ &\tau_{n+1} = \tau_n + \lambda\ln\frac{32}{|w_{n+1}|} \quad (\bmod\ 2\pi); \end{aligned} \tag{4}$$

with the parameters

$$\lambda = \frac{\Omega}{\omega_0}, \tag{5}$$

$$\begin{aligned} W^+(\lambda, \eta) &= \varepsilon\lambda\left(A_2(\lambda) + \eta A_2(-\lambda)\right), \\ W^-(\lambda, \eta) &= \varepsilon\lambda\left(\eta A_2(\lambda) + A_2(-\lambda)\right), \end{aligned} \tag{6}$$

where $\varepsilon = \frac{a}{\mathcal{F}}$, $\eta = \frac{b}{a}$. The Melnikov–Arnold integral $A_2(\lambda)$ is given by the relation

$$A_2(\lambda) = 4\pi\lambda\frac{\exp(\pi\lambda/2)}{\sinh(\pi\lambda)}, \tag{7}$$

see (Chirikov 1979; Shevchenko 1998b; Shevchenko 2000).

The quantity w denotes the relative (with respect to the separatrix value) pendulum energy: $w \equiv \frac{H_0}{\mathcal{F}} - 1$. The variable τ keeps its meaning of the phase

angle of perturbation. One iteration of map (4) corresponds to one half-period of pendulum's libration or one period of its rotation.

If $a = b$ (symmetric case), the separatrix algorithmic map reduces to the ordinary separatrix map

$$w_{i+1} = w_i - W \sin \tau_i,$$
$$\tau_{i+1} = \tau_i + \lambda \ln \frac{32}{|w_{i+1}|} \quad (\text{mod } 2\pi), \tag{8}$$

first written in the given form in (Chirikov 1979); the expression for W (Shevchenko 1998b; Shevchenko 2000) is

$$W = \varepsilon\lambda \left(A_2(\lambda) + A_2(-\lambda) \right) = 4\pi\varepsilon \frac{\lambda^2}{\sinh \frac{\pi\lambda}{2}}. \tag{9}$$

An equivalent form of Eqs. (8), used e. g. in (Chirikov & Shepelyansky 1984; Shevchenko 1998a), is

$$y_{i+1} = y_i + \sin x_i,$$
$$x_{i+1} = x_i - \lambda \ln |y_{i+1}| + c \quad (\text{mod } 2\pi), \tag{10}$$

where $y = w/W$, $x = \tau + \pi$; and

$$c = \lambda \ln \frac{32}{|W|} \tag{11}$$

(compare with Eqs. (1)).

The applicability of the theory of separatrix maps for description of the motion near the separatrices of the perturbed nonlinear resonance in the full range of the relative frequency of perturbation, including its low values, was discussed and shown to be legitimate in (Shevchenko 2000).

The half-width y_b of the main chaotic layer of the separatrix map (10) in the case of the least perturbed border of the layer is presented as a function of λ in fig. 1 in (Shevchenko 2004a). The observed dependence follows the piecewise linear law with a transition point at $\lambda \approx 1/2$. This transition takes place not only in what concerns the width of the layer, but also in other characteristics of the motion, in particular, in the maximum Lyapunov exponent. The clear sharp transition at this point manifests a qualitative distinction between two types of dynamics, namely, "slow" and "fast" chaos.

The parameter $\lambda = \Omega/\omega_0$ of the separatrix map measures the separation of the perturbing and guiding resonances in the units of one quarter of the width of the guiding resonance, because the resonances separation in frequency space is equal to Ω, while the guiding resonance width is equal to $4\omega_0$ (Chirikov 1979). Therefore, λ can be regarded as a kind of the resonance overlap parameter (Shevchenko 2011). In the asymptotic limit of the adiabatic (slow) case the resonances in the multiplet strongly overlap, while in the asymptotic limit of the "fast" case the resonances are separated. However note that the border $\lambda \approx 1/2$ (Shevchenko 2008a) between slow and fast chaos does not coincide with the borderline between the cases of overlapping and non-overlapping of resonances: the latter borderline lies much higher in λ. E.g., in phase space of

the standard map the integer resonances start to overlap (on decreasing λ) at $\lambda \approx 2\pi/0.97 \approx 6.5$ (Chirikov 1979).

We consider four generic resonance types: the fastly chaotic resonance triplet, fastly chaotic resonance doublet, slowly chaotic resonance triplet, slowly chaotic resonance doublet (we call them, respectively, the "ft", "fd", "st", "sd" resonance types).

5.2.   Fast chaos. Resonance triplet

Consider the case $a = b$ and $\lambda > 1/2$. There is a symmetric (due to the first condition) triad of interacting resonances, and chaos is fast (due to the second condition).

The case of the fastly chaotic triad is completely within the range of applicability of the method presented in (Shevchenko 2002a). The perturbed nonlinear resonance is modelled by Hamiltonian (3) with $a = b$. Following (Shevchenko 2002a), we take the dependence of the maximum Lyapunov exponent of the separatrix map (10) upon λ in the form

$$L_{\text{sx}}(\lambda) \approx C_h \frac{2\lambda}{1 + 2\lambda}, \tag{12}$$

where $C_h \approx 0.80$ is Chirikov's constant (Shevchenko 2004b).

The average increment of τ (proportional to the average libration half-period, or rotation period) inside the chaotic layer is (Chirikov 1979; Shevchenko 2002a):

$$T_{\text{sx}}(\lambda, W) \approx \lambda \ln \frac{32e}{\lambda |W|}, \tag{13}$$

where e is the base of natural logarithms. From Eq. (2), one has for the Lyapunov time for the "ft" resonance type:

$$T_{\text{L}} = \frac{T_{\text{pert}}}{2\pi} \frac{T_{\text{sx}}}{L_{\text{sx}}} \approx T_{\text{pert}} \frac{(1 + 2\lambda)}{4\pi C_h} \ln \frac{32e}{\lambda |W|}, \tag{14}$$

where $T_{\text{pert}} = 2\pi/\Omega$ is the period of perturbation.

5.3.   Fast chaos. Resonance doublet

The previous analysis of the symmetric case $a = b$ sets a foundation for an analysis of the more general asymmetric case $a \neq b$, since the Lyapunov exponents in the asymmetric case can be found by averaging the contributions of the separate components of the chaotic layer.

Calculation of the average constitutes a complicated problem. In particular, one should know the relative average times of residence of the system in three different components of the layer corresponding to direct rotation, reverse rotation, and libration of the pendulum. The relative times of residence depend on the asymmetry of perturbation. A simple heuristic method of averaging was proposed in (Shevchenko 2002a). We consider the limit case of a or b equal to zero. It means that one of the two perturbing resonances simply does not exist, and instead of the resonance triad we have a duad.

If $\lambda > 1/2$, the equality $b = 0$ implies $|W^-| \ll |W^+|$, and, vice versa, $a = 0$ implies $|W^-| \gg |W^+|$. We designate the dominating quantity by W.

Consider first the libration side of the chaotic layer. Then W^- and W^+ alternate (replace each other) at each iteration of the separatrix algorithmic map (4). It is straightforward to show that, if W^- or W^+ is equal to zero, the separatrix algorithmic map (4) on the doubled iteration step reduces to the ordinary separatrix map (8) with the doubled value of λ and the same non-zero value of W. One iteration of the new map corresponds to two iterations of the old one. Since the half-width of the chaotic layer of map (8) in the case of fast chaos is $\approx \lambda W$ (Chirikov 1979, Shevchenko 2008a), the layer's extent in w on the side of librations doubles and becomes $\approx 2\lambda W$. Note that the parameters λ and W are considered here as independent from each other.

Consider then the circulation sides of the chaotic layer. The side corresponding to reverse (or direct) rotations does not exist, if W^- (or, respectively, W^+) is equal to zero; its measure is zero. The other side, corresponding to direct (or reverse) rotations is described by the ordinary separatrix map (8) with the parameters λ, W; its extent in w is $\approx \lambda W$.

The averaged (over the whole layer) value of the maximum Lyapunov exponent is the sum of weighted contributions of the layer components corresponding to the librations, direct rotations and reverse rotations of the pendulum. The weights are directly proportional to the times that the trajectory spends inside the components, and, via supposed approximate ergodicity, to the relative measures of the components in the phase space. Taking into account the just made estimates of the widths of the chaotic layer's components in the duad case, one can expect that the relative weights of librations and circulations in the "fd" case are respectively 4 and 1.

Hence the formula for the Lyapunov exponent for the "fd" resonance type is

$$L = \frac{\Omega}{\mu_{\mathrm{libr}} + 1} \left(\mu_{\mathrm{libr}} \frac{L_{\mathrm{sx}}(2\lambda)}{T_{\mathrm{sx}}(2\lambda, W)} + \frac{L_{\mathrm{sx}}(\lambda)}{T_{\mathrm{sx}}(\lambda, W)} \right), \tag{15}$$

and

$$T_{\mathrm{L}} = \frac{T_{\mathrm{pert}}}{2\pi} \cdot \frac{\mu_{\mathrm{libr}} + 1}{\mu_{\mathrm{libr}} \frac{L_{\mathrm{sx}}(2\lambda)}{T_{\mathrm{sx}}(2\lambda, W)} + \frac{L_{\mathrm{sx}}(\lambda)}{T_{\mathrm{sx}}(\lambda, W)}}, \tag{16}$$

where $\mu_{\mathrm{libr}} \approx 4$, and W, L_{sx}, T_{sx} are given by formulas (9, 12, 13).

5.4. Slow chaos. Resonance triplet

In the case of $\lambda < 1/2$, the diffusion across the layer is slow, and on a short time interval the phase point of the ordinary separatrix map (10) follows close to some current curve. We call this curve guiding. Let us derive an analytical expression for the guiding curve with an irrational winding number far enough from the main rationals. We approximate the winding number by the rationals m/n. Thus $c \approx 2\pi m/n$. Noticing that at an iteration n of the map the phase point hits in a small neighborhood of the starting point, one obtains for the derivative:

$$\frac{dy}{dx} = \frac{1}{nc - 2\pi m} \sum_{k=0}^{n-1} \sin(x + kc) =$$

$$= \frac{1}{nc - 2\pi m} \sin \frac{nc}{2} \operatorname{cosec} \frac{c}{2} \sin\left(x + \frac{n-1}{2}c\right). \qquad (17)$$

Integrating and passing to the limit $n \to \infty$, one obtains:

$$y = -\frac{1}{2} \operatorname{cosec} \frac{c}{2} \cos\left(x - \frac{c}{2}\right) + \mathcal{C}, \qquad (18)$$

where $\mathcal{C}$ is an arbitrary constant of integration.

The motion is chaotic only when curve (18) crosses the singular line $y = 0$. Hence the half-width of the chaotic layer is $y_b = \left|\operatorname{cosec}\frac{c}{2}\right|$. Averaging (by taking an integral analytically) the quantity $-\ln|y_{i+1}|$ (equal to $(\langle\Delta x\rangle - c)/\lambda$, Eqs. (10)), where y_{i+1} is substituted by y of Eq. (18), over the chaotic layer in the derived boundaries, we find the approximate analytical expression for $\Theta \equiv (T_{\mathrm{sx}} - c)/\lambda$:

$$\Theta \approx \ln\left|4\sin\frac{c}{2}\right|. \qquad (19)$$

Then, we need an expression for $L_{\mathrm{sx}}(\lambda)$. We explore the λ dependence of L_{sx} in a numerical experiment. At each step in λ (namely, $\Delta\lambda = 0.005$, $\lambda \geq 0.005$) we find the value of c corresponding to the case of the least perturbed layer and plot the value of L_{sx}. At $\lambda < 0.3$, the dependence turns out to be practically linear. The linear fit $L_{\mathrm{sx}}(\lambda) = a\lambda$ gives $a = 1.01132 \pm 0.00135$, and the correlation coefficient $R = 0.9998$.

We set $L_{\mathrm{sx}} \approx \lambda$ for the generic (non-resonant) values of c and for $\lambda < 1/2$. Then, from Eqs. (2, 19) one has the following approximate formula for the maximum Lyapunov exponent:

$$L \approx \frac{\Omega}{\ln\left|4\sin\dfrac{c}{2}\right| + \dfrac{c}{\lambda}}, \qquad (20)$$

where $c = \lambda \ln\frac{32}{|W|}$.

For $\lambda \ll 1$ one has $W \approx 8\varepsilon\lambda$, hence the formula for the Lyapunov time for the "st" resonance type:

$$T_{\mathrm{L}} \approx \frac{T_{\mathrm{pert}}}{2\pi} \ln\left|\frac{16}{\varepsilon\lambda} \sin\left(\frac{\lambda}{2}\ln\frac{4}{|\varepsilon|\lambda}\right)\right|. \qquad (21)$$

5.5. Slow chaos. Resonance doublet

Utilizing the approximation of the Melnikov–Arnold integral $A_2(\lambda) \approx 2\pi\lambda + 4$ at $\lambda \ll 1$, $\eta = 0$, one has: $W^{\pm} \approx \varepsilon\lambda(4 \pm 2\pi\lambda) \approx 4\varepsilon\lambda$. So, in the "sd" case, the separatrix algorithmic map (4) degenerates to the ordinary separatrix map (8) with $W \approx 4\varepsilon\lambda$, i.e. mathematically the case is equivalent to the "st" case, but with a different (halved) value of W.

Following the lines of the previous Subsection, it is then straightforward to write down the formula for the Lyapunov time for the "sd" resonance type:

$$T_{\rm L} \approx \frac{T_{\rm pert}}{2\pi} \ln \left| \frac{32}{\varepsilon\lambda} \sin \left(\frac{\lambda}{2} \ln \frac{8}{|\varepsilon|\lambda} \right) \right|. \tag{22}$$

5.6. Separatrix map theory versus numerical experiment

To check the theory, the Lyapunov exponents of the chaotic motion near the separatrices of Hamiltonian (3) have been directly computed in (Shevchenko 2007b) by means of the program package (Shevchenko & Kouprianov 2002; Kouprianov & Shevchenko 2003) utilizing the HQRB method by von Bremen, Udwadia, & Proskurowski (1997). The power of this method is far greater than that necessary in the present computation. It allows one to evaluate the full Lyapunov spectrum of a multidimensional system. The spectrum of our system consists of a sole pair of Lyapunov exponents — the maximum one and its negative counterpart. The integration of the equations of motion has been performed by the integrator by Hairer, Nørsett, & Wanner (1987). It is an explicit 8th order Runge–Kutta method due to Dormand and Prince, with the step size control. We choose the time unit in such a way that $\omega_0 = 1$.

The results of the computations for $\lambda = 0.1$ and 2 are shown in Fig. 7 in decimal logarithmic scales. The integration time interval has been chosen to be equal to 10^6. This is sufficient for the computed values of the Lyapunov exponents to saturate in each case. Local wave-like patterns represent prominent features of the constructed dependences; they are conditioned by the process of encountering resonances, while ε changes.
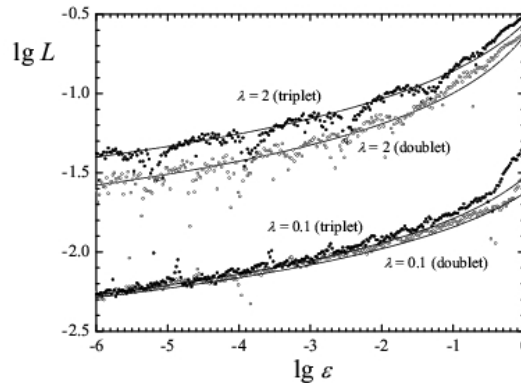

Figure 7. The maximum Lyapunov exponent of the chaotic motion of system (3) in dependence on the amplitude of perturbation: the results of direct computation and the theoretical curves (Shevchenko 2007b). Filled and white dots denote the numerical-experimental data for triplets and doublets, respectively.

Theoretical dependences are shown in Fig. 7 as solid curves. $T_{\rm L}$ are given by formulas (14, 16, 21, 22); $L = 1/T_{\rm L}$. Close correspondence is observed between the theory and experimental data for each value of λ.

5.7. Lyapunov exponents for the motion in supermultiplets

Now let us consider the case when the number of resonances in a multiplet is greater than 3. We assume that the guiding resonance is not at the border of the multiplet. In the case of fast chaos ($\lambda > 1/2$) one can still apply, as an approximation, the theory developed in Subsection 5.2. for the triplet case, because the influence of the "far away" resonances is exponentially small with λ. However, if chaos is slow ($\lambda < 1/2$), the triplet approximation does not work and one has to develop an approach totally different from the previous analysis. Let us consider the limit case — the case of infinitely many interacting equal-sized and equally spaced resonances.

The standard map

$$
\begin{aligned}
y_{i+1} &= y_i + K \sin x_i \quad (\text{mod } 2\pi), \\
x_{i+1} &= x_i + y_{i+1} \quad (\text{mod } 2\pi)
\end{aligned}
\tag{23}
$$

describes the motion in an infinite multiplet of equal-sized and equally spaced resonances, as it is clear from its Hamiltonian

$$
H = \frac{y^2}{2} + K \sum_{n=-\infty}^{\infty} \cos(x - nt)
\tag{24}
$$

(Chirikov 1979). The variables x_i, y_i of the map (23) correspond to the variables $x(t_i)$, $y(t_i)$ of the continuous system (24) taken stroboscopically at time moduli 2π (see e.g. Chirikov 1979).

A formula for the maximum Lyapunov exponent of the standard map was derived in (Chirikov 1979). By means of linearizing the map near some point (not necessarily fixed) and solving the characteristic equation one obtains the following formula for the maximum eugenvalue at $K \gg 1$:

$$
l^+ \approx K|\cos x|.
$$

At $K \gg 1$ the relative measure of the regular component of the phase space of the standard map is close to zero; elliptic periodic trajectories (e.g. the so-called *accelerator modes*) and small regular islands, associated to them, can exist, but the measure of these islands rapidly tends to zero with $K \to \infty$; see (Chirikov 1979; Giorgilli & Lazutkin 2000; Shevchenko 2007c). Therefore one can assume complete ergodicity of the motion at large values of K and obtain the maximum Lyapunov exponent by averaging of l^+ over the phase plane, equivalently over x:

$$
L_{\text{st}} \approx \frac{1}{2\pi} \int_0^{2\pi} \ln K|\cos x| dx = \ln \frac{K}{2}.
\tag{25}
$$

As found by Chirikov in a numerical experiment, already at $K = 6$ the difference between the theoretical and observed values is less than $\approx 2\%$.

The adiabaticity parameter $\lambda \equiv \Omega/\omega_0$ is related to the standard map parameter by the formula $\lambda = 2\pi/\sqrt{K}$ (see Chirikov 1979). Therefore, for equal-sized resonance multiplets, parameterized by λ, one has

$$L_{\mathrm{mp}} \approx \ln \frac{2\pi^2}{\lambda^2}. \tag{26}$$

Since the threshold for global chaos onset is $K \approx 1$ (the condition of marginal overlap; see Chirikov 1979; Lichtenberg & Lieberman 1992), this formula can be used already at λ less than ≈ 6 and, with much higher precision, at λ less than ≈ 3. In other words, the domain of its applicability covers not only the realm of adiabatic chaos, but an actual portion of that of fast chaos as well. However, the number of resonances in the multiplet, for the model to work, should be large, because for the standard map this number is infinite.

Using formulas obtained in (Shevchenko 2004a; Shevchenko 2004b) for the dependence $L(K)$ for the standard map at $K < 1$ and $K > 4.5$, one arrives at the approximation

$$L = \Omega \cdot \begin{cases} 0.1333K, & \text{if } K \leq 2, \\ \ln \frac{K}{2} + \frac{1}{K^2}, & \text{if } K > 2, \end{cases} \tag{27}$$

where $K = (2\pi/\lambda)^2$ and $\Omega = \lambda/(2\pi)$. Note that we glue two asymptotic dependences at $K = 2$, thus somewhat underestimating the actual values of L in the interval $1 > K > 4.5$; see Fig. 8.

Formula (27) can be straightforwardly applied for estimating maximum Lyapunov exponents in resonance multiplets, when the number of resonances in a multiplet is large. The λ dependences, both numerical-experimental and theoretical, of the maximum Lyapunov exponent in multiplets of equal-sized and equally spaced resonances are shown in Fig. 9. One can see that the dependence for the septet occupies an intermediate position between the dependence for the triplet and the dependence for the "infinitet", i. e. for the standard map. The numerical data for the triplet are described well by the separatrix map theory (the lower solid curve) described above in Subsections 5.2. and 5.4., notwithstanding the large amplitude of perturbation $\varepsilon = 1$.

6. Chaos in rotation of planetary satellites

In 1984 Wisdom, Peale, & Mignard predicted, on a basis of numerical experiments and theoretical estimates of the width of a chaotic layer, that the seventh satellite of Saturn, Hyperion, rotates chaotically with respect to its centre of masses: the orientation of the satellite and its rotation rate vary with time chaotically (Wisdom *et al.* 1984).

The cause of chaos in the rotational dynamics of satellites consists in interaction of *spin-orbit resonances*, i. e. resonances between the rotational and orbital motions (see e. g. Wisdom *et al.* 1984; Celletti & Chierchia 2000). It turned out that in the case of Hyperion this interaction is especially strong, due to the strong asymmetry of its figure and a substantial deviation of its orbit from the circular one, i. e. its appreciable eccentricity.

A primary source of observational data on the rotational dynamics of minor planetary satellites is provided by the analysis of their optical lightcurves — the series of the light flux values fixed by observers during extended intervals of time. In 1989 James Klavetter constructed accurate lightcurves of Hyperion with high
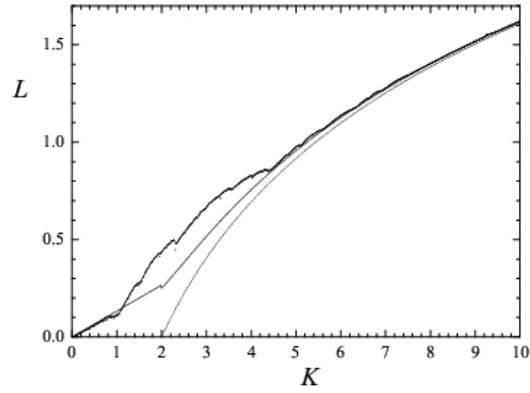
Figure 8.   The dots show the numerical-experimental dependence $L(K)$ for the standard map at $0 < K < 10$ (Shevchenko 2004a; Shevchenko 2004b). The dashed curve shows the function $\ln \frac{K}{2}$. The solid curve is given by Eqs. (27) with $\Omega = 1$.
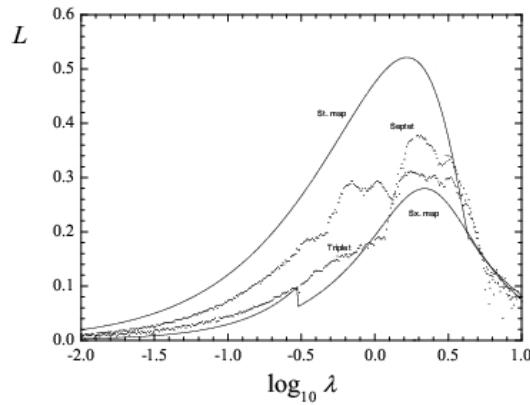

Figure 9.   The λ dependences of the maximum Lyapunov exponent in multiplets. The dots denote numerical-experimental data, and the solid curves denote theoretical dependences. The upper solid curve corresponds to the standard map theory (for the infinite multiplet), whereas the lower one corresponds to the separatrix map theory (for the triplet).

time resolution and carried out their analysis and modelling (Klavetter 1989). He concluded that this satellite was, most likely, indeed in the irregular rotation state. In 2002 Alexander Melnikov at the Pulkovo Observatory with the help of especially developed software modelled the lightcurves of Hyperion (Melnikov 2002), using both the data of Klavetter's observations and the data of Pulkovo observations by A.V.Devyatkin *et al.* By means of computation of the Lyapunov

exponents of the motion he arrived to rigorous conclusion that Hyperion was in the chaotic mode of rotation. In Fig. 10 the model lightcurves of Hyperion are presented.
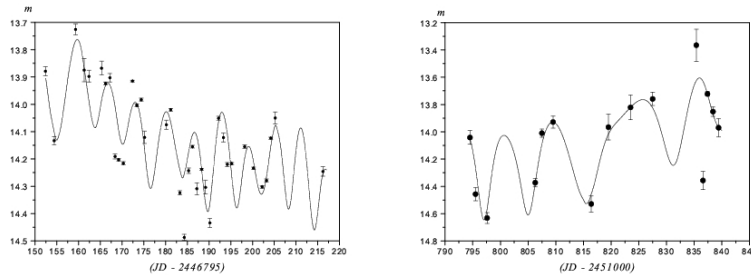

Figure 10. Model lightcurves of Hyperion: for Klavetter's observational data (on the left), and for the Pulkovo observational data obtained in 2000 (on the right). The dots denote the observational data. At the graph axes: JD is the Julian date, m is the stellar magnitude.

The planar (in the orbit plane) oscillations and rotations of a satellite near *synchronous* resonance (the resonance at which the period of rotation of a satellite is equal to its orbital period, as in the case of the Moon) are described by an equation of the perturbed pendulum, where the role of the angle of the pendulum is played by the angle specifying orientation of the satellite with respect to the direction to the planet. Therefore the described above method of analytical estimating the maximum Lyapunov exponent is applicable. The resulting theoretical estimates of the Lyapunov time (≈ 30 days) are in accord with the numerical-experimental ones (Shevchenko & Kouprianov 2002; Kouprianov & Shevchenko 2005).

Do there in the Solar system exist any other (besides Hyperion) chaotically tumbling satellites? For many satellites the character of rotation is unknown. The satellites for which it is established, in the majority rotate synchronously with the orbital motion. Similar to the Moon, they permanently face their planet with one side. However, in the course of the long-term dynamical evolution any satellite at some moment of time enters a state of chaotic rotation, e. g. when passing the separatrix of synchronous resonance, because otherwise it cannot be captured in synchronous rotation. Minor and large satellites have different characters of chaotic rotation. For the minor (irregular-shaped) satellites such rotation represents three-dimensional tumbling; for the majority of large satellites (close to spherical in shape) the rotational motion in the chaotic layer in the vicinity of the separatrix of synchronous resonance preserves planar character: the axis of rotation remains to be approximately orthogonal to the orbit plane (Kouprianov & Shevchenko 2005).

Our theoretical research (Kouprianov & Shevchenko 2005) showed that, besides Hyperion, in the state of chaotic rotation might also reside the 16th and 17th Saturnian satellites Prometheus and Pandora. If they rotate chaotically, the Lyapunov times of their rotation are very small: 0.5–0.6 days. The character of their rotation is not known yet. The following problem is important: whether in the course of the "chaotic tumbling" there exists a preferable orientation of

the satellite, or all orientations are equiprobable? Our calculations (Melnikov & Shevchenko 2008) testify that Prometheus and Pandora in the case of chaotic rotation do have preferable orientations of the largest axis of the figure in direction to Saturn. This complicates making conclusions about the character of rotation of these satellites from observations, because the chaotic mode is similar in some sense to the usual synchronous rotation.

So, the Lyapunov times of chaotic rotation of minor planetary satellites of the Solar system can be very small: as small as a half-day, i. e. chaos in case of its presence is quite observable. However, till now the chaotic rotation was observed only in the case of Hyperion. More than 160 satellites of planets have been discovered till now, but the state of rotation is determined from observations for only 33 of them. Amongst these 33 satellites, 25 are in synchronous resonance and 7 have the periods of rotation much less than the orbital periods.

And what about the Moon, which is of course most important to us, amongst all the planetary satellites? As it is well-known to everyone, the Moon always faces the Earth with one and the same side, i. e. it is in synchronous resonance. It is not so widely known that the Moon is subject to small oscillations with respect to the exact synchronous state. This phenomenon carries the name of *physical libration* of the Moon. How much regular and predictable are these small oscillations of our satellite, if one considers large intervals of time? Unfortunately, this problem has not been considered till now, maybe due to the complexity of the problem.

7.   Satellite systems

If the orbital frequencies of planets in a planetary system or satellites in a satellite system are approximately commensurable, i. e. their ratio is approximately equal to the ratio of two integers, the system is either close to orbital resonance,[2] or resides in it. In the Solar system, many planetary satellites belong to resonant or close-to-resonant systems. E.g., in the satellite system of Jupiter, Io and Europa, as well as Europa and Ganymede, are in the *mean motion resonance* 2/1. In the system of Saturn, Mimas and Tethys, as well as Enceladus and Dione, are also in the mean motion resonance 2/1, Dione and Rhea are close to resonance 5/3, Titan and Hyperion are in resonance 4/3. In the system of Uranus, all resonances are approximate: Miranda and Umbriel move close to resonance 3/1, Ariel and Umbriel — 5/3, Umbriel and Titania — 2/1, Titania and Oberon — 3/2.

Captures of satellite systems in orbital resonances are natural stages of tidal evolution of these celestial-mechanical systems. Tittemore & Wisdom (1988, 1989) and Dermott, Malhotra, & Murray (1988) showed that such capture can significantly affect the further orbital dynamics of satellites and even their internal structure. In the course of capture in orbital resonance or in the course of escape from it, the system crosses the chaotic layer in the vicinity of the resonance separatrices. Inside the layer the system moves chaotically. However,

---

[2]The expression "close to resonance" means that the frequencies satisfy the given integer relation only approximately, and thus the corresponding resonant phase rotates, instead of libration.

the cause of chaos, as a rule, consists not in the interaction of separate mean motion resonances, but in the interaction of *subresonances* in a multiplet corresponding to a single mean motion resonance. The form and orientation of orbits in satellite systems are subject to slow variations, including monotonous (secular) precession of orbits. The splitting of orbital resonances into subresonance multiplets is caused by this precession.

The second (Miranda) and fifth (Umbriel) satellites of Uranus do not reside now in exact orbital resonance with each other, but they are rather close to resonance 3/1. According to Tittemore and Wisdom's conclusions, it is rather probable that in the past, in the course of tidal evolution of their orbits, these satellites spent some time in resonance 3/1. An example of the phase space section of the motion for one of the stages of evolution of this system is presented in Fig. 11. Numerical-experimental and analytical (obtained with the help of the separatrix map theory) estimates of the Lyapunov time were made by Melnikov & Shevchenko (2005). The Lyapunov time of the chaotic system at various stages of its evolution turned out to be equal to 50–100 yr.

Chaotic modes in the dynamics of satellite systems took place not only during the former epochs of the history of the Solar system. Champenois & Vienne (1999a, 1999b) considered the dynamics of the first and third satellites of Saturn, Mimas and Tethys, residing in orbital resonance 2/1. These dynamics might be chaotic due to the large amplitude of librations on resonance. Champenois and Vienne reduced averaged equations of the system motion to the equation of pendulum with periodic perturbations. Using this analytical representation, it is possible to obtain analytical estimates of the Lyapunov time. Such estimates were obtained by Melnikov & Shevchenko (2005), and not only analytically, but also numerically, by means of direct integration of the equations of motion. The Lyapunov times turned out to be 300–600 yr in various models. Thus, chaos in this system does not manifest itself on timescales accessible to observations.

The behavior of the Prometheus–Pandora system (the 16th and 17th Saturnian satellites — the shepherd satellites of the ring F) is much more chaotic. In 2003 Goldreich and Rappaport estimated, by means of numerical experiments, the Lyapunov time of the orbital motion of this system in the vicinity of the mean motion resonance 121/118 (Goldreich & Rappaport 2003a, 2003b). According to their results, it is equal to ≈ 3.3 yr. Farmer & Goldreich (2006) reduced the equations of the system motion to the equation of the pendulum with periodic perturbations. It is possible to obtain an analytical estimate of T_L (Shevchenko 2008b). It is inside the narrow range of the numerical-experimental estimates.

Therefore, the available estimates of the Lyapunov time for the Prometheus–Pandora (≈ 3 yr), Miranda–Umbriel (50–100 yr) and Mimas–Tethys (300–600 yr) systems testify that the range of the times of predictable dynamics in the chaotic satellite systems is rather broad: in the known cases it extends from one year to thousand years (by the order of magnitude). In case of the Prometheus–Pandora system, orbital chaos, due to the small Lyapunov time, is potentially observable.
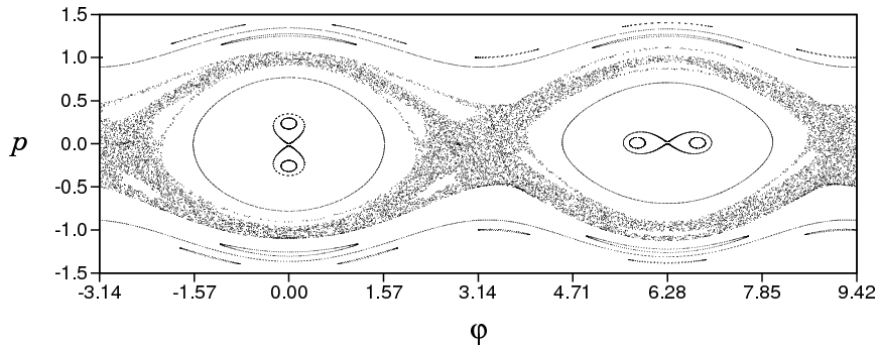
Figure 11.   The phase space section for the Miranda–Umbriel system (Melnikov & Shevchenko 2005). Broad chaotic layers in the vicinity of resonance separatrices are prominent.

8.   Orbital dynamics of asteroids and comets

The important role of resonances in the dynamics of asteroids became evident since the time of discovery of gaps (sparsely populated intervals in the distribution of mean motions) in the belt of asteroids in 1867 by Daniel Kirkwood. The most expressed minima in the distribution of the semimajor axes of the asteroidal orbits correspond to the mean motion resonances 2/1, 3/1, 4/1, 5/2 and 7/3 with Jupiter (see e. g. Murray & Dermott 1999; Morbidelli 2002). The role of chaos in "clearing" the gaps from the asteroids was found out by Wisdom in the beginning of eighties of the last century, at least for the gap corresponding to resonance 3/1 (this one and that corresponding to resonance 2/1 are the most prominent): most of the chaotic orbits close to this commensurability demonstrate sporadic jumps of the eccentricity, resulting in crossings of the orbits with the orbit of Mars. Therefore, sooner or later the asteroids are ejected from the gap area owing to close encounters with this planet. However, as was discovered later on in the works by S. Ferraz-Mello & J. C. Klafke (1991), B. Gladman *et. al.* (1997), the overlap of the 3/1 resonance with secular resonances ν_5 and ν_6 at moderate and high eccentricities provides much greater pumping (up to unity) of the asteroidal eccentricity and consequently a much higher rate of the population depleting, than in Wisdom's scenario, due to encounters with Earth and Venus and falling onto the Sun. The same phenomenon of overlap with secular resonances provides fast depleting of population in other important mean motion resonances in the asteroid belt, such as 2/1, 4/1, 5/2, and 7/3 (Moons & Morbidelli 1995; Gladman *et. al.* 1997; Moons, Morbidelli, & Migliorini, 1998; for a review see Morbidelli 2002).

The domain of chaos corresponding to the 3/1 resonance, computed by means of Wisdom's (1983) map in the planar elliptic restricted three-body problem Sun–Jupiter–asteroid, is shown in Fig. 12 on a representative set of initial values of the semimajor axis a and eccentricity e of the asteroidal orbits. The borders of this domain are determined by the calculation of the maximum Lyapunov exponent of the asteroidal motion. The initial data corresponding to the chaotic trajectories with and without jumps of the eccentricity are shown re-

spectively in black and grey. The graph reveals an evidently significant extent of the near-resonant domain of chaos.
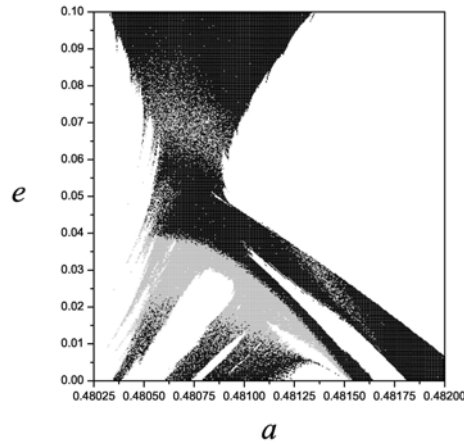

Figure 12.   The region of the 3/1 mean motion resonance with Jupiter in the planar elliptic three-body problem Sun–Jupiter–asteroid (Shevchenko 2002b); a is the semimajor axis, e is the eccentricity of an asteroidal orbit.

The orbital resonances in the motion of asteroids subdivide into the *mean motion resonances* and *secular resonances*. The mean motion resonances represent the commensurabilities between the mean frequencies of the orbital motions of an asteroid and a planet, and the secular resonances represent the commensurabilities between the rates of precessions of the orbits of an asteroid and a planet. Besides the usual (two-body) mean motion resonances, an essential role in the dynamics of asteroids is played by the so-called *three-body mean motion resonances* (Murray, Holman, & Potter 1998; Nesvorný & Morbidelli 1998, 1999; Morbidelli 2002; Cachucho, Cincotta, & Ferraz-Mello 2010). In this case the resonant phase represents a combination of the angular coordinates of an asteroid and two planets (e. g. an asteroid, Jupiter and Saturn). Both for the two-body and three-body resonances, the equations of motion in typical cases are reduced approximately to the equation of pendulum with periodic perturbations; therefore, it is possible to estimate analytically the Lyapunov times of the motion (Shevchenko 2007b).

As follows from the orbital data placed at the AstDyS web service[3] (Knežević & Milani 2000) for hundreds of thousands asteroids, the Lyapunov times of the asteroidal motion in the main belt that are less than 400 yr are absent among the numerical-experimental estimates. It is usually not known what is the resonant multiplet that serves as the cause of the chaotic behavior of a particular asteroid, because it is rather difficult to identify three-body resonances (which are most numerous among the important resonances).

---

[3]http://hamilton.dm.unipi.it/cgi-bin/astdys/

In studies of chaotic dynamics of celestial bodies, the motion in perturbed highly-eccentric orbits is of special interest, because it is often strongly chaotic. The asteroids and comets crossing the orbits of planets are by all means amongst the most chaotic objects of the Solar system. Giovanni Valsecchi (2007) showed convincingly that the history of researches of chaotic dynamics of the minor bodies of the Solar system counts from the works on the dynamics of the Lexell comet accomplished by Leonhard Euler's pupil Andrej Lexell in the seventies of the 18th century and by U. Le Verrier in the forties and fifties of the 19th century. The works of Lexell were published in the proceedings of Academia Petropolitana — the Russian Academy of Sciences. In Valsecchi's saying, "It can be said that the work of Lexell started the modern understanding of the dynamics of small solar system bodies" (Valsecchi 2007). This modern understanding consists in taking into account the paramount role of resonances and close encounters with planets. In celestial-mechanical calculations of Le Verrier, the concept of essential dependence on initial conditions appeared for the first time: very small model variations (about several metres per second) of the velocity of the Lexell comet in its orbit perihelion resulted in *qualitative* changes of the orbit.

In 1770, the Lexell comet passed at an unprecedentedly close distance from the Earth. Nine years later, it was most probably ejected from the Solar system as a result of its close encounter with Jupiter. In the history of astronomical observations it became the first outstanding example of an object closely encountering the Earth. This example is enough to make it clear that the problem of assessment of the degree of predictability of the motion of planet-crossing bodies has not only theoretical but also important practical significance.

Such an assessment for the orbital motion of the asteroids passing close to the Earth (near-Earth asteroids — NEAs), and other objects potentially hazardous to the Earth, is one of the most complicated aspects of the problem of asteroid-cometary hazard. Some asteroids move in virtually predictable orbits, others do not. In A. Whipple's (1995) saying, "The existence of a significant population of extremely chaotic Earth-crossing asteroids must be factored into the thinking about the potential hazard posed by these objects. An asteroid with a Lyapunov time of 20 years may be considered as an example. If the initial error in its position is 100 km (a very optimistic assumption) then that error will grow to one Earth radius in 83 years and to an Earth–Moon distance in 165 years. Assessments of the threat from specific objects like this can be made for only short spans of time."

The Lyapunov exponents for this kind of objects can be estimated by means of analysis of the so-called *Kepler map*. The equations of this map were deduced by T. Petrosky (1986), and independently by Chirikov & Vecheslavov (1986, 1989), in connection with apparition of the Halley comet. The model of the comet motion consisted in the assumption that the main perturbing effect of Jupiter is concentrated on a rather short interval of time when the comet is close to its orbit perihelion. This effect is determined by the phase of Jupiter at this short time span. The deduction of the formula for the Lyapunov time of the motion of comets and other planet-crossing objects (such as NEAs) is based on the Kepler map concept as a generalized separatrix map, the non-perturbed parabolic orbit playing the role of the non-perturbed separatrix (Shevchenko 2007b). This formula gives the lower bound for the possible values of the Lya-

punov time. As far as it is known, at present none of the observed objects breaks this border.

As a whole, the planet-crossing asteroids (NEAs among them) and comets are amongst the most chaotic objects of the Solar system. The Lyapunov times of these objects in comparison with the Lyapunov times of the majority of usual asteroids can be very small — as low as several years.

8.1. Two-body mean motion resonances

Let us consider analytical estimating the maximum Lyapunov exponents of the orbital motion of asteroids in two-body mean motion resonances, following Shevchenko (2007b).

The Hamiltonian of the motion of a zero-mass test particle in the gravitational field of the Sun and Jupiter, in the plane of Jupiter's orbit, in the vicinity of a mean motion resonance with Jupiter can be represented in some approximation in the following form (Holman & Murray 1996; Murray & Holman 1997):

$$H = \frac{1}{2}\beta\Lambda^2 - \sum_{p=0}^{q} \varphi_{k+q,k+p,k} \cos(\psi - p\omega), \qquad (28)$$

where $\beta = 3k^2/a^2$, $\Lambda = \Psi - \Psi_{\rm res}$, $\Psi = (\mu_1 a)^{1/2}/k$, $\Psi_{\rm res} = (\mu_1^2/(k^2(k+q)n_{\rm J}))^{1/3}$, $\mu_1 = 1-\mu$, $\omega \equiv -\varpi$ (i. e. ω is minus the longitude of asteroid's perihelion; its time derivative is assumed to be constant); a and e are asteroid's semimajor axis and eccentricity. The integer non-negative numbers k and q define the resonance: the ratio $(k+q)/k$ equals the resonant ratio of the orbital frequencies of the particle and Jupiter. The phase $\psi \equiv kl - (k+q)l_{\rm J}$, where l and $l_{\rm J}$ are the mean longitudes of an asteroid and Jupiter.

Here the units are chosen in such a way that the total mass (Sun plus Jupiter), the gravitational constant, Jupiter's semimajor axis $a_{\rm J}$ are all equal to one; $\mu = 1/1047.355$, $\mu_1 = 1 - \mu$. Jupiter's mean longitude $l_{\rm J} = n_{\rm J}t$, eccentricity $e_{\rm J} = 0.048$. Jupiter's mean motion $n_{\rm J} = 1$, i. e. the time unit equals $\frac{1}{2\pi}$th part of Jupiter's orbital period.

According to Eq. (28), the resonance $(k+q)/k$ splits in a cluster of $q+1$ subresonances $p = 0, 1, \ldots, q$. The coefficients of the resonant terms are

$$|\varphi_{k+q,k+p,k}| \approx \frac{\mu}{q\pi a_{\rm J}} \binom{q}{p} \left(\frac{\varepsilon}{2}\right)^p \left(\frac{\varepsilon_{\rm J}}{2}\right)^{q-p}, \qquad (29)$$

where $\varepsilon = ea_{\rm J}/|a - a_{\rm J}|$, $\varepsilon_{\rm J} = e_{\rm J}a_{\rm J}/|a - a_{\rm J}|$ (Holman & Murray 1996; Murray & Holman 1997). The approximation (29) is good, if $\varepsilon q < 1$ (Holman & Murray 1996). Besides, the model is restricted to the resonances of relatively high order, $q \geq 2$.

The signs of the coefficients $\varphi_{k+q,k+p,k}$ alternate with changing p; therefore, the coefficients with numbers p and $p+2$ are always of the same sign. This means that whatever is the choice of the guiding resonance in the multiplet, its closest neighbors have coefficients of equal signs, and η is always non-negative.

The frequency of small-amplitude oscillations on subresonance p is

$$\omega_0 = (\beta|\varphi_{k+q,k+p,k}|)^{1/2} \approx \frac{a_{\mathrm{J}}}{|a-a_{\mathrm{J}}|} n_{\mathrm{J}} \left(\mu_1 \mu \frac{4q}{3\pi} \binom{q}{p} \left(\frac{a}{a_{\mathrm{J}}} \right) \left(\frac{\varepsilon}{2} \right)^p \left(\frac{\varepsilon_{\mathrm{J}}}{2} \right)^{q-p} \right)^{1/2},$$

(30)

and the perturbation frequency is

$$\Omega = \dot{\omega} \approx \frac{\mu_1 \mu}{2\pi} n_{\mathrm{J}} \left(\frac{a}{a_{\mathrm{J}}} \right)^{1/2} \left(\frac{a_{\mathrm{J}}}{a-a_{\mathrm{J}}} \right)^2$$

(31)

(cf. Holman & Murray 1996; Murray & Holman 1997). The ratio of Ω and ω_0 gives the value of λ.

Table 1.   Numerical and analytical estimates of Lyapunov times for two-body mean motion resonances, T_{L} is in years, $e = 0.1$ (Shevchenko 2007b)

$\frac{k+q}{k}$	λ	$\log_{10} T_{\mathrm{L}}^{\mathrm{num}}$ †	$\log_{10} T_{\mathrm{L}}^{\mathrm{theor}}$	Res. type
3/1	0.093	3.8–4.3	4.3	sd
5/2	0.192	3.5–3.8	4.1	st
7/3	0.415	3.8–4.2	4.0	st
9/4	0.932	3.9–4.3	4.2	fd*
11/5	1.970	3.9–4.3	4.3	ft*
9/5	0.323	3.6–3.8	3.7	st*
7/4	0.166	3.2–3.3	3.7	st
12/7	0.594	3.6–4.0	3.9	fd*
5/3	0.101	2.5–3.3	3.7	sd
8/5	0.156	2.5–3.3	3.6	st*
11/7	0.264	3.3–3.6	3.5	st*

† (Morbidelli & Nesvorný 1999; Holman & Murray 1996)

Now we are able to apply the theory developed in Section 5. For comparison, we take the data on the numerical values (based on integrations) of T_{L} for the motion near the mean motion resonances presented in Fig. 1 in (Morbidelli & Nesvorný 1999) and Fig. 6 in (Holman & Murray 1996). The theoretical estimates are made by means of formulas (14, 16, 21, 22). Before they are used, the guiding resonance in the multiplet is identified (it has the maximum value of $|\varphi_{k+q,k+p,k}|$), and its two closest neighbors are considered as the perturbing resonances. Then the formula is chosen in accord with the resonance type (fastly chaotic triad "ft", fastly chaotic duad "fd", slowly chaotic triad "st", or slowly chaotic duad "sd"). If the amplitudes of the neighbors differ from each other less than twice, the model resonance is considered to be a triad, otherwise a duad. Those resonances which have $\varepsilon q > 1$ are marked in Table 1 by an asterisk.

The analytical estimates of the maximum Lyapunov exponent are generally in agreement with the numerical ones. However, some differences can be clearly seen, especially in the domain of slow chaos. This should be attributed to the imperfectness of model (28), and mainly to the fact that the coefficients

$\varphi_{k+q,k+p,k}$ are treated as constants. They fix the frequencies ω_0 of small amplitude oscillations on subresonances, and if the period of perturbation is large in relation to the period of these oscillations, the variations of ω_0 can have greater dynamical effect.

Also another effect can be important. The differential distribution, built by Shevchenko, Kouprianov, & Melnikov (2003) for a representative set of starting values (the trajectories were computed by means of Wisdom's (1983) map in the planar elliptic restricted three-body problem), demonstrates that the maximum Lyapunov exponent of the trajectories near the 3/1 mean-motion resonance with Jupiter has two preferable numerical values: the distribution has a bimodal peak structure. This signifies that there are two distinct domains of chaos in phase space; thus the perturbed pendulum model as applied to this low-order mean motion resonance turns out to be too approximate. This example shows also that generally a closer look at the numerical data may be necessary when comparing it with a theory.

Now consider an example of an estimate for an actual asteroid, namely 522 Helga. This object is famous to be the first example of "stable chaos" among asteroids (Milani & Nobili 1993): while its Lyapunov time is relatively small (6900 yr), its orbit does not exhibit any gross changes on cosmogonic time scales, according to numerical experiments. It is known to be in the 12/7 mean motion resonance.

Let us apply our method. The necessary data on a, e, and the perihelion frequency $g = \dot{\varpi}$ are taken from the "numb.syn" catalogue of the AstDyS web service. T_{pert} is defined by the value of g. One finds that the guiding subresonance in the resonance sextet is the third one ($p = 2$), consequently the perturbing neighbors in our model correspond to $p = 1$ and 3. The quantity $\varepsilon q = 0.624 < 1$, therefore the potential model is valid. The derived separatrix map parameters are: $\lambda = 2.325$, $\eta = 0.812$, consequently the model resonance type is the fastly chaotic triplet "ft". Applying formula (14), one has $T_{\text{L}} = 9700$ yr. The agreement with the values, obtained in integrations in the full problem, (6900 yr, according to Milani & Nobili (1993), and 6860 yr, according to the AstDyS data) are quite satisfactory.

8.2. Three-body mean motion resonances

Let us consider analytical estimating the maximum Lyapunov exponents of the orbital motion of asteroids in three-body mean motion resonances, following Shevchenko (2007b).

The three-body resonances can be described by the perturbed pendulum model (Murray, Holman, & Potter 1998; Nesvorný & Morbidelli 1998; Nesvorný & Morbidelli 1999). The Hamiltonian of the motion of a zero-mass test particle near a three-body resonance $\{m_{\text{J}} m_{\text{S}} m\}$ with Jupiter and Saturn in the planar-elliptic problem can be expressed, in some approximation, in the following form (Nesvorný & Morbidelli 1999):

$$H = \alpha S^2 + \sum_{p_{\text{J}}, p_{\text{S}}, p} \beta_{p_{\text{J}} p_{\text{S}} p} \cos \sigma_{p_{\text{J}} p_{\text{S}} p} \,, \tag{32}$$

where the resonant argument conjugated to S is $\sigma_{p_{\text{J}} p_{\text{S}} p} = m_{\text{J}} l_{\text{J}} + m_{\text{S}} l_{\text{S}} + m l + p_{\text{J}} \varpi_{\text{J}} + p_{\text{S}} \varpi_{\text{S}} + p \varpi$. It is assumed that the time derivatives of l_{J}, l_{S}, ϖ_{J}, ϖ_{S} are

constants; $\alpha = -(3/2)n^2 a_{\mathrm{res}}^{-2}$. Analytical expressions $\beta_{p_{\mathrm{J}}p_{\mathrm{S}}p}(e)$ for some important three-body resonances are given in Tables 3–6 in (Nesvorný & Morbidelli 1999). It is clear from Eq. (32) that the three-body resonance $\{m_{\mathrm{J}}m_{\mathrm{S}}m\}$ splits in a cluster of subresonances with various $\{p_{\mathrm{J}}p_{\mathrm{S}}p\}$ combinations.

The frequency of small-amplitude oscillations on subresonance $\{p_{\mathrm{J}}p_{\mathrm{S}}p\}$ is (Nesvorný & Morbidelli 1999):

$$\omega_0 = 2\pi n (3\beta_{p_{\mathrm{J}}p_{\mathrm{S}}p})^{1/2} a_{\mathrm{res}}^{-1}. \tag{33}$$

The perturbation frequency Ω is generally an algebraic combination of perihelion frequencies of Jupiter, Saturn and the asteroid. The formula for this combination is defined by the choice of the guiding subresonance (see below). The ratio of Ω and ω_0 gives the value of λ.

Table 2. Numerical and analytical estimates of Lyapunov times for asteroids in three-body mean motion resonances (Shevchenko 2007b)

Asteroid	Resonance $\{m_{\mathrm{J}}m_{\mathrm{S}}m\}$	λ	$T_{\mathrm{L}}^{\mathrm{num}}$, yr †	$T_{\mathrm{L}}^{\mathrm{num}}$, yr ‡	$T_{\mathrm{L}}^{\mathrm{theor}}$, yr	Res. type
258 Tyche	$2 + 2 - 1$	0.536	35900	–	43100	ft
485 Genua	$3 - 1 - 1$	0.376	6550	6500	35700	sd
1642 Hill	$3 - 1 - 1$	0.643	36100	–	43300	fd
936 Kunigunde	$6 + 1 - 3$	0.624	22200	–	54600	fd
490 Veritas	$5 - 2 - 2$	0.546	10200	8500	9100	fd
2039 Paine-Gaposchkin	$5 - 2 - 2$	0.449	22000	–	6020	sd
3460 Ashkova	$5 - 2 - 2$	0.433	65100	8300	5940	sd

† AstDyS
‡ (Nesvorný & Morbidelli 1998; Nesvorný & Morbidelli 1999; Milani, Nobili, & Knežević 1997)

We consider the asteroids residing close to the three-body resonances studied in (Nesvorný & Morbidelli 1999) and utilize the analytical data in Tables 3–6 in (Nesvorný & Morbidelli 1999) on the coefficients of the resonant terms. The theoretical estimates of Lyapunov times are made by means of formulas (14, 16, 21, 22). Before they are used, the guiding resonance in the multiplet is identified (it has the maximum value of $|\beta_{p_{\mathrm{J}}p_{\mathrm{S}}p}|$), and its two closest neighbors are considered as the perturbing resonances. Then, the formula is chosen in accord with the resonance type ("ft", "fd", "st", "sd").

The identification of the guiding resonances in the multiplets reveals that the three-body resonances under study subdivide in two distinct classes: those for which the perturbation frequency Ω in model (3) is equal to $\dot{\varpi} - \dot{\varpi}_{\mathrm{J}}$, and those for which it is equal to $\dot{\varpi}_{\mathrm{S}} - \dot{\varpi}_{\mathrm{J}}$. The resonances $5 - 2 - 2$ and $3 - 1 - 1$ belong to the first class, while $2 + 2 - 1$ and $6 + 1 - 3$ to the second. We use $\dot{\varpi}_{\mathrm{J}} = 4.257''/\mathrm{yr}$ and $\dot{\varpi}_{\mathrm{S}} = 28.243''/\mathrm{yr}$ (Bretagnon 1990). The data on $\dot{\varpi}$, a, and e are taken from the "numb.syn" catalogue of the AstDyS web service.

Theoretical estimates obtained in this way are presented in Table 2. Some of them are in accord with the numerical ones (in particular, in the case of Veritas), others are not. From the fact of disagreement in the cases of Genua and Paine-

Gaposchkin, one can judge that these objects do not, most probably, reside in the chaotic layers of the prescribed resonance multiplets. Thus analytical estimating the Lyapunov times represents a promising tool for discerning between possible models of chaos in the motion of actual asteroids, and, generally, celestial bodies.

9. Multiple asteroids

In the beginning of the nineties of the 20th century, B. Chauvineau, P. Farinella, and F. Mignard found out that the dynamics of a small satellite of a rotating irregular-shaped asteroid can be highly chaotic (Chauvineau *et al.* 1993). In 1997, J.-M. Petit and co-workers made a unique (up to present time) estimation of the Lyapunov time of the possible chaotic dynamics of a double asteroid: by means of a numerical experiment they established that the orbital motion of Dactyl, a satellite of the asteroid 243 Ida, can be chaotic with the Lyapunov time ranging from 9 d to 4 yr (Petit *et al.* 1997). In highly chaotic modes of the motion, the satellite either falls on the surface of Ida or escapes very soon.

The population of double asteroids includes, among others, the asteroid 762 Pulcova. Although more than 160 multiple asteroids have been discovered already, and new and new are discovered, one should not expect, apparently, that chaos in such systems can be often observed. The systems for which the type of dynamics is already known are regular, and, moreover, as a rule they possess double synchronization: the periods of rotation of both components about their centres of masses are equal to the orbital period, thus both components are always facing each other by the same sides, i. e. one deals with the systems that have reached the final stage of tidal evolution.

10. In brief about planets

And at the highest level of the structural hierarchy of the Solar system there are resonances and chaos. The approximate orbital commensurabilities Jupiter–Saturn (the ratio of mean motions $\approx 5/2$), Saturn-Uranus ($\approx 3/1$), Uranus–Neptune ($\approx 2/1$), and the Neptune–Pluto resonance $(3/2)$ are well known. At the end of the eighties G. Sussman and J. Wisdom and, independently, J. Laskar in complicated numerical experiments obtained the first estimates of the Lyapunov time of the Solar system (Sussman & Wisdom 1988; Laskar 1989). It turned out that it is not at all infinite, i. e. the motion of the Solar system is not regular. Moreover, the Lyapunov time is rather small: it is three orders of magnitude less than the age of the Solar system. According to Sussman and Wisdom's calculations, the Lyapunov time of the outer Solar system (that from Jupiter to Pluto) is equal to ≈ 10 million years. And for the system of all planets, either with Pluto or without it, $T_{\rm L} \approx 5$ million years.

At a first glance it might seem that the basic contribution to chaos must be brought by the planets with relatively small masses, i. e. the planets of terrestrial group, as well as Pluto ranked as a planet until recently. However, if the calculations were limited only to four planets-giants, then, as it was found out in 1992 by Sussman and Wisdom and confirmed in 1999 by N. Murray and M. Holman, chaos remains and, moreover, the Lyapunov time practically does not change: $T_{\rm L} \approx 5$–7 million years (Sussman & Wisdom 1992; Murray & Holman 1999).

Murray & Holman (1999) found that chaos could be due to a multiplet of subresonances associated to a particular three-body resonance Jupiter–Saturn–Uranus. This conclusion, however, has a preliminary character, because no full agreement of the analytical model with the numerical experiments has been achieved up to now. If it is true, the degree of chaoticity of the Solar system has, in some sense, an arbitrary character: if the semimajor axis of the orbit of Uranus differed from the present value by only several diameters of Uranus, the chaoticity would sharply decrease, if not at all practically disappear. Recent numerical-experimental studies by W. Hayes and co-workers (Hayes 2007, 2008; Hayes *et al.* 2010) show that the spatial interplay of chaos and order may be present even on much finer scales. In the future researches, if the guiding three-body resonance is accurately identified, then analytical estimating the Lyapunov time T_{L} will be possible to accomplish using the separatrix map theory.

In conclusion

The idea of full predefiniteness of the motion of both large and small bodies of the Solar system was fundamental in celestial mechanics until the eighties of the 20th century, but now we know that the time of predictable motion of any body in the Solar system is limited, by the order of magnitude, by the value of its Lyapunov time. For some bodies this time is large and the motion is practically regular; for others, on the contrary, dynamical chaos manifests itself on small timescales and, therefore, it can be observed directly.

Direct astronomical observations of chaos in case of its presence are possible in investigations of the rotational dynamics of minor planetary satellites, where the Lyapunov times are small enough for the chaotic character of dynamics to manifest itself on the intervals of time accessible to observations. The presence of chaos can be as well identified in the dynamics of comets, if, as in case of the Halley comet, historical chronicles about the dates of comet's returnings are available. As a rule, highly chaotic is the motion of the planet-crossing asteroids and comets. Judging by the Lyapunov times, chaos (if it is present) is also observable in the motion of some satellite systems and in the inner orbital dynamics of multiple asteroids.

In the dynamics of many other bodies of the Solar system, chaos is obscure, due to either too large Lyapunov times at nowadays or to remoteness of the time epochs when the motion was strongly chaotic, but it is important to remember that in the long-term evolution its role is great and it imprints and sculptures the present appearance and structure of the Solar system, e. g. in the form of Kirkwood gaps in the main belt of asteroids.

References

Bretagnon, P. 1990, A&A, 231, 561

Cachucho, F., Cincotta, P. M., & Ferraz-Mello, S. 2010, Celest. Mech. Dyn. Astron., 108, 35

Celletti, A., & Chierchia, L. 2000, Celest. Mech. Dyn. Astron., 76, 229

Champenois, S., & Vienne, A. 1999a, Icarus, 140, 106

Champenois, S., & Vienne, A. 1999b, Celest. Mech. Dyn. Astron., 74, 111

Chauvineau, B., Farinella, P., & Mignard, F. 1993, Icarus, 105, 370

Chirikov, B.V. 1959, Atomnaya Energiya, 6, 630 (in Russian)

Chirikov, B.V. 1979, Phys. Rep., 52, 263

Chirikov, B.V. 1982, Priroda, No. 7, 15 (in Russian)

Chirikov, B.V., & Shepelyansky, D.L. 1984, Physica D, 13, 395

Chirikov, B.V., & Vecheslavov, V.V. 1986, INP Preprint, 86–184

Chirikov, B.V., & Vecheslavov, V.V. 1989, A&A, 221, 146

Dermott, S.F., Malhotra, R., & Murray, C.D. 1988, Icarus, 76, 295

Farmer, A.J., & Goldreich, P. 2006, Icarus, 180, 403

Ferraz-Mello, S., & Klafke, J.C. 1991, in: Predictability, Stability and Chaos in N-body Dynamical Systems, ed. by A.E.Roy, New York: Plenum Press, 177

Froeschlé, Cl. 1984, Celest. Mech., 34, 95

Giorgilli, A., & Lazutkin, V.F. 2000, Phys. Lett., A 272, 359

Gladman, B., Migliorini, F., Morbidelli, A., *et al.* 1997, Science, 277, 197

Goldreich, P., & Rappaport, N. 2003a, Icarus, 162, 391

Goldreich, P., & Rappaport, N. 2003b, Icarus, 166, 320

Hairer, E., Nørsett, S.P., & Wanner, G. 1987, Solving Ordinary Differential Equations I. Nonstiff Problems, Berlin: Springer-Verlag

Hayes, W.B. 2007, Nat. Phys., 3, 689

Hayes, W.B. 2008, MNRAS, 386, 295

Hayes, W.B., Malykh, A.V., & Danforth, C.M. 2010, MNRAS, 407, 1859

Holman, M.J., & Murray, N.W. 1996, AJ, 112, 1278

Kapitsa, P.L. 1951, Zh. Exp. Teor. Fiz., 21, 588 (in Russian)

Kapitsa, P.L. 1954, Uspekhi Fiz. Nauk, 44, 7 (in Russian)

Klavetter, J.J. 1989, AJ, 98, 1855

Knežević, Z., & Milani, A. 2000, Preprint

Kouprianov, V.V., & Shevchenko, I.I. 2003, A&A, 410, 749

Kouprianov, V.V., & Shevchenko, I.I. 2005, Icarus, 176, 224

Laskar, J. 1989, Nature, 338, 237

Lichtenberg, A.J., & Lieberman, M.A. 1992, Regular and Chaotic Dynamics, New York: Springer-Verlag

Melnikov, A.V. 2002, IAA Transactions, 8, 131

Melnikov, A.V., & Shevchenko, I.I. 2005, Sol. Sys. Res., 39, 322 [Astronomicheskii Vestnik, 39, 364]

Melnikov, A.V., & Shevchenko, I.I. 2008, Celest. Mech. Dyn. Astron., 101, 31

Milani, A., & Nobili, A.M. 1993, Celest. Mech. Dyn. Astron., 56, 323

Milani, A., Nobili, A.M., & Knežević, Z. 1997, Icarus, 125, 13

Moons, M., & Morbidelli, A. 1995, Icarus, 114, 33

Moons, M., Morbidelli, A., & Migliorini, F. 1998, Icarus, 135, 458

Morbidelli, A. 2002, Modern Celestial Mechanics, London: Taylor & Francis

Morbidelli, A., & Nesvorný, D. 1999, Icarus, 139, 295

Murray, C.D., & Dermott, S.F. 1999, Solar System Dynamics, Cambridge: Cambridge University Press

Murray, N.W., & Holman, M.J. 1997, AJ, 114, 1246

Murray, N., & Holman, M. 1999, Science, 283, 1877

Murray, N., Holman, M., & Potter, M. 1998, AJ, 116, 2583

Nesvorný, D., & Morbidelli, A. 1998, AJ, 116, 3029

Nesvorný, D., & Morbidelli, A. 1999, Celest. Mech. Dyn. Astron., 71, 243

Pannekoek, A. 1961, A History of Astronomy, London: George Allen and Unwin Ltd.

Petit, J.-M., Durda, D.D., & Greenberg, R., *et al.* 1997, Icarus 130, 177

Petrosky, T.Y. 1986, Phys. Lett., A 117, 328

Poincaré, H. 1899, Les Méthodes Nouvelles de la Mécanique Céleste III, Paris: Gauthier-Villars

Shevchenko, I.I. 1998a, Phys. Lett., A 241, 53

Shevchenko, I.I. 1998b, Phys. Scr., 57, 185

Shevchenko, I.I. 1999, Celest. Mech. Dyn. Astron., 73, 259

Shevchenko, I.I. 2000, J. Exp. Theor. Phys., 91, 615 [ZhETP, 118, 707]

Shevchenko, I.I. 2002a, Cosmic Res. 40, 296 [Kosmich. Issled., 40, 317]

Shevchenko, I.I. 2002b, in: Asteroids, Comets, Meteors 2002, ed. by B.Warmbein, Berlin: ESA, 367

Shevchenko, I.I. 2004a, Phys. Lett., A 333, 408

Shevchenko, I.I. 2004b, JETP Lett., 79, 523 [Pis'ma Zh. Eksp. Teor. Fiz. 79, 651]

Shevchenko, I.I. 2007a, in: Astronomy: Its Tradition, Present and Future, Saint Petersburg: Saint Petersburg Univ., 284 (in Russian)

Shevchenko, I.I. 2007b, in: Near Earth Objects, Our Celestial Neighbors: Opportunity and Risk (IAU Symp. 236), eds A.Milani *et al.*, Cambridge: Cambridge Univ. Press, 5

Shevchenko, I.I. 2007c, Physica, A 386, 85

Shevchenko, I.I. 2008a, Phys. Lett., A 372, 808

Shevchenko, I.I. 2008b, MNRAS, 384, 1211

Shevchenko, I.I. 2010, Priroda, No. 4, 12 (in Russian)

Shevchenko, I.I. 2011, ApJ, 733, 39

Shevchenko, I.I., & Kouprianov, V.V. 2002, A&A, 394, 663

Shevchenko, I.I., Kouprianov, V.V., & Melnikov, A.V. 2003, Sol. Sys. Res., 37, 74 [Astronomicheskii Vestnik, 37, 80]

Stephenson, A. 1908, Manchester University Memoirs, v. LII, No. 8, March 5th, 1

Sussman, G.J., & Wisdom, J. 1988, Science, 241, 433

Sussman, G.J., & Wisdom, J. 1992, Science, 257, 56

Tittemore, W., & Wisdom, J. 1988, Icarus, 74, 172

Tittemore, W., & Wisdom, J. 1989, Icarus, 78, 63

Valsecchi, G.B. 2007, in: Near Earth Objects, Our Celestial Neighbors: Opportunity and Risk (IAU Symp. 236), eds A.Milani *et al.*, Cambridge: Cambridge Univ. Press, xvii

von Bremen, H.F., Udwadia, F.E.S, & Proskurowski, W. 1997, Physica, D 101, 1

Whipple, A.L. 1995, Icarus, 115, 347

Wisdom, J. 1983, Icarus, 56, 51

Wisdom, J., Peale, S.J., & Mignard, F. 1984, Icarus, 58, 137

LECTURE

The periodic and chaotic regimes of motion in the exoplanet 2/1 mean-motion resonance

T.A. Michtchenko[1], S. Ferraz-Mello[1], C. Beaugé[2]

(1) Instituto de Astronomia, Geofísica e Ciências Atmosféricas, USP, São Paulo, Brazil
(2) Observatorio Astronómico, Universidad Nacional de Córdoba, Córdoba, Argentina

Abstract.   We present the dynamical structure of the phase space of the planar planetary 2/1 mean-motion resonance (MMR). Inside the resonant domain, there exist two families of periodic orbits, one associated to the librational motion of the critical angle (σ-family) and the other related to the circulatory motion of the angle between the pericentres ($\Delta\varpi$-family). The well-known apsidal corotation resonances (ACR) appear at the intersections of these families. A complex web of secondary resonances exists also for low eccentricities, whose strengths and positions are dependent on the individual masses and spatial scale of the system.

Depending on initial conditions, a resonant system is found in one of the two topologically different states, referred to as *internal* and *external* resonances. The internal resonance is characterized by symmetric ACR and its resonant angle is $2\lambda_2 - \lambda_1 - \varpi_1$, where λ_i and ϖ_i stand for the planetary mean longitudes and longitudes of pericentre, respectively. In contrast, the external resonance is characterized by asymmetric ACR and the resonant angle is $2\lambda_2 - \lambda_1 - \varpi_2$. We show that systems with more massive outer planets always envolve inside internal resonances. The limit case is the well-known asteroidal resonances with Jupiter. At variance, systems with more massive inner planets may evolve in either internal or external resonances; the internal resonances are typical for low-to-moderate eccentricity configurations, whereas the external ones for high eccentricity configurations of the systems. In the limit case, analogous to Kuiper belt objects in resonances with Neptune, the systems are always in the external resonances characterized by asymmetric equilibria.

1.   Introduction

Asteroids and exoplanets are amongst the most striking sources of problems in Celestial Mechanics where chaos and order play an important role. Asteroids have always puzzled astronomers by their peculiar distribution: in some regions, they avoid resonances, but in others, they prefer to be exactly where the resonances are found. The famous problem of the Kirkwood gaps, solved in the last quarter of the past century, is the better example. Asteroids have benefited from more than one century accurate observations and the number of aster-

oids whose orbits are known with high precision is presently of order of many thousands. Accurate observations allowed us to map the phase space where the asteroids evolve and obtain the boundaries of regular and chaotic motions. In addition, the increasing computer capacity allowed us to understand why they are distributed in such a way.

Exoplanets on the other hand are being observed at the limit of our technical capabilities. We are far from having accurate knowledge of their orbits. We know presently more than 700 planets. Most of them are in systems composed of only one planet. But the number of multi-planet systems is increasing and may reach its first hundred in a short time. Besides, what is important, they show a striking diversity. They range from the small systems of super-Earths and mini-Neptunes, as those discovered by the space missions CoRoT and Kepler in circular orbits very close to their central stars, to systems of giant planets in very elongated elliptic orbits reaching enormous distances from their host stars, as far as the outskirts of our Solar System. Eccentricities larger than 0.5 are not infrequent: in multi-planet systems they reach up to 0.75 and in single-planet systems they reach up to 0.97.

Many multi-planet extrasolar systems exhibit resonant behavior. Systems of resonant planets is a novelty for the Celestial Mechanician. We were used to study the resonant problems of small objects, whose mass can be assumed as zero (restricted three-body problem), or a few planetary satellites in which restricted models cannot be used but which are dominated by the oblateness of the central planet. So, exoplanets are a prime source for cases of the three-body problem in which we cannot neglect the mass of the bodies and in which the mutual point-like gravitational interaction is by far the dominant force. The most frequent resonance found is the 2/1 mean-motion resonance (e.g. GJ 876 **c-b**, HD 40307 **c-d**, HD 73526 **b-c**, HD 82943 **c-b** and HD 128311 **b-c**).

Although other resonance sites may also be inhabited (e.g. HD 60532 **b-c** in the 3/1 or HD 45364 **b-c** in the 3/2 resonance), the 2/1 is presently the most populated. Notwithstanding the limited accuracy of the observations, resonant configurations have at least two strong arguments in their favor: The long-term stable motion of close planets in high eccentricity configurations is possible only if the planets are locked in (and protected by) MMRs. Secondly, resonance trapping appears to be a natural outcome of planetary migration processes due to planet–disk interactions, which are believed to take place in the latest stage of the planet system formation (e.g. Kley 2000; Snellgrove *et al.* 2001; Kley *et al.* 2005).

The classical studies of the general three-body problem with their averaged Hamiltonians have been quickly extended to the case of exoplanets (Hadjidemetriou 2002; Lee & Peale 2002; Beaugé & Michtchenko 2003; Ferraz-Mello *et al.* 2003). The periodic solutions, dubbed as corotation resonances or apsidal corotation resonances (ACR), were obtained to cover all possible families in low-order resonances (Lee 2004; Beaugé *et al.* 2006; Michtchenko *et al.* 2006b, Giuppone *et al.* 2010). ACR are stationary states of the resonant Hamiltonian averaged over the synodic period and correspond to equilibrium solutions of the averaged equations of motion.

Notwithstanding the attention devoted to ACR and periodic orbits, not much was known on the topology of the phase space outside their vicinity. A

detailed analysis was presented first by Michtchenko & Ferraz-Mello (2001) for the 5/2 MMR, by Callegari *et al.* (2004) for the 2/1 MMR and by Callegari *et al.* (2006) for the 3/2 MMR. However, these works were limited to the study of a few specific systems and are only valid for very small eccentricities. Even so, results showed the complex structure of the phase space, populated by several different families of periodic orbits, modes of oscillation and possible regimes of motion. ACR appear to be only one of several distinct types of stable configurations.

It was necessary to extend our knowledge of the resonant dynamics of the MMR beyond the ACR. This has been done for the 2/1 resonance in two papers published by Michtchenko *et al.* (2008 a,b), for arbitrary mass ratios and with no restrictions in the orbital eccentricities, providing a complete map of the phase space around the ACRs. These results, summarized in this paper, help us to understand the results of numerical simulations and pinpoint where future (or even currently known) extrasolar systems may be found. Even if all planetary systems in this commensurability seem to be restricted to the very close vicinity of ACR, the maps allow to see the possible evolutionary routes inside the 2/1 MMR from initially non-resonant configurations.

2.   The model of the resonant three-body problem

When the ratio of the orbital periods of two planets is close to a ratio of two simple integers, we say that the planets are close to a mean-motion resonance (MMR). A mean-motion resonance is often written in a generic form as

$$n_1/n_2 = (p+q)/p,$$

where n_i are the planetary mean motions and p and q are integers, the latter one representing the order of the resonance. The critical (or resonant) angles σ_i are defined as

$$
\begin{aligned}
\sigma_1 &= (1+p/q)\lambda_2 - (p/q)\lambda_1 - \varpi_1, \\
\sigma_2 &= (1+p/q)\lambda_2 - (p/q)\lambda_1 - \varpi_2,
\end{aligned}
\tag{1}
$$

where λ_i and ϖ_i are mean longitudes and longitudes of pericentre of the planets; index 1 denotes the inner body, while index 2 is reserved for the outer body. The behavior of the critical angles defines the location of the system with respect to the resonance: when one of these angles is librating, the system is said to be *inside* the resonance. It is worth noting that the secular angle defined as a difference in longitudes of pericentre of the planets, is $\Delta\varpi = \sigma_1 - \sigma_2$.

The dynamics of two resonant planets, with masses m_1 and m_2, orbiting a star of the mass m_0 is defined by the averaged Hamiltonian $\overline{\mathcal{H}}_{\rm res}$ and two integrals of motion, $\mathcal{AM}$ and $\mathcal{K}$, where the first is the total angular momentum and the second is the so-called spacing parameter. These analytical functions are given by the expressions, up to second order in masses:

$$
\begin{aligned}
\overline{\mathcal{H}}_{\rm res} &= -\sum_{i=1}^{2} \frac{Gm_0 m_i}{2\,a_i} - \frac{1}{2\pi}\int_0^{2\pi} \mathcal{R}(a_i, e_i, \sigma_i, Q)\,dQ, \\[2mm]
\mathcal{AM} &= m_1\,n_1\,a_1^2\,\sqrt{1-e_1^2} + m_2\,n_2\,a_2^2\,\sqrt{1-e_2^2}, \\[2mm]
\mathcal{K} &= (p+q)\,m_1\,n_1\,a_1^2 + p\,m_2\,n_2\,a_2^2,
\end{aligned}
\tag{2}
$$

where $\mathcal{R}$ is the disturbing function. All orbital elements, including the semi-major axes a_i and eccentricities e_i, are canonical astrocentric variables; their relation to the usual osculating heliocentric orbital elements can be found in (Ferraz-Mello *et al.* 2006). The averaging of the disturbing function $\mathcal{R}$ is done with respect to the synodic angle $Q = \lambda_2 - \lambda_1$. In the vicinity of the MMR, Q is much faster than the resonant and secular angles, and does not influence significantly the long-term evolution of the system. Thus, all periodic terms dependent on Q can be eliminated (i.e. averaged out) of the Hamiltonian function, and only secular and resonant terms need to be retained.
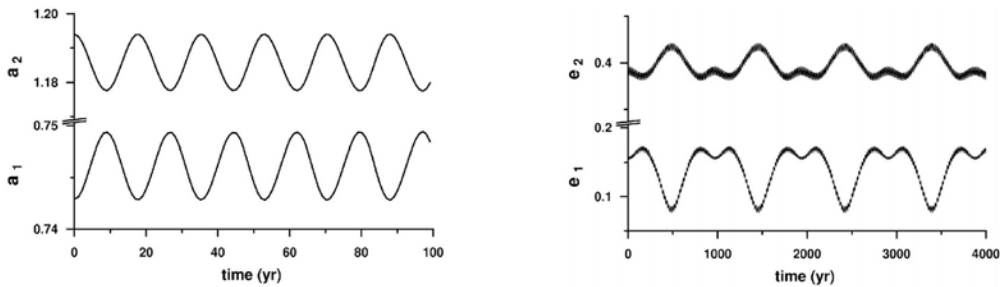
Figure 1. Numerical simulation of two planets displaying large-amplitude librations around a $(0,0)$-ACR, with masses and initial conditions drawn from Fit B of Ferraz-Mello *et al.* (2005) for the HD 82943 system. Plot shows the resonant evolution of the semimajor axes (left panel) and eccentricities (right panel), averaged over short-term oscillations. It is worth comparing the resonant oscillation timescale (left panel) to the secular one (right panel).

The invariance of $\mathcal{AM}$ and $\mathcal{K}$ has important consequences for the orbital evolution of the system. It indicates that, after the averaging process, from the set of the variables (a_1, a_2, e_1, e_2), only two are independent, and the planar resonant problem has two degrees of freedom. Each degree of freedom is characterized by its proper mode of motion. Note that $\mathcal{K}$ depends only upon masses and semimajor axes; therefore, the invariance of the spacing parameter $\mathcal{K}$ defines a coupling of the semimajor axes: they oscillate with opposite phases and with amplitudes that are inversely proportional to the planetary mass. We say that the semimajor axes have no secular variation and their evolution occurs in timescales associated with the *resonant mode* of motion. This effect can be seen in Figure 1 (left panel).

On the other hand, the integral of the total angular momentum $\mathcal{AM}$ causes a coupling of the planetary eccentricities, in such a way that, when one eccentricity grows, the other decreases, as seen in Figure 1 (right panel). The oscillation generally occurs in timescales much longer when compared to those of the resonant mode. This mode is defined as *secular mode* of motion of the system and is associated with the secular angle $\Delta\varpi$. $\mathcal{AM}$ is a function of both semimajor axes and eccentricities, that implies that both modes of motion, resonant and secular, are present in eccentricity variations. In the example shown on the right panel in Figure 1, the amplitudes of the resonant component in the eccentricity variations are much smaller (and frequencies are much higher) than those of the secular mode. However, as will be shown below, for some initial conditions, both amplitudes and frequencies can be comparable.

3. Stationary orbits of the averaged 2/1 resonant problem

The Hamiltonian function given in Eq.(2) is very complicated, even in the planar case, and generally possesses several extrema, which define stationary solutions of the averaged problem. These solutions are often referred to as *Apsidal Corotation Resonances* (ACR). To obtain these special solutions over a large domain of the parameters of the problem, we employ the geometrical method presented in Michtchenko et al. (2006b), where each equilibrium solution is identified looking at local maxima of the Hamiltonian function, for given values of the total angular momentum $\mathcal{AM}$ and the spacial parameter $\mathcal{K}$. The location of the global maxima of the 2/1 resonance Hamiltonian calculated for all possible values of $\mathcal{AM}$ and a fixed $\mathcal{K}$, are shown on the plane (e_1, e_2) in Figure 2, in the form of families parameterized by the mass ratio m_2/m_1.
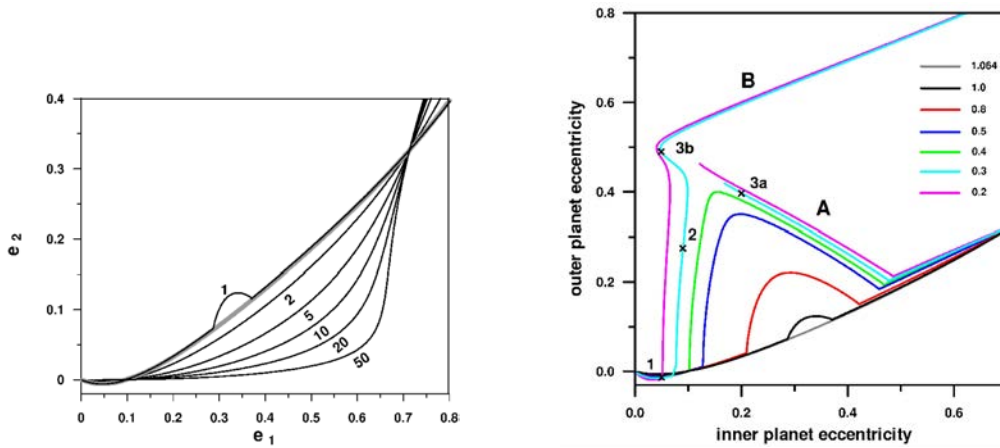

Figure 2.   The 2/1 resonance stable apsidal corotation solutions parameterized by the mass ratio m_2/m_1. The cases with $m_2/m_1 \geq 1$ are shown on the left panel, while those with $m_2/m_1 \leq 1$ on the right panel. The mass ratio 1.064 corresponding to the HD 82943 system is illustrated by a gray curve.

Usually, ACR are classified in two types: symmetric and asymmetric solutions (e.g. Beaugé *et al.* 2003, Lee 2004, Voyatzis & Hadjidemetriou 2005). The symmetric solutions are characterized by stationary values of both critical angles (1) at 0 or 180°. The symmetric ACR–solutions evolve monotonically with the increasing eccentricities of the planets. For instance, for $m_2/m_1 \geq 1.015$ (Ferraz-Mello *et al.* 2003), all ACR families are symmetric; some are plotted on the left graph in Figure 2.

For mass ratios smaller than this critical value, the smooth evolution along a symmetric family is interrupted by a sudden increase of the outer planet's eccentricity (the case of the family with $m_2/m_1 = 1$ on the left graph). The corresponding symmetric solutions become unstable and the stationary systems evolve now along an asymmetric segment of the corresponding family. Asymmetric ACR are characterized by stationary values of both critical angles different from zero or 180°. The families shown on the right graph in Figure 2 exhibit this feature.

A second critical value of the mass ratio occurs at $\simeq 0.36$ and is illustrated by cyan and magenta curves in Figure 2 (right panel). Below this limit, asymmetric ACR–families split into two disconnected branches on the (e_1, e_2)–plane, marked by labels **A** and **B**. In these cases, there may exist two distinct stationary configurations described by different sets of orbital elements, but leading to the same values of total angular momentum and the spacing parameter.

We analyze the evolution of the proper periods (inverse of the proper frequencies) along each ACR–family, employing a method of dynamic power spectra (Michtchenko *et al.* 2006a). The left graph in Figure 3 shows, in logarithmic scale, the period of the resonant proper mode of motion (T_σ) and the period of the secular mode ($T_{\Delta\varpi}$), both obtained along the ACR-family parameterized by the mass ratio of the HD 82943 system. The ratio between both periods, plotted by a dashed line on the same panel, indicates that the characteristic times of the resonant component of motion are generally shorter (at least in order 1) than the secular ones. Both periods are comparable solely for very small eccentricities when, for some initial configurations of the planets, they can be commensurable originating a complex structure of *secondary resonances* inside the primary 2/1 resonance.
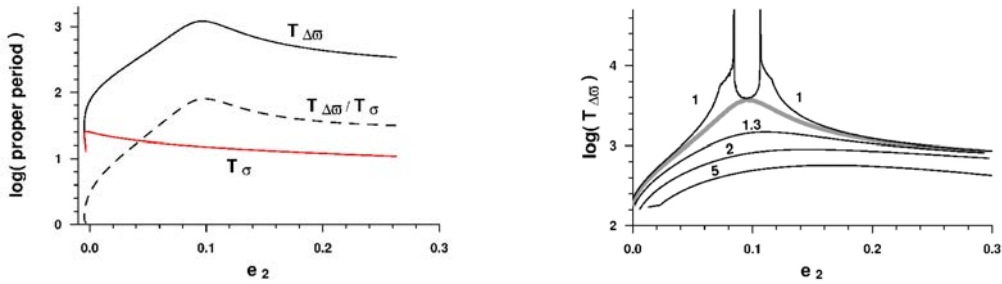

Figure 3.   Left: Evolution of the period T_σ of the resonant mode of motion and the period $T_{\Delta\varpi}$ of the secular mode along the ACR family parameterized by $m_2/m_1 = 1.064$, with the individual planet masses of the HD 82943 system. The ratio $T_{\Delta\varpi}/T_\sigma$ is shown by dashed line. Right: Evolution of the secular period $T_{\Delta\varpi}$ along several ACR families. The values of the mass ratio are indicated by numbers.

The right graph in Figure 3 shows the secular period, in logarithmic scale, as a function of the outer planet eccentricity obtained along several ACR–families with different mass ratios. For large values of m_2/m_1, there is a little change in the period with a maximum value at e_2 roughly between 0.1 and 0.2. However, for decreasing values of m_2/m_1, the maximum becomes more pronounced until it reaches a singularity near $e_2 \simeq 0.1$, for the mass ratio $m_2/m_1 \cong 1.015$. In Figure 3 (right panel) we show discontinuities which appear along the ACR–family parameterized by $m_2/m_1 = 1$. A comparison with the plot of Figure 2 shows that these discontinuities are associated to the bifurcation of the symmetric ACR leading to the origin of the asymmetric stable solutions. The existence of the singularities indicate that the difference between symmetric and asymmetric ACR is not restricted to the equilibrium values of the angles, but constitutes qualitatively distinct solutions, each belonging to separate regions of the phase space.
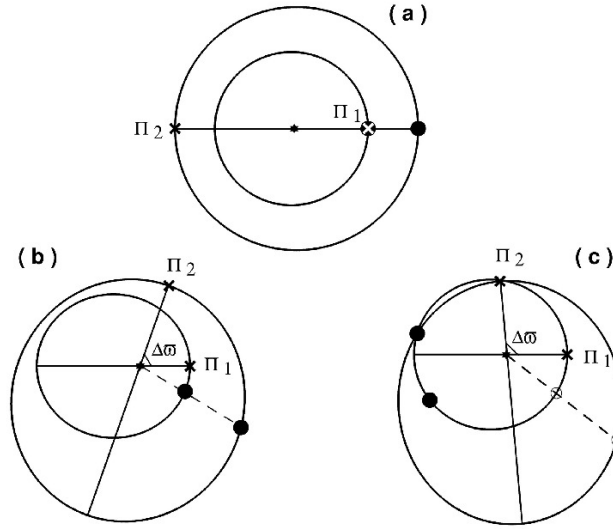
Figure 4. Examples of stable symmetric **(a)** and asymmetric **(b)** and **(c)** stationary configurations of two planets in the 2/1 resonance, with the mass ratio $m_2/m_1 = 0.3$. The continuous lines are apsidal lines, while the dashed lines are conjunction lines (in the case **(a)** two lines coincide). The full circles show the positions of the planets on the orbits at the instant of the closest approach between them.

4. Geometry of the stationary orbits

Figure 4 shows three different configurations of the ACR–solutions of the 2/1 MMR, obtained for $m_2/m_1 = 0.3$. The case **(a)** illustrates a symmetric solution with anti-aligned pericentre lines. In this case, the closest approach between two planets occurs when the planets are in conjunction: the inner planet in the pericentre and the outer planet in the apocentre of their orbits ($\Delta\varpi = 180°$).

Two asymmetric ACR-solutions are shown in Figure 4 (b) and (c). Both configurations are characterized by the same set of the constants $\mathcal{A}M$ and $\mathcal{K}$, but the solution **(b)** belongs to the branch **A** of the ACR–family, while the solution **(c)** corresponds to the branch **B** in Figure 2 (right panel). The secular angle $\Delta\varpi$ defines the relative orientation of the pericentre lines of two orbits. The positions of the planets on their orbits are strongly tied by the 2/1 resonance relationship between the planetary mean anomalies which are related to the critical angles through $M_1 = \sigma_1 + (p + q)\,Q$ and $M_2 = \sigma_2 + p\,Q$, where Q is the synodic angle and $p = q = 1$ (Beaugé & Michtchenko 2003). Full circles in Figure 4 show positions of the planets, for which their mutual distance is minimal.

In the configuration **(b)**, when the orbits are not intersecting, the closest approaches occur when the planets are in conjunction, forming a co-linear configuration with the central star. It is interesting to observe that the planetary conjunctions avoid positions between the two pericentra. As a consequence, conjunctions always occur when two planets are far away from positions where the two orbits are very close one to another; this feature can be seen as a protection mechanism due to the 2/1 mean-motion resonance.

In the case (**c**), two orbits are intersecting and the mutual distance is not more minimal when the planets (open circles) are in conjunction. The closest approaches occur when the central star and the two planets (full circles) are in a triangle configuration and the outer planet is localized close to the intersection of the two orbits. In this case, again, the protection mechanism prevents very close approaches between two planets (see also Lee *et al.* 2006, for similar analysis).

5. Two regimes of resonant behavior: interior and exterior resonances

In the previous sections, we have seen that the resonant behavior is actually composed of two proper modes of motion, the resonant and secular ones. It is generally characterized by the libration of one angle and the circulation of the other. We refer to the librating angle as the *resonant angle* and to the oscillation/circulation angle as the *secular angle*. The secular theories provide us with the secular angle: it is defined by the difference in longitudes of pericentre $\Delta\varpi = \varpi_2 - \varpi_1$ or, from our definitions (1), as $\Delta\varpi = \sigma_1 - \sigma_2$. The last expression shows that only one of two critical angles defined in Eq. (1) is independent and is a truly resonant angular variable of the problem (the other independent angular variable is $\Delta\varpi$). The question which rises now is which one of two critical angles can be chosen as a resonant variable.

When the outer body is more massive ($m_2/m_1 > 1.015$), the situation is similar to that of an asteroid evolving in a resonance with Jupiter, when the librating angle is σ_1. In this case, all ACR-solutions are symmetric. Thus, we choose the critical angle σ_1 as a resonant angular variable of the problem and adopt the name *interior resonance* for this regime of motion.

For the planets with the mass ratios $m_2/m_1 < 1.015$, both kinds of ACR-solutions are possible, symmetric and asymmetric ones. If $m_2/m_1 << 1$, which is the case of a Kuiper belt object in resonance with Neptune, the asymmetric solutions are dominating. Therefore, in the case of asymmetric ACR, we choose the critical angle σ_2 as a resonant angular variable of the problem. We refer to the regime of motion characterized by asymmetric ACR as *exterior resonance.*

The behavior of the secular angle $\Delta\varpi$ has also some special characteristics. In both interior and exterior resonances, it may be either a circulation or an oscillation. However, the oscillation in this case should be differentiated from the regime of motion defined as a libration in Celestial Mechanics. Indeed, the motions are akin to a family of concentric curves around one center displaced from the origin; the curves which are close to the center, do not enclose the origin – they correspond to oscillations, while the outer curves enclose the origin and correspond to circulations. The separation between them is not a dynamical separatrix, but just one curve passing through the origin. The whole set of curves forms a homeomorphic family of solutions and the distinction between oscillations and circulations in this case is merely kinematical. To stress this behavior we use the word *oscillatiory/circulatory*, when describing this regime of motion.

6. Dynamics around the ACR solutions: Interior resonance

Independent of the mass ratio, the dynamics associated to all symmetric ACR is qualitatively similar, showing interior resonances. To portray the phase space of the interior resonance, we introduce a symmetric representative plane $(n_1/n_2, e_2)$, with the resonant angle σ_1 fixed at 0 and the secular angle $\Delta\varpi$ at 0 (positive values on the e_2-axis) or 180° (negative values on the e_2-axis). A typical example of the dynamical map of the phase space around one symmetric ACR is shown in Figure 5.

The map contains one center, which represents the ACR solution; its position shown by a red circle. Around the center, we find a region of quasi-periodic motion, coded in gray scale on the dynamical maps. Light gray color indicates regular motion, while darker tones correspond to increasingly chaotic motions.
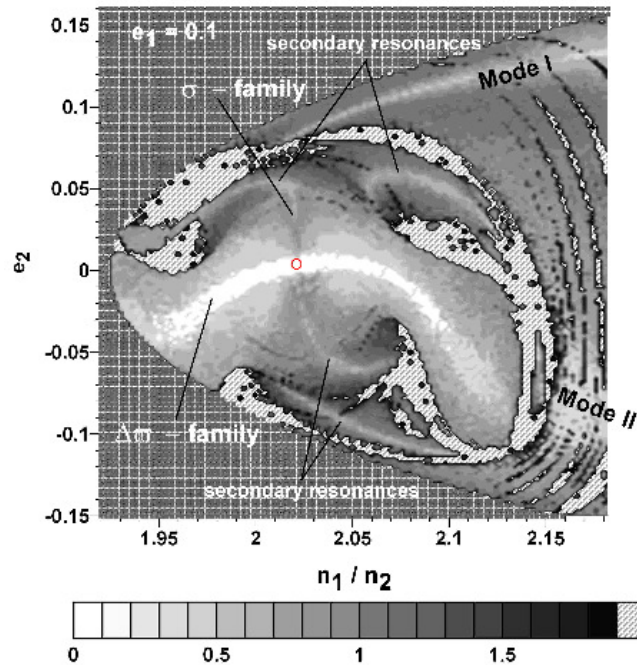

Figure 5.   Dynamical map of the domain around the stable symmetric $(0,0)$–ACR (red dot) with $e_1 = 0.1$ and $m_2/m_1 = 1.064$. The initial value of σ_1 is fixed at 0 and the initial value of $\Delta\varpi$ is fixed at 0 (positive values on the e_2-axis) or 180° (negative values on the e_2-axis).

The quasi-periodic behavior is the composition of two independent modes of motion: the resonant (with the frequency f_σ) and the secular (with the frequency $f_{\Delta\varpi}$) ones. The first one is associated to the resonant angle σ_1, while the second one to the difference in longitudes of pericentre, $\Delta\varpi$. Thus, any regular solution will be given by a combination of two periodic terms and their harmonics, each with a given amplitude and phase angle. Generally, both modes are well separated in the frequency space, with f_σ much higher than the secular frequency $f_{\Delta\varpi}$ (see Figure 3 left panel). However, at small eccentricities, both

frequencies may have same order, and it is possible to find initial conditions corresponding to low-order commensurabilities between them. They give origin to secondary resonances inside the primary 2/1 resonance, whose locations on the representative plane are indicated in Figure 5.

For some initial conditions, the amplitude (not the frequency) of one mode tends to zero, and the solution changes from quasi-periodic to periodic. Since we have two independent frequencies, we can also have two independent families of periodic motion inside the 2/1 resonance, which we refer to as σ-*family* (the amplitude of the resonant mode is equal to zero) and $\Delta\varpi$–*family* (the amplitude of the secular mode is equal to zero). On the dynamical map shown in Figure 5, the periodic families appear as continuous narrow white strips inside the gray tone domains of quasi-periodic motion. By definition, the intersection of two families gives us the position of the central ACR solution.

Apart from the characteristic frequencies, the resonant and secular modes also differ in dynamical behavior. The resonant mode is always librating and the passage to circulation must occur at a separatrix, which gives origin to domains of highly chaotic motion, with possible disruption of the system. At variance, the secular mode is generally oscillating/circulating; only in the domains of very high eccentricities it is possible to find initial conditions which lead to a true secular resonance inside the 2/1 MMR (see Section 7.2.).

The domains of chaotic motion (and large instabilities) are always present on dynamical maps of the 2/1 resonance. The chaotic behavior is associated with the existence of separatrices between: *(i)* the 2/1 resonance region and the regions of near-resonant and purely secular motion; *(ii)* the regions of qualitatively distinct regimes of motion inside the 2/1 mean-motion resonance, and *(iii)* secondary resonances inside the primary 2/1 resonance. The domains of highly chaotic motion are always shown in dark tones on the dynamical maps, while the hatched regions are regions of large-scale instabilities followed by disruption of the system within the time interval of each simulation (130,000 years). Finally, the domains of forbidden motion are filled by dark gray color in Figure 5.

7.   Dynamics around the ACR solutions: Exterior resonance

The dynamics associated to all asymmetric ACR is characteristic of exterior resonances. Asymmetric stationary solutions arise for mass ratios m_2/m_1 lesser than 1.015 (see Figure 2, right panel). Moreover, for $m_2/m_1 < 0.36$ and at high eccentricities, each asymmetric family bifurcates into two disconnected branches on the (e_1, e_2)–plane, referred to as branches **A** and **B** in our work. The ramification of the asymmetric families indicates the advent of the true secular resonance inside the mean-motion resonance. As a consequence, the phase space of the exterior resonance shows a very complex picture marked by the presence of several distinct regimes of resonant and non-resonant motion, crossed by families of periodic orbits and separated by chaotic zones.

7.1.   Dynamical map around the one-branch asymmetric ACR

One-branch asymmetric solutions are characteristic for the mass ratio values in the range from 0.36 to 1.015. The dynamics around a single asymmetric ACR is illustrated in Figure 6 on the two representative planes: the left graph is a plane
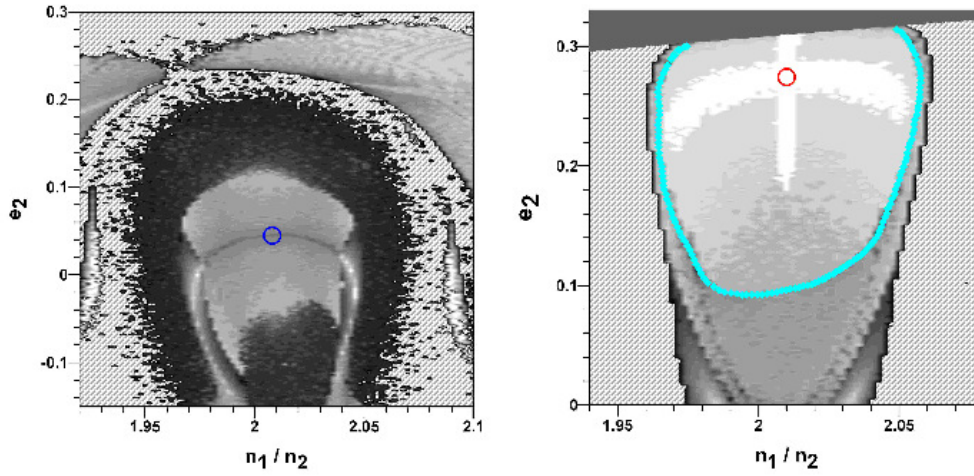
Figure 6.   Left panel: Dynamical map on the representative plane obtained for σ_1 fixed at 0 and $\Delta\varpi$ at 0 (positive values of the e_2-axis) or 180° (negative values of the e_2-axis). The position of the unstable ACR is shown by a blue circle. Right panel: Dynamical map on the asymmetric representative plane obtained with initial values of the angular variables equal to those an asymmetric ACR: $\sigma_1 = 45.4°$ and $\sigma_2 = -68.4°$, for $m_2/m_1 = 0.3$. The position of the stable ACR is shown by a red circle.

analogous to one introduced for the interior resonance and shown in Figure 5; the right graph is a plane defined by the initial values of the resonant angles fixed exactly at those of the stable asymmetric stationary solution.

The plane defined by symmetric resonant angles does not allow us to fully describe the dynamics of the exterior resonance. The basic properties of the resonant phase space, such as the existence of the intersecting families of periodic orbits and, thus, of the stable centers, are not seen on the resulting dynamical map. However, the analysis of this map allows us to detect the co-existence of the two regimes of motions named as interior resonance and exterior resonance for the same set of the constants m_2/m_1, $\mathcal{AM}$ and $\mathcal{K}$. Note the $\Delta\varpi$-families of symmetric periodic solutions (with $\Delta\varpi = 0$ or 180°), which appear as narrow light strips inside the chaotic regions on the map in Figure 6, left panel.

To better represent the dynamics in the neighborhood of the asymmetric centers, we re-calculated the dynamical map for the same sets of constants of motion, but with the initial values of the resonant angles fixed exactly at those of the stable asymmetric stationary solution (right graph in Figure 6). Now we can observe the existence of the stable center at the intersection of the two periodic families (white strips), σ and $\Delta\varpi$. The stable domain surrounding the center (in light tones), is composed of quasi-periodic motions with two independent modes: resonant libration around the σ–family and secular oscillation/circulation around the $\Delta\varpi$–family. Generally, the global dynamics of the single exterior resonance is similar to the structure of the interior resonance presented in the previous section. However, there are two main differences: *(i)* the librating angle is σ_2 and the oscillating/circulating angle is σ_1 (and $\Delta\varpi$), and *(ii)* the angles oscillate around values that are not simply 0 or 180°.
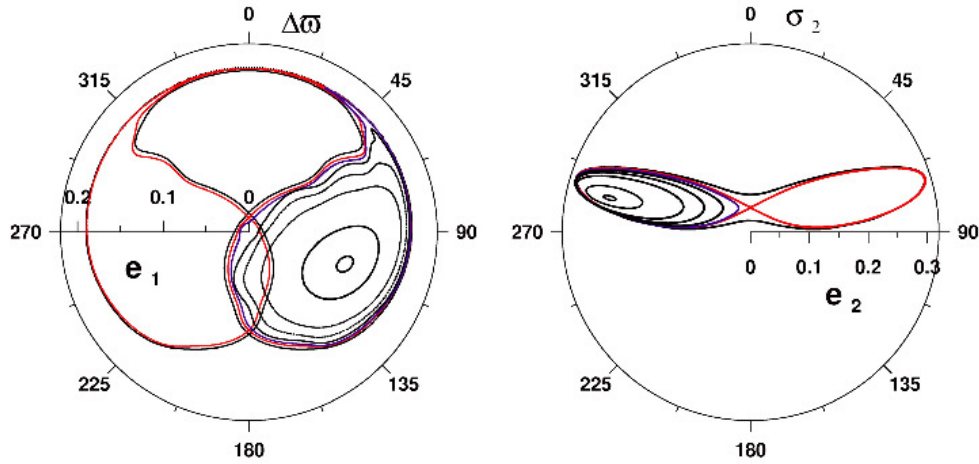
Figure 7.   Smoothed planetary paths around one-branch asymmetric ACR in the $(e_1, \Delta\varpi)$ (left panel) and (e_2, σ_2) (right panel) polar coordinates calculated with $e_1 = 0.09$ and $m_2/m_1 = 0.3$ and the same energy. The smoothing has been done using a low-pass filter.

The planetary motions around a stable asymmetric ACR are shown in Figure 7, in the polar coordinates $(e_1, \Delta\varpi)$ on the left graph and in the polar coordinates (e_2, σ_2) on the right graph. Only the secular components are presented; the fast resonant oscillations were eliminated using a low-pass filter. All planetary paths enclose the asymmetric ACR whose coordinates are $e_1 = 0.13$, $e_2 = 0.26$, $\sigma_2 = 284°$ and $\Delta\varpi = 108°$. The orbits close to the ACR are presented by black curves: they show oscillations of both the resonant angle σ_2 and the secular angle $\Delta\varpi$. For initial conditions far away from the stable center, the amplitudes of the oscillations increase and the angle $\Delta\varpi$ starts to circulate, as shown by blue curve in Figure 7, left panel. In this case, the curve separating the oscillatory and circulatory solutions represents a solution passing through the origin $e_1 = 0$; we say that the secular angle $\Delta\varpi$ (and σ_1) is in the oscillatory/circulatory regime of motion.

At variance, the red curve in Figure 7, right panel, separating the oscillations of the resonant angle σ_2 around the asymmetric ACR from the oscillations around zero, is formed by solutions asymptotic to the unstable symmetric ACR. Hence this curve is a true separatrix and the oscillations of σ_2 around the stable asymmetric ACR are true librations.

Outside the separatrix, the resulting orbits (as an example, the external black curve in Figure 7) are *horseshoe-like orbits* encompassing both asymmetric centers and appearing to oscillate around zero (analogous of orbits known from the dynamics of the asteroidal 1:1 mean-motion resonance with Jupiter). On the dynamical map on the right panel in Figure 6, the domain of the horseshoe orbits is separated from the central resonance region by the blue curve; this curve is the locus of the initial conditions corresponding to the true separatrix corresponding to the red color path in Figure 7. The detailed analysis of the dynamics in its close vicinity reveals that the transition across the blue curve is

topologically discontinuous, but the associated large-scale instabilities followed by disruption of the system are not observed on the dynamical map.

7.2. Dynamical maps around the two-branch asymmetric ACR

Two-branch asymmetric solutions are characteristic for mass ratio values lesser than 0.36. In this case, there exist two distinct stationary configurations described by different sets of orbital elements, but leading to the same value of total angular momentum. To better illustrate the bifurcation phenomenon, we analyze and compare planetary paths in the vicinity of the two-branch ACR: one belongs to the branch **A** and other to the branch **B**. This is done in Figure 8, where, in order to simplify the visualization, we concentrate solely on the secular behavior of the system. The paths of both planets were smoothed using low-pass filtering (resonant oscillations were eliminated) and were plotted in polar coordinates $(e_1, \Delta\varpi)$.
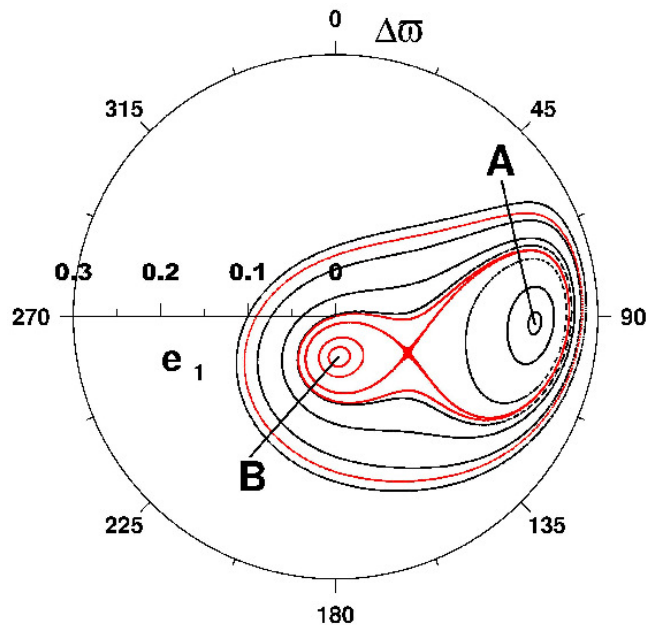

Figure 8.   Smoothed planetary paths around two-branch asymmetric ACR in the $(e_1, \Delta\varpi)$ polar coordinates calculated with $e_1 = 0.09$ and $m_2/m_1 = 0.3$ and the same energy. The smoothing has been done using a low-pass filter.

Two centers are now clearly observed in Figure 8, instead of one center as in the case of the one-branch asymmetric ACR (see Figure 7). The black curves surround the center associated to the branch **A** of the periodic solutions, while the red curves surround the center on the branch **B**. Among these paths, we find an infinite-period separatrix formed by solutions asymptotic to an unstable ACR (saddle point). Outside the separatrix, all paths follow similar patterns and are structurally stable. The existence of the separatrix, not present in the case of a single ACR, implies that there are motions near one of the two-branch asymmetric ACR that actually correspond to a true libration around the ACR and no longer a simple circulation/oscillation. This asymmetric ACR belongs to

the branch **B**. In its neighborhood therefore both $\Delta\varpi$ and σ_2 show true librations around the equilibrium values. In other words, the two-branch asymmetric ACR generates, inside the $2/1$ mean-motion resonance, a *true resonance* of the secular angle $\Delta\varpi = \sigma_1 - \sigma_2$. This novel behavior is absent in the cases of symmetric and one-branch asymmetric ACR (except for very-high, close to 1, eccentricities of the inner planet).
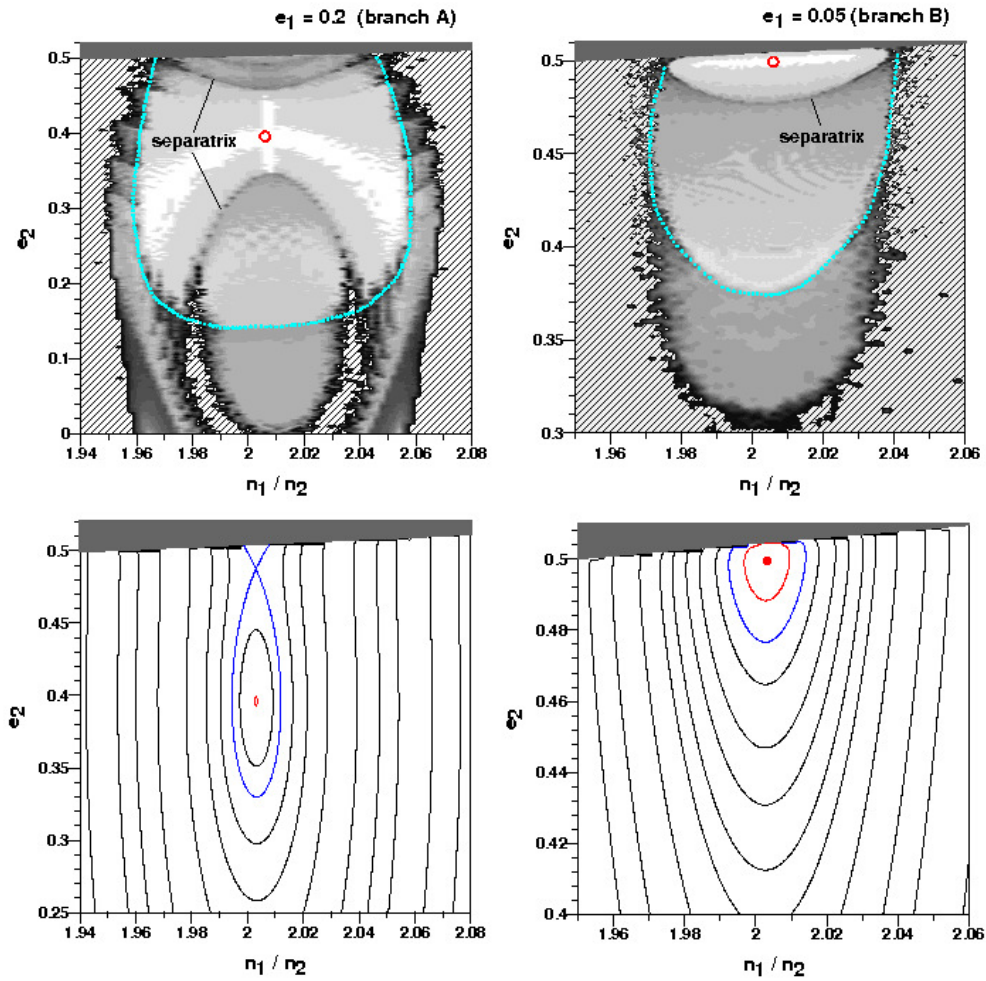

Figure 9.   Top: Dynamical maps of the two possible asymmetric solutions: one with $\sigma_1 = 32.2°$ and $\sigma_2 = -62.4°$ belongs to the branch **A** (left panel), and other, with $\sigma_1 = 120.7°$ and $\sigma_2 = -69.1°$, to the branch **B** (right panel). The positions of the stable asymmetric centers on the corresponding dynamical maps are marked by red dots. Bottom: The levels of the resonant Hamiltonian around the branch **A** ACR (left panel) and the branch **B** ACR (right panel).

Using the same constants of motion, but two distinct sets of critical angles, we constructed the dynamical maps on the two asymmetric representative planes, shown in Figure 9 (top panels). The positions of the stable asymmetric centres on the corresponding dynamical maps are marked by red dots. The

dynamical map of the branch **A** solution (left-top panel) displays qualitative similarities with the one-branch ACR map shown in Figure 6, right-top panel, mainly the existence of two families of periodic motion intersecting at the center and a large zone of the quasi-periodic motion inside the exterior resonance regime. The main difference is the existence of the separatrix whose location is indicated on the left-top graph in Figure 9.

In order to better visualize a separatrix feature, we present in the same figure (bottom panels) the neighborhoods of the stable asymmetric ACR, plotting the energy levels of the resonant Hamiltonian. A bifurcation phenomenon is clearly observed in the behavior of the Hamiltonian around the branch **A** ACR on the left-bottom panel, where the energy level in blue color presents a saddle-type structure. The continuation of the same level in the domain of the branch **B** ACR appears on the right-bottom graph as a blue curve. The fixed point of the centre (red dot) on the left-bottom panel appears as the energy level in red color on the right-bottom graph; that is, the branch **B** stationary solution is located at the global maximum of the resonant Hamiltonian.

The region surrounding the branch **B** ACR (right-top panel) is a domain of true secular resonance, when the passage from oscillation to circulation of the angle $\Delta\varpi$ is chaotic. In this region, both the resonant angle σ_2 and the secular angle $\Delta\varpi$ librate around the corresponding asymmetric values (in the case shown in Figure 9, they are 120.7° and 189.1°, respectively). The equilibrium configurations of the planetary orbits are similar to those shown in Figure 4: the closest approaches occur when the central star and the two planets form a triangle configuration. Our numerical simulations have shown that the planetary motions in this region are very stable, even for very high eccentricities of the orbits.

Finally, on both maps in Figure 9 there exist regions of the horseshoe-like orbits at initial conditions situated outside the cyan curves. The horseshoe orbits are not centered on any single ACR, but actually contain all of them including the single (or double) stable asymmetric ACR and the unstable symmetric ACR.

References

Beaugé, C. & Michtchenko, T.A. 2003, MNRAS, 341, 760

Beaugé, C., Michtchenko T.A. & Ferraz-Mello S. 2006, MNRAS, 365, 1160

Callegari, N. Jr, Michtchenko, T.A. & Ferraz-Mello, S. 2004, Celest. Mech. Dyn. Astron., 89, 201

Callegari, N. Jr, Ferraz-Mello, S. & Michtchenko T. A. 2006, Celest. Mech. Dyn. Astron., 94, 381

Ferraz-Mello, S., Beaugé, C. & Michtchenko, T. A. 2003, Celest. Mech. Dyn. Astr., 87, 99

Ferraz-Mello, S., Michtchenko, T.A. & Beaugé C. 2005, AJ, 621, 473

Ferraz-Mello, S., Michtchenko, T.A. & Beaugé, C. 2006, In: Chaotic Worlds: From Order to Disorder in Gravitational N-Body Systems (Steves B.A. et al., eds.), NATO Science Series II, Springer, 227, 255 (astro-ph 0402335)

Giuppone, C.A., Beaugé, C., Michtchenko, T. A. & Ferraz-Mello, S. 2010, MNRAS, 407, 390

Hadjidemetriou, J. D. 2002, Celest. Mech. Dyn. Astron., 83, 141

Kley, W. 2000, MNRAS, 313, L47

Kley, W., Lee M. H., Murray N., Peale S. J. 2005, A&A, 437, 727

Lee, M.H. 2004, ApJ, 611, 517

Lee, M. H. & Peale, S. J. 2002, ApJ, 567, 596

Michtchenko, T. A. & Ferraz-Mello S. 2001, Icarus, 149, 357

Michtchenko, T.A., Ferraz-Mello, S. & Beaugé C., 2006a, Icarus, 181, 555

Michtchenko, T.A., Beaugé, C. & Ferraz-Mello, S. 2006b, Celest. Mech. Dyn. Astr., 94, 411

Michtchenko, T.A., Beaugé, C. & Ferraz-Mello, S. 2008a, MNRAS, 387, 747

Michtchenko, T.A., Beaugé, C. & Ferraz-Mello, S. 2008b,MNRAS, 391, 215

Snellgrove, M.D., Papaloizou, J.C.B. & Nelson R. P. 2001, A&A, 374, 1092

Voyatzis, G. & Hadjidemetriou, J.D. 2006, Celest. Mech. Dyn. Astr., 95, 259

LECTURE

Chaos in elliptical galaxies

J.C. Muzzio[1,2]

(1) Facultad de Ciencias Astronómicas y Geofísicas - UNLP
(2) Instituto de Astrofísica de La Plata (CCT CONICET La Plata - UNLP)

Abstract. Here I present a review of the work done on the presence and effects of chaos in elliptical galaxies plus some recent results we obtained on this subject. The fact that important fractions of the orbits that arise in potentials adequate to represent elliptical galaxies are chaotic is nowadays undeniable. Alternatively, it has been difficult to build selfconsistent models of elliptical galaxies that include significant fractions of chaotic orbits and, at the same time, are stable. That is specially true for cuspy models of elliptical galaxies which seem to best represent real galaxies. I argue here that there is no physical impediment to build such models and that the difficulty lies in the method of Schwarzschild, widely used to obtain such models. Actually, I show that there is no problem in obtaining selfconsistent models of elliptical galaxies, even cuspy ones, that contain very high fractions of chaotic orbits and are, nevertheless, highly stable over time intervals of the order of a Hubble time.

1. Models of elliptical galaxies

How can we build a dynamical model of an elliptical galaxy? The basic constituents of galaxies are dark matter, stars and gas (plus a little dust). Gas dynamics plays a significant role in the formation process of all galaxies, but their influence in full-fledged elliptical galaxies is negligible, so that only dark matter and stars need to be taken into account for models that represent them. Thus, Newtonian dynamics is all that is needed to model elliptical galaxies and even a black hole can be represented simply as a point mass for our purposes.

The effects of star-star interactions can be neglected for these models, as is shown in any textbook on galactic dynamics (see, e.g., Binney & Tremaine 2008), and we can assume that there is a smooth distribution of mass density (dark matter plus stars) that creates a similarly smooth potential. Let us consider now the types of orbits we found in elliptical galaxies.

1.1. Typical orbits in elliptical galaxies

A significant breakthrough in elliptical galaxies dynamics was done with the use of Stäckel potentials that are fully integrable, so that all orbits are regular. In particular, de Zeeuw & Lynden-Bell (1985), de Zeeuw (1985) and Statler (1987) investigated the "perfect ellipsoid", whose density is given by:

$$\varrho(r) = \frac{\varrho_o}{(1 + m^2)^2} \tag{1}$$

where m is the ellipsoidal radius:

$$m^2 = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}. \tag{2}$$

Although that model is no longer regarded as an adequate representation of elliptical galaxies, its main kinds of orbits are also found in the more realistic models we use nowadays. They are: a) Boxes, that can come as close as one wishes to the center of the model and do not keep the sense of rotation; b) Minor axis tubes, that never come close to the center and rotate always in the same sense around the minor axis of the model; c) Major axis tubes, that never come close to the center either and always rotate in the same sense around the major axis of the model, they are in turn divided in inner and outer major axis tubes.

There is no reason, other than mathematical simplicity, to adopt the perfect ellipsoid, or any other Stäckel potential for that matter, to model elliptical galaxies. On the contrary, it has a flat central density distribution and we know nowadays that cuspy galaxies are the rule (we will return to this point later on). In cuspy models the box orbits tend to be replaced by "boxlets", i.e., resonant orbits that avoid the center of the model (Miralda-Escudé & Schwarzschild 1989).

All these are regular orbits and, of course, in generic potentials we have also chaotic orbits but, since chaos is the main subject of the present lectures, they will be analyzed in much more detail later on.

1.2. The problem of self-consistency

The galactic models must be self-consistent, that is, the mass distribution creates a gravitational field where certain kinds of orbits are possible, and those orbits should be such that they yield that mass distribution. The Poisson and Boltzmann equations can be used to obtain simple spherical or disk models (see, e.g., Binney and Tremaine 2008), but special methods are required for the triaxial models needed for elliptical galaxies.

One of the most popular of those methods is the one due to Schwarzschild (1979). First, one chooses a reasonable mass distribution and obtains its potential. A library of orbits is then computed in that potential using a variety of initial conditions in order to cover all the possible orbits in that potential, and weights are assigned according to the time that a particle on each orbit would spent in every region of the available space. Those weights are subsequently used to set a system of linear equations that relate the density in every region to the fraction of orbits that make up the system. Finally, solving that system, the fraction of every kind of orbit is obtained.

This method certainly offers a straightforward way of obtaining a self-consistent model, but there are several details that should be taken into account. To begin with, there is not a single solution, and one does not know *a priori*

towards which solution the method will converge. Besides, the chosen initial mass distribution conditions the results, for example, choosing the mass distribution of the perfect ellipsoid given above will force the system to have constant axial ratios, b/a and c/a, from the center through the outermost regions of the model. It is not easy to guarantee that all the possible kinds of orbits have been included in the library and, to make this problem worse, the libraries are usually not very large, from several hundreds of orbits at Schwarschild's time and up to a few tens of thousands nowadays (see, e.g., Capuzzo-Dolcetta *et al.* 2007).

The most serious problem to the method of Schwarzschild is the one posed by chaotic orbits, which can display very different behavior at different time intervals, even over long periods of time. Sticky orbits, in particular, can mimic regular orbits over long intervals, only to reveal their true chaotic nature later on. Schwarzschild (1993) realized that chaotic orbits were all too frequent in triaxial models and he tried to include them in his models but, then, the models did not remain stationary. He computed models with orbits integrated over one Hubble time and, then, new models using the same orbits but integrated between two and three Hubble times. Differences beween equivalent models were revealed, e.g., by differing axial ratios, as shown by Table 1.

Table 1. Axial ratio changes in the models of Schwarzschild (1993).

c/a	b/a	$\Delta(c/a)$	$\Delta(b/a)$
0.7	0.8631	-4.3%	-7.3%
0.5	0.7906	2.6%	-3.9%
0.3	0.7382	10.0%	0.9%
0.3	0.9534	4.3%	0.6%
0.3	0.4254	16.7%	16.4%
0.3	0.3289	6.7%	3.4%

Of course, it would be obviously hopeless to try to use observations to detect such small changes, so that we might ask: If we cannot recognize a truly stable galaxy from a slowly evolving one, why care? Nevertheless, from a theoretical point of view, it is certainly important. Therefore, the question we should be actually asking is: Is it possible to have a stable self-consistent system if a significant fraction of its material is in chaotic motion? No doubt, most dynamicists will feel inclined to answer "No" to that question (see, e.g., Siopis and Kandrup 2000) but, as I will show, they will be wrong.

From the observational point of view there is proof, both statistical (e.g., Ryden 1996) and on individual galaxies (e.g., Statler *et al.* 2004), that at least some elliptical galaxies are triaxial, and not merely rotationally symmetric (either oblate or prolate). Moreover, the surface brightness of elliptical galaxies increases towards the center forming a "cusp" (e.g., Crane et al. 1993, Moller et al. 1995) that reveals the presence of central mass concentration and, probably, a black hole. The problem is that it has long been known (see, e.g., Gerhard & Binney 1985) that central cusps and black holes can disrupt the box orbits that are necessary to keep the triaxial form. The strong central field reorients the box and chaotic orbits that come close to the center, so that a triaxial model

will quickly evolve into a rotationally symmetric one (we will return to this point later on). Is it then possible to have cuspy triaxial galaxies? Can such galaxies have significant amounts of matter on chaotic orbits? These questions are obviously important from an observational point of view and for fitting models to the observations.

1.3.   The problem of diffusion

Several authors haven been worried by the "problem" of chaotic mixing and, for that reason, they have tried to avoid including chaotic orbits in their models. For example, Merritt and Fridman (1996) refer to solutions that contain chaotic orbits as *quasi equilibrium* and indicate that in quasi equilibrium models chaotic mixing will produce a slow evolution of the model figure, specially near the center. That idea comes from the assumption that, in the long run, chaotic orbits will cover uniformly all the space available to them by the energy integral, i.e., a region much rounder than the triaxial model, but that assumption is simply wrong. Chaotic orbits do not occupy all the space available to them and live there forever happily. First, they do not occupy all the space allowed to them by the energy integral because there are usually "islands of stability" where chaotic orbits cannot enter and, second, they may at times behave similarly to regular orbits. The more or less chaotic behavior of a chaotic orbit can be easily followed, for example, computing the finite time Lyapunov numbers (FTLNs hereafter) that increase, or decrease, as the orbit behaves more, or less, chaotically. As shown by Muzzio *et al.* (2005), if one sorts the fully chaotic orbits according to the values of those numbers, one finds that the orbits with lower values give a more triaxial distribution than the orbits with higher values.

   The actual problem posed by the changing behavior of chaotic orbits is the one they pose to the method of Schwarzschild. The method will favor the inclusion of orbits elongated in the direction of the major axis but, when those orbits begin to behave more chaotically, the shape of the model will become rounder. This effect could be compensated if the model includes chaotic orbits that initially had a rounder distribution and, later on, a more elongated one. In other words, one might have a *dynamic equilibrium*: a given chaotic orbit will not occupy always the same region of space (as regular orbits do) but, as it moves out to occupy a different zone, another chaotic orbit will fill in the region left vacant by the former. I will show later on that it is perfectly possible to achieve such dynamical equilibrium and to get stable models that include large fractions of chaotic orbits, but one has to resort to methods different from Schwarzschild's. One cannot see any simple way to take into account the abovementioned effect in that method, and the situation is even worse when constant axial ratios are used throughout the model. As indicated by Muzzio *et al.* (2005), in that case one has no rounder halo that could act as a reservoir of chaotic orbits that, as time goes by, become more elongated and compensate for those that, on the contrary, become rounder. In fact, stable triaxial models that include high fractions of chaotic orbits become rounder as one goes farther from the center of the system (see, e.g., Muzzio *et al.* 2005, Aquilano *et al.* 2007).

1.4. Central black holes

The effects of a central growing black hole on the dynamics of a triaxial system were well described by Merritt and Quinlan (1998), who built an N-body model of a triaxial galaxy and verified that it was stable. They then let a black hole grow at the center of the model and, as could be expected, the central density distribution became much steeper. But it also turned out that the model lost its triaxiality and the evolution was faster for more massive black holes. The central regions became rounder, but even the outermost regions were affected.

Rather different results were obtained by Poon & Merrit (2002, 2004) who obtained equilibrium models of the central regions of triaxial galaxies containing black holes and checked their stability. They were able to find stable configurations that persisted even within the sphere of influence of the black hole. Interestingly, their models included chaotic orbits as well as regular ones. These results seem to be at odds with other work, like that of Merritt & Quinlan (1998). It is possible that the fact that the models were built including the black hole, rather than letting it grow after the model was built, might help to explain the difference.

2. Chaotic orbits in elliptical galaxies

The presence of chaotic orbits in triaxial potentials was noticed early on, even in the original paper of Schwarzschild (1979). Moreover, in Schwarzschild (1993), which dealt with the singular (i.e., cuspy) logarithmic potential, he showed that many orbits that resulted from his "initial condition spaces" were indeed chaotic.

More recently, after the acceptance that elliptical galaxies are cuspy and that most of them may be harboring central black holes, it was found that both cuspiness and central black holes enhance the chaotic effects. These studies of orbits are usually performed on fixed and smooth potentials, such as the one arising from the triaxial generalization of the model of Dehnen (1993), whose density distribution is:

$$\varrho(m) = \frac{(3-\gamma)aM}{4\pi} m^{-\gamma}(m+d)^{-(4-\gamma)} \quad \text{with} \quad 0 \leq \gamma < 3, \qquad (3)$$

where m is the triaxial radius already defined by equation (2), M is the total mass, d is a scale parameter proportional to the effective radius (i.e., the radius containing half of the mass) and γ parametrizes the slope of the central cusp. Merritt and Fridman (1996) provided the corresponding equations for the potential, the forces and the derivatives of the forces (which are needed for the variational equations that allow the computation of the Lyapunov exponents). Those expresions involve rather complicated integrals and must be solved numerically.

The axial ratios b/a and c/a give an idea of the flatness of the model, and it is usual to measure the triaxiality, T, as:

$$T = \frac{(a^2 - b^2)}{(a^2 - c^2)}. \qquad (4)$$

It goes from 0, for an oblate spheroid, to 1, for a prolate spheroid, and the ellipsoids with $T = 0.5$ are called "maximally triaxial" ellipsoids.

Merritt and Fridman (1996) investigated maximally triaxial models with $\gamma = 1$ (weak-cusp) and $\gamma = 2$ (strong-cusp). Their libraries of orbits contained high fractions of chaotic orbits, particularly in the strong-cusp case and for orbits with zero initial velocity. Thus, it is not surprising that the attempts to build stable models without chaotic orbits were doomed to failure. Using the method of Schwarzschild, Merritt and Fridman managed to build "fully mixed" models for their weak-cusp case, but not for the strong-cusp one.

2.1.   The contribution of Kandrup and Siopis

Siopis and Kandrup (2000) and Kandrup and Siopis (2003) did two very comprehensive investigations of orbits in the triaxial Dehnen potential, allowing also for the presence of a central black hole. They considered different axial ratios, cusp slopes and black hole masses and, in addition to the chaoticity of the orbits, they also investigated chaotic diffusion and how it is affected by noise.

They found that chaotic orbits tend to be extremely sticky for all cusp slopes, but specially for the steepest ones. This fact was revealed by visual inspection of the orbits, but also by the bimodal distribution of the FTLNs. Besides, if the orbits were not sticky, the FTLNs should fall as $t^{-1/2}$, where t is the integration time, but they found a much lower slope, corroborating the stickiness. Moreover, the presence of a black hole increased the stickiness of the orbits. Except for the steepest cusp, the *fractions* of chaotic orbits were not much affected by either the steepness of the cusp or the mass of the black hole. Nevertheless, the *values* of the FTLNs were significantly affected by both: steeper cusps and more massive black holes led to larger FTLN values. They investigated the effect of triaxiality adopting $a = 1.0$, $c = 0.5$ and selecting different b values. As could be expected, the fractions of chaotic orbits decreased both toward $b = 0.5$ (prolate spheroid) and toward $b = 1.0$ (oblate spheroid). Besides, going toward the innermost regions of the model the fraction of chaotic orbits increased, and also increased the effect of the mass of the black hole on that fraction. They also found that prolate spheroids tend to have larger fractions of chaotic orbits, and with larger FTLNs, than oblate spheroids; that effect was more pronounced for the outermost shells. Finally they considered changing both b and c adopting $b = 1 - D$ and $c = 1 - 2D$, i.e., from a spherical system for $D = 0$, to a disk for $D = 0.5$. The fraction of chaotic orbits turned out to increase monotonically with D, and even a very small departure from sphericity yielded chaos. Without a black hole, the size of the FTLNs also increased monotonically with D. Larger black hole masses resulted in larger FTLN values, but the increase with D leveled off for the most massive ones.

Chaotic mixing is another interesting aspect of the work of Kandrup and Siopis. They showed that different ensembles of chaotic orbits diffuse in different ways and end up occupying different regions even after long times. A very interesting aspect of their work is that they recognized that galaxies are subject to several perturbations that can be modelled as "noise" and profoundly affect chaotic diffusion. They considered: a) Periodic driving, to simulate the effect of a satellite orbiting the galaxy; b) Friction and white noise to simulate discreteness

effects; c) Coloured noise to simulate the effects of encounters with other galaxies. They showed that all these effects tend to increase the diffusion rate.

2.2. Partially and fully chaotic orbits

Since we are interested in equilibrium models of elliptical galaxies, the energy integral always holds. Regular orbits have two additional isolating integrals, but we can have chaotic orbits either without any other isolating integral (fully chaotic orbits), or with just one (partially chaotic orbits). This difference had been recognized long ago (Contopoulos *et al.* 1978, Pettini & Vulpiani 1984), but little importance was assigned to it in galactic dynamics studies despite being found in several investigations of triaxial systems (e.g., Goodman and Schwarzschild 1981, Merritt and Valluri 1996). These studies used a sort of three dimensional (3-D) Poincaré map. If the velocity on an orbit is computed every time that the particle returns to the same point, then: a) The velocities of fully chaotic orbits will adopt any possible direction; b) Those of partially chaotic orbits will fall on a curve; c) The velocities of regular orbits will have always the same, or at most a finite number, of directions. Needless to say, this sort of 3-D Poincaré map is extremely computer time consuming, because one has to follow the orbit over very long time intervals to have a significant number of returns (almost) to the same point. A more efficient, albeit also computer time consuming, method is to compute all (and not just the largest) Lyapunov exponents: the exponents of regular orbits are all zero, partially chaotic orbits have one positive exponent, and fully chaotic orbits have two positive exponents. Using Lyapunov exponents Muzzio (2003) showed that, in triaxial models of elliptical galaxies, the spatial distributions of partially and fully chaotic orbits are diferent, so that it is important to separate them in studies of galactic dynamics. Similar results were obtained by Muzzio and Mosquera (2004) with models of galactic satellites and by Muzzio *et al.* (2005), Aquilano *et al.* (2007), Muzzio *et al.* (2009) and Zorzi (2011) with models of elliptical galaxies.

The Lyapunov exponents should be obtained integrating the orbit over an infinite time interval but, as that is impossible for numerically integrated orbits, the FTLNs are used instead. That rises the problem that one cannot reach zero values, even for regular orbits: if T is the integration interval, then the lowest FTLN values will be of the order of $\ln T/T$. Besides, one might ask which is the practical limiting value to separate regular from chaotic orbits. The inverse of the Lyapunov exponent is called the Lyapunov time, and it gives the time scale for the exponential divergence of the orbit. At first sight, orbits with Lyapunov times longer than the Hubble time could be regarded as regular, but one should recall that what actually interests in galactic dynamics is the orbital distribution, rather than the exponential divergence of orbits. Therefore, the right question to ask is which is the limiting value of the Lyapunov exponents that separates orbits whose distribution is similar to that of regular orbits from those that have a clearly different distribution. Aquilano *et al.* (2007), Muzzio *et al.* (2009) and Zorzi (2011) investigated this problem and found that, for their models, the limiting value corresponds to a Lyapunov time about six or seven times larger than the Hubble time. In other words, orbits not chaotic enough to experience significant exponential divergence on a Hubble time are, nevertheless, chaotic enough to have a distribution significantly different from

that of regular orbits. It should be noted, however, that the computed fraction of chaotic orbits is not substantially altered by the limiting value selected: for the models of the authors cited, adopting the limit corresponding to six or seven Hubble times may result in a fraction of chaotic orbits of, say, 40% as compared to 35% for the limit corresponding to one Hubble time.

3.   The N-body method

The method of Schwarzschild (1979) is not the only one available to obtain equilibrium models of elliptical galaxies. Actually, self-consistent models of triaxial systems had been obtained earlier using the N-body method (see, e.g., Aarseth & Binney 1978). One starts with a certain distribution of mass points and follows its evolution using an N-body code to integrate the equations of motion. A wise selection of the initial distribution, or a little tinkering with the code, allows one to obtain a stable triaxial system whose self-consistency is assured by the use of the N-body code.

One can begin, for example, with a spherical distribution of mass points with very small velocities. Gravity will force the collapse of such a system and the radial orbit instability will lead it toward a triaxial equilibrium distribution (see, e.g., Aguilar & Merritt 1990). One can change the initial velocity dispersion to obtain systems with different triaxiality: the smaller the dispersion, the larger the triaxiality. This method has been used by Voglis *et al.* (2002), Kalapotharakos and Voglis (2005), and by my coworkers and myself. Another possibility is to launch one stellar system against another and let them merge, yielding a triaxial system, a strategy followed, e.g., by Jesseit *et al.* (2005).

Code tinkering was used, for example, by Holley-Bockelmann *et al.* (2001) who started with a Dehnen (1993) spherical distribution of mass points and added to the N-body code a fictitious force that slowly squeezed the system, first in the z direction, and then in the y direction. The original N-body code was finally used to let the system relax toward a final triaxial equilibrium distribution.

After building the triaxial system, one has to investigate its orbital structure. The N-body code is no longer useful because its potential is neither smooth (because it is the sum of the potentials of the individual particles) nor time independent (due to the statistical changes in the distribution of the finite number of particles). Therefore, a smooth and time independent approximation should be adopted for the potential and then, orbits in that potential can be computed using as initial conditions the positions and velocities of all, or a random sample of, the mass points. One can then separate regular from chaotic orbits, and partially from fully chaotic orbits, with an adequate chaos indicator and, finally, the different kinds of regular orbits can be obtained with an automatic classification code based, e.g., on the analysis of the orbital frequencies.

Both Schwarzschild's and the N-body method end up with the same result: an equilibrium self-consistent triaxial system and the knowledge of its orbital content. The steps to reach those results are, however, different. One uses the orbits to obtain the system in Schwarzschild's method, while in the N-body method the system is built by the gravitational interactions mimicked by the N-body code and thereafter one performs the orbital analysis.

3.1. The work of the Athens' group

Voglis *et al.* (2002) obtained two non-cuspy stable triaxial models, Q (Quiet) and C (Clumpy), from cold collapses of initially spherical particle distributions. The fractions of chaotic orbits were 32% for the former and 26% for the latter. A similar model was built and investigated by Kalapotharakos and Voglis (2005), who found 32% of chaotic orbits. An interesting point that they raised is that the real particles in their model did not cover all the space that one could cover with test particles. It is a natural consequence of self-consistency, that only allows the presence of certain orbits, and Kalapotharakos and Voglis showed this effect beautifully on the frequency map. A very clever aspect of their work is that they determined the orbital frequencies of the chaotic orbits, in addition to those of the regular orbits, and showed that, due to the stickiness, many of them fall close to the loci of the regular orbits on the frequency map. Again, their model is stable but, after the sudden introduction in it of a central black hole, it quickly evolves toward an oblate shape.

More recently, Kalapotharakos (2008) investigated the evolution of two stable, non-cuspy, triaxial systems after the introduction of a central black hole. He linked that evolution to the presence of chaotic orbits and introduced a parameter, dubbed effective chaotic momentum, that correlates well with the rate of evolution of the system.

3.2. The puzzles of the work of Holley-Bockelmann et al.

As already mentioned, Holley-Bockelmann *et al.* (2001) obtained cuspy triaxial systems through adiabatical deformation of Dehnen (1993) spherical models. Their systems turned out to be very stable, but they found very little chaos (less than 1%, and they even attributed that little to noise in their potential approximation), in complete contradiction with the results of the Athen's group and our own. Kandrup & Siopis (2003) pointed out that the Fourier technique used by Holley-Bockelmann *et al.* could have made them miss many chaotic orbits, a likely possibility considering that such technique is not as good as others for chaos detection (see, e.g., Fig. 2 of Kalapotharakos & Voglis 2005). Nevertheless, one should recall that the collapses used by the Athens' group and ourselves to build triaxial models yield predominantly radial orbits, which have long been known to favor chaos (see, e.g., Martinet 1974), while the adiabatical squeezing used by Holley-Bockelman *et al.* probably resulted in a more isotropic velocity distribution, so that the discrepancy among the chaotic fractions might be real.

In a second paper, Holley-Bockelmann *et al.* (2002) let a black hole grew at the center of one of their models. As could be expected, they found that the central cusp became steeper and the central regions rounder. But the outer regions (even those that contained the inner 10% of the total mass) retained the triaxial shape. Just as mentioned before about the work of Poon and Merritt (2002, 2004), these results seem to be at odds with other investigations, like that of Merritt & Quinlan (1998).

3.3. The work of our La Plata-Rosario group

In our first works we used 100,000 particles to build our *N*-body models, and for the past few years we have been using one million, guaranteeing in any case the

accuracy of stability studies. Besides, we typically classify between 3,000 and 5,000 orbits per model, which provides adequate statistics. To build our models we just use the simple cold collapse technique and we do not try to follow the actual galaxy formation process, because we are not interested in that process but only in obtaining models morphologically similar to elliptical galaxies. For example, although it is not physically realistic, we let our models relax during intervals of several Hubble times to ensure that our subsequent studies refer to equilibrium models.

Muzzio *et al.* (2005) obtained a model of a non-cuspy E6 galaxy that contained 52.7% of chaotic orbits. Later on, Muzzio (2006) found that the model displayed very slow figure rotation (i.e., it rotated despite having zero angular momentum). Taking the rotation into account, the fraction of chaotic orbits raised to 56.6%, probably because of the symmetry breaking caused by the rotation, despite its small value. Later on, Aquilano *et al.* (2007) obtained three models of non-cuspy galaxies resembling elliptical galaxies of types E4 through E6, all of them very stable and with between one third and two thirds of chaotic orbits. All these models had axial ratios that increased from the center toward the outer parts of the galaxy, which may have contributed to them being stable and chaotic at the same time. Besides, all of them corroborated the need to distinguish partially from fully chaotic orbits because their spatial distributions were different.

Muzzio *et al.* (2009) managed to obtain cuspy models of E4 and E6 galaxies that, again, were highly stable despite having about two thirds of chaotic orbits. Nevertheless, those models had pushed to the limit the use of the *N*-body code of Aguilar (see, White 1983 and Aguilar & Merritt 1990) and, in fact, they had to introduce a small additional potential to compensate for the softening needed by the code. Therefore, in order to continue investigating cuspy models, we switched to the code of Hernquist (Hernquist & Ostriker 1992) that needs no softening and uses a radial expansion of the potential in functions related to the potential of Hernquist (1990) which is itself cuspy with $\gamma = 1$. Our first work with this code is the PhD thesis of Zorzi (2011), defended last June at the Universidad Nacional de Rosario.

Kalapotharakos *et al.* (2008) investigated the approximation of *N*-body realizations of models of Dehnen (1993) for different values of γ using the self-consistent field method (the method of Hernquist & Ostriker (1992) is just a special case of that method), and they found that the choice of the radial basis functions seriously affected the results obtained on chaotic orbits. Nevertheless, the method of Hernquist & Ostriker turned out to be adequate for models with $\gamma \simeq 1$, and we had selected the number of terms in the potential expansion so as to have models whose cusps had that slope (Zorzi & Muzzio 2009).

Zorzi (2011) built four groups of models that mimic E2, E3, E4 and E5 galaxies. Each group consisted of three models that differed only in the seed number used to randomly generate the initial conditions. In other words, for each galaxy type she had three statistically equivalent models, which is important to control the consistency and robustness of the results to be obtained from them. In fact, except for the figure velocity rotation, all her results were essentially the same for statistically equivalent models. Figure 1 gives the logarithmic density vs. radius plot for the central regions of her E4 models and we can notice the

excellent agreement among the different realizations of the model, as well as the $\gamma \simeq 1$ cusp.
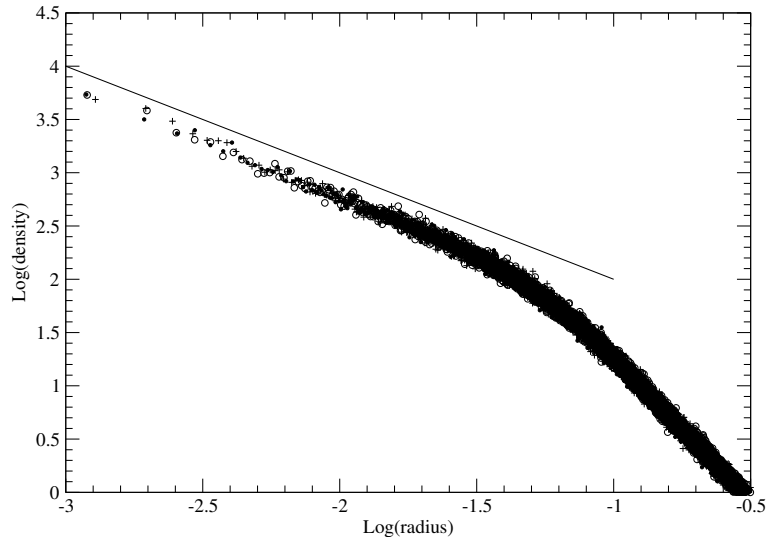

Figure 1.   The logarithm of the density vs. the logarithm of the radius for the E4 models of Zorzi (2011). Filled and open circles and crosses correspond to three statistically equivalent realizations of the same model. The slope of the straight line is equal to 1.

To check the stability of the models, she computed their central density and their three moments of inertia over intervals of the order of five Hubble times. The changes in those quantities were very small, indeed, smaller than 2.7% in one Hubble time in all cases. Moreover, she showed that most of that change is most likely due to relaxation effects in the N-body code (see Hernquist & Barnes 1990). Not only are these values smaller than most of those of Schwarzschild shown in Table 1, but Zorzi's had been obtained with self-consistent models, while Schwarzschild's correspond to a fixed potential. Actually, when the potential is fixed, the changes in the models of Zorzi are an order of magnitude smaller than those obtained self-consistently. The fractions of chaotic orbits are extremely high in Zorzi's models, up to more than 85% for her E2 and E5 models, and larger that 75% for the other two. The values of the FTLNs are also very high, almost twice the values of the equivalent non-cuspy models of Aquilano *et al.* (2007). And, as in our previous work, she also found that partially and fully chaotic orbits have different distributions and should be analyzed separately.

4.   Conclusions

The works of the Athens' group and our own show that it is perfectly possible to build triaxial stellar systems, even cuspy ones, that include large fractions of chaotic orbits and are, nevertheless, highly stable over intervals of the order of one Hubble time. All these models were obtained using the N-body method, so that the difficulty to obtain such models with the method of Schwarzschild should be attributed to the method itself and not to physical causes.

We certainly need more models built with the adiabatic squeezing technique of Holley-Bockelmann *et al.* (2001). The orbits in those models are probably less radial than those in the models arising from cold collapses and that might result in lower fractions of chaotic orbits and models more capable of preserving triaxiality when harboring a black hole.

The distributions of partially and fully chaotic orbits are certainly different, but much remains to be done on this subject. Are the partially chaotic orbits merely confined to the stochastic layers around resonances? Or do they fill in connected regions where an isolating integral, or pseudo integral, holds? And, certainly, we need better and faster methods than that of the Lyapunov exponents to separate partially from fully chaotic orbits.

Modern observations suggest not only that elliptical galaxies have significant rotation, but that they even rotate in different directions at different distances from the center, perhaps a consequence of past mergers. Nevertheless, there are not many works on rotating ellipticals (a recent exception is that of Deibel et al. 2011) and more work is clearly warranted on this subject.

The figure rotation we found in many of our models is certainly puzzling. All the tests we made support that it is real and, besides, these models can be seen as the stellar counterpart of the Riemann ellipsoids of fluid dynamics. But why, appart from the tendency of flatter systems to rotate faster, the rotation shows no obvious connection to the properties of the model and, worst, why do statistically equivalent models display different rotation velocities? This is another subject where more research is clearly warranted.

Last but no least, let me end recalling that, although I have only briefly mentioned regular orbits here, their study is also very interesting. Not only are they important for the dynamics of elliptical galaxies, but they also offer technical challenges, like the design of fast and accurate methods of automatic classification.

References

Aarseth, S.J., & Binney, J. 1978, MNRAS, 185, 227

Aguilar, L.A., & Merritt, D. 1990, ApJ, 354, 33

Aquilano, R.O., Muzzio,J.C., Navone, H.D., & Zorzi, A.F. 2007, Celest. Mech. & Dynam. Astron., 99, 307

Binney, J., & Tremaine, S.J. 2008, Galactic Dynamics, Princeton University Press

Capuzzo-Dolcetta, R., Leccese, L., Merritt, D., & Vicari, A. 2007, ApJ, 666, 165

Contopoulos, G., Galgani, L., & Giorgilli, A. 1978, Phys. Rev. A, 18, 1183

Dehnen W. 1993, MNRAS, 265, 250

Deibel, A.T., Valluri, M., & Merritt, D. 2011, ApJ, 728, 128

de Zeeuw, P.T. 1985, MNRAS, 216, 273

de Zeeuw, P.T., & Lynden-Bell, D. 1985, MNRAS, 215, 713

Gerhard, O.E., & Binney, J. 1985, MNRAS, 216, 467

Goodman, J., & Schwarzschild, M. 1981, ApJ, 245, 1087

Hernquist, L. 1990, ApJ, 356, 359

Hernquist, L., & Barnes, J. 1990, ApJ, 349, 562

Hernquist, L., & Ostriker, J.P. 1992, ApJ, 386, 375

Holley-Bockelmann, K., Mihos, J.C., Sigurdsson, S., & Hernquist, L. 2001, ApJ, 549, 862

Holley-Bockelmann, K., Mihos, J.C., Sigurdsson, S., Hernquist, L., & Norman, C. 2002, ApJ, 567, 817

Jaffe, W., Ford, H.C., O'Connell, R.W., van den Bosch, F.C., & Ferrarese, L. 1994, AJ, 108, 1567

Jesseit, R., Naab, T., & Burkert, A. 2005, MNRAS, 360, 1185

Kalapotharakos, C. 2008, MNRAS, 389, 1709

Kalapotharakos, C., Efthymiopoulos, C., & Voglis, N. 2008, MNRAS, 383, 971

Kalapotharakos, C., & Voglis, N. 2005, Celest. Mech. & Dynam. Astron., 92, 157

Kandrup, H.E., & Siopis, C. 2003, MNRAS, 345, 727

Merritt, D., & Quinlan, G.D. 1998, ApJ, 498, 625

Merritt, D., & Valluri, M., 1996, ApJ, 471, 82

Miralda-Escudé, J., & Schwarzschild, M. 1989, ApJ, 339, 752

Moller, P., Stiavelli, M., & Zeilinger, W.W. 1995, MNRAS, 276, 979

Martinet, L. 1974, A&A, 32, 329

Muzzio, J.C. 2003, BAAA, 45, 69

Muzzio, J.C. 2006, Celest. Mech. & Dynam. Astron., 96, 85

Muzzio, J.C., Carpintero, D.D., & Wachlin, F.C. 2005, Celest. Mech. & Dynam. Astron., 91, 173

Muzzio, J.C., & Mosquera, M.E. 2004, Celest. Mech. & Dynam. Astron., 88, 379

Muzzio, J.C., Navone, H.D., & Zorzi, A.F. 2009, Celest. Mech. & Dynam.Astron., 105, 379

Pettini, M., & Vulpiani, A. 1984, Phys. Lett., 106A, 207

Poon, M.Y., & Merritt, D. 2002, ApJ, 568, 89

Poon, M.Y., & Merritt, D. 2004, ApJ, 606, 774

Ryden, B.S. 1996, ApJ, 461, 146

Schwarzschild, M. 1979, ApJ, 232, 236

Schwarzschild, M. 1993, ApJ, 409, 563

Siopis, C., & Kandrup, H.E. 2000, MNRAS, 319, 43

Statler, T.S. 1987, ApJ, 321, 113, 1987

Statler, T.S., Emsellem, E., Peletier, R.F., & Bacon, R. 2004, MNRAS, 353, 1

Voglis N., Kalapotharakos C., & Stavropoulos I. 2002, MNRAS, 337, 619

White, S.D.M. 1983, ApJ, 274, 53

Zorzi, A.F. 2011, PhD Thesis, Universidad Nacional de Rosario, Argentina

Zorzi, A. F., & Muzzio, J. C. 2009, Memorias del II Congreso de Matemática Aplicada, Computacional e Industrial, Rosario

276

LECTURE

Faraway matter as a possible substitute for dark matter

A. Carati[1], L. Galgani[1]

(1) Dipartimento di Matematica, Università degli Studi di Milano

Abstract.   A review is given of an attempt, made in two recent papers, to estimate the gravitational action of faraway matter on a test particle, in connection with the velocity dispersion in clusters of galaxies and with the rotation curves of spiral galaxies, respectively. Under the assumptions that faraway matter has a fractal distribution and that the gravitational action has a correlation length of the order of some kiloparsec, the gravitational action of faraway matter appears to be sufficient to explain the observations relative to such two phenomena, without invoking any local, dark matter contribution.

1.   Introduction

The thesis illustrated in this paper is that the gravitational action of faraway matter may be a substitute for the local action of invisible, transparent (usually called dark) matter, at least in the two cases in which dark matter was first introduced in order to save the phenomena, namely, the velocity dispersion in clusters of galaxies and the rotation curves of spiral galaxies. The main idea underlying such a thesis came to our mind quite occasionally, in connection with one of our main themes of research, foundational features of classical electrodynamics. Indeed, motivated by a critique (Carati & Galgani 2004) of the way in which Planck was dealing with microscopic models of a black body,[1] we were involved in the problem of a microscopic foundation of dispersion of light in crystals[2]. In this connection it occurred to us to understand how dispersion of light in matter is strictly related to the so–called Wheeler–Feynman identity, which was conjectured by such authors (Wheeler & Feynman 1945) and we were able to prove in a simple model (Carati & Galgani 2003, Marino Carati & Galgani 2007). Now, such an identity has an evident global character, inasmuch as it

---

[1]Planck was considering the elementary model in which one oscillator is acted upon by an external field, and was unable to deal with the full dynamical system of N oscillators with mutual retarded interactions. He assumed all single oscillators to be independent from each other. This fact plagues all of his work, from which one would deduce an emission of radiation proportional to volume rather than to surface.

[2]A full microscopic treatment of the problem is lacking also in the celebrated book of Born and Huang, because radiation reaction force is neglected. Such a force was present in the previous studies of Planck, who actually was the scientist that discovered it. However, he did not recognize the existence of a cancellation which plays a fundamental role and was first proved by Oseen in the year 1916, and later conjectured on general grounds by Wheeler and Feynman.

takes into consideration the retarded and the advanced potentials "created" by all the charges present in the Universe, and states that the sum of the semidifferences of the retarded and of the advanced fields created by all charges exactly vanishes. This introduced us into the frame of concepts of a global character involving the Universe.

Quite naturally we were thus led, especially after conversations with G. Contopoulos and C. Efthymiopoulos, to consider the analogy between the roles far fields play in electrodynamics and in gravitation theory. Obviously, this analogy is evident to everybody, and was particularly pointed out for example by Einstein himself in his Princeton lectures (Einstein 1922), when he was commenting on the fact that the perturbation to the flat metric satisfies the d'Alembert equation, so that even in general relativity one should deal with retarded potentials, as in electrodynamics. He also commented how one might think of implementing in such a way Mach's idea on the role of faraway matter. However, he estimated that the inertial force due to faraway matter was too small. Now, such considerations were made before Hubble's law on galaxies recession was established. In brief, our idea just amounts to implement Einstein's estimate when Hubble's law is assumed as a phenomenological fact.

This idea was implemented in the paper (Carati Cacciatori & Galgani 2008), where it was shown how the gravitational effect of the faraway matter vanishes if the matter is assumed to be uniformly distributed, while it does not if the distribution is assumed to be fractal. In fact the latter hypothesis had been previously suggested by some authors on an observational basis (Sylos Labini *et al.* 1998), although there is an open debate on this point. So, taking a fractal distribution of dimension 2, the force per unit mass on a test particle was shown to have the typical value of $0.2cH_0$, which appears to agree with the observed one (Milgrom 1983, Milgrom & Beckenstein 1987). An application to the velocity dispersion in the Coma cluster of galaxies was also given in that paper.

Finally, in the work (Carati 2011) an estimate of the gravitational effect of faraway matter on the rotation curves of spiral galaxies was given. To this end, a new assumption was needed, namely that of the field due to distant matter should be uncorrelated beyond a length l, which constitutes a free parameter of the theory. With such an assumption is was shown that the theory fits pretty well the observations, with l of the order of 1 kpc. And this, not only for the most common cases in which the rotation curves decay more slowly than expected from the Newtonian action of the local matter, but also for the few cases in which the decay is faster, to which dark matter cannot provide a solution.

So our "theory" is of a conservative character, entirely framed within classical general relativity, the only new idea being that of estimating the gravitational action of matter when the latter is described as constituted of a discrete system of point–galaxies which are assumed on empirical grounds to obey Hubble's law. The computations then show that the relevant contribution comes from the galaxies which are near the border of the visible Universe, which we call here the faraway matter. Actually, their gravitational action turns out to be negligible if matter is assumed to have a homogeneous distribution, while turns out to be of the correct order of magnitude if the distribution is assumed to be fractal of dimension 2. So our "theory" has little to do with other ones, such

as for example MOND (Milgrom 1983) which constitutes a kind of "effective" theory, or TeVeS (Beckenstein 2004) which is one among the theories alternative to classical general relativity.

In the present paper a short review of the works (Carati Cacciatori & Galgani 2008) and (Carati 2011) is given. In section 2 the analogy between far fields in electrodynamics and in gravitation theory is recalled. In section 3 the model is described, and the first significant result is illustrated, namely, a deduction of the FRW metric. Moreover, the Friedmann–Robertson–Walker (FRW) metric is deduced as a mean metric, and an estimate is obtained for an effective local density, which turns out to be five times the observed one. Finally, the applications to velocity dispersion in clusters of galaxies and to rotation curves in spiral galaxies are illustrated in section 4.

2.   Far fields in electrodynamics and in gravitation theory

The relevance of faraway matter can be illustrated through the example of the electromagnetic field. As is well known, the electromagnetic field due to a charge e can be split, according to Maxwell's equations, as the sum of two terms:

- The Coulomb field (or near field) $E \simeq \frac{e}{r^2}$

- The far field $E \simeq \frac{ea}{c^2 r}$, where a is the charge acceleration and c the speed of light.

Clearly, the electromagnetic interaction between distant bodies just reduces to the far field, which decreases with the distance much more slowly than the Coulomb one, and manifests itself as a radiation emitted by the source. If one tries to take into account the radiation emitted by all charges present in the Universe, one meets with paradoxes as that of Olbers, which is just due to the radiation emitted by the faraway objects. It is true that modern cosmological theories allow one to escape such a paradox, but at any rate the far away sources still play a role in producing some background field, the $3 \, {}^0K$ cosmic background radiation. In a similar way we will try to take into account the background gravitational field due to distant galaxies, estimating its magnitude, and discussing its possible effects.

First of all one has to recall that, in the electromagnetic case, the far field comes into play because the 4–potential A_μ due to a point charge, of position vector $\mathbf{q}$, is a solution of the d'Alembert equation

$$\Box A_\mu = 4\pi e \dot{\mathbf{q}}_\mu \, \delta(\mathbf{x} - \mathbf{q})$$

($\delta()$ being the Dirac delta function), i.e., is a relativistic effect. So, in order to compute the gravitational effects of far away matter, one has to go beyond Newton's theory, and make use of general relativity.

This cannot be done in full generality, as the full problem is intractable. In fact, in computing the field produced by distant sources, one is faced with a non linear coupled problem, namely,

- The gravitational field is a solution of the Einstein equations having as source the energy–momentum tensor corresponding to the galaxies, dealt with as "particles";

- the motion of the particles (and thus the corresponding energy–momentum tensor) is determined by the force field, that they themselves create as sources

3.   The model. First result: the effective FRW metric and the effective matter density

The problem is too complicated, and no one is able to say anything definite about it. One can simplify it, as we will do, by making the following assumptions:

1. the motion of the galaxies is assigned, according to the observation (this corresponds in electrodynamics to the antenna problem, in which the currents are assigned)

2. the Einstein field equation is linearized.

The motion of the galaxies (of position vectors $\mathbf{q}_j$) is assigned according to Hubble's Law $\dot{\mathbf{q}}_j = H_0\,\mathbf{q}_j$, where, for the sake of simplicity, Hubble's constant H_0 is taken time–independent. So, if one thinks of the galaxies as point sources, for the energy–momentum tensor one gets the expression

$$T^{\mu\nu} = \sum_{j=1}^{N} \frac{1}{\sqrt{g}}\,\frac{M_j}{\gamma_j}\,\delta(\mathbf{x} - \mathbf{q}_j)\dot{\mathbf{q}}_j^{\mu}\dot{\mathbf{q}}_j^{\nu}$$

where N is the number of galaxies, of mass M_j, and γ_j is the usual Lorentz factor, while g is the modulus of the determinant of the metric. All derivatives are meant with respect to proper time. In this expression one has to think of the positions $\mathbf{q}_j$ of the galaxies as distributed at "random", i.e., the vectors $\mathbf{q}_j$ are random variables, distributed according to some definite law. The galaxy masses M_j too could in principle be thought of as distributed at random, but we will instead take them all equal, just in order to simplify the discussion.

The second step is the linearization of Einstein's equation for the metric $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ about an unperturbed solution $\eta_{\mu\nu}$. For $\eta_{\mu\nu}$ we take the vacuum solution, i.e., the Minkowsky metric, so that the perturbing metric $h_{\mu\nu}$ has to satisfy the equation (with G the gravitational constant)

$$\Box\big[h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h\big] = -\frac{16\pi G}{c^4}T_{\mu\nu} \ ,$$

i.e., essentially the d'Alembert equation, with the energy–momentum tensor $T_{\mu\nu}$ as source. Then the gravitational force will contain a far field term as in the electromagnetic case.

However, before addressing this problem in the next section, we will preliminarily look here at the expression for the metric. As solution of the Einstein equations we consider here (as commonly made in electromagnetism) the retarded one. This is given by

$$h_{\mu\nu} = \frac{-2G}{c^4}\ \sum_{j=1}^{N} \frac{M_j}{\gamma_j}\ \frac{2\dot{q}_\mu^{(j)}\dot{q}_\nu^{(j)} - c^2\eta_{\mu\nu}}{|\mathbf{x} - \mathbf{q}_j|}\Bigg|_{t=t_{\mathrm{ret}}} \ , \tag{1}$$

where the time is the retarded one, i.e., $t_{\text{ret}} = t - |\mathbf{q}_j - \mathbf{x}|/c$, which is a function of the position $\mathbf{x}$. We note in passing that this is not the unique solution, because for example one could consider the semisum of the retarded and the advanced potentials, as was recently done by Romero and Pérez (Romero & Pérez 2011). Now, with the retarded solution one deals with the observed positions of galaxies at time t, while the advanced solution requires to know their positions in the remote future, which are unknowable. Thus the estimate of the advanced potentials is quite difficult, and we choose the retarded ones.

Now the metric $h_{\mu\nu}$ is a random variable, because such were assumed to be the positions of the source–galaxies. Then, on averaging, one can get the "mean metric", which should give the properties of the metric in the large: the actual value of the metric will "fluctuate" about the mean value, and such a fluctuation turns out to produce the peculiar effects we will describe later.

If one assumes the distribution of galaxies to be isotropic, for the mean metric one gets the expression

$$\langle\, g_{\mu\nu}\,\rangle\, dx^\mu dx^\nu = (1 - \alpha - 3\beta)\, c^2 dt^2 - (1 + \alpha + \beta)dl^2 \;,$$

where $dl^2 = dx^2 + dy^2 + dz^2$, and

$$\alpha = \frac{2G}{c^2}\,\langle\, \sum_j \frac{M_j}{|\mathbf{q}_j|}\,\rangle\;,\quad \beta \lesssim \frac{4GH_0^2}{3c^4}\,\langle\, \sum_j M_j|\mathbf{q}_j|\,\rangle\;. \tag{2}$$

So the mean metric turns out to be a Friedmann–Robertson–Walker one.

We meet here with a consistency problem, because in a Friedmann–Robertson–Walker metric the Hubble constant is related to the coefficients α and β by the relation

$$H_0 = \frac{1}{2}\frac{d}{dt}\log\frac{1 + \alpha + \beta}{1 - \alpha - 3\beta}\;.$$

So the value of the r.h.s. has to coincide with the value of H_0 we have assumed phenomenologically for the motion of the galaxies.

On the other hand, one also meets with a big difference with respect to the usual treatment in which matter is dealt with as a continuum. Indeed, in our case α and β depend heavily on the distribution of faraway matter (as the expression (2) explicitly shows), while in the continuum approximation they depend only on the local density. In fact the sums in (2) diverge, as one sees for example in the case of a uniform distribution of galaxies, so that the larger contribution comes from the faraway galaxies.

Introduce an effective density of matter ϱ_{eff}, defined by requiring that one has

$$\langle\, \sum \frac{M_j}{|\mathbf{q}_j|}\,\rangle \simeq 4\pi\varrho_{\text{eff}}\,\frac{R_0{}^2}{2}\;,\quad \langle\, \sum M_j|\mathbf{q}_j|\,\rangle \simeq 4\pi\varrho_{\text{eff}}\frac{R_0{}^4}{4}$$

where R_0 is the "radius of the universe" (or better of our chart of it).

Then one finds that, between effective density and Hubble's constant, one has the relation

$$\varrho_{\text{eff}} \simeq \frac{1}{4}\frac{3H_0^2}{8\pi G}\;.$$

Thus, using the accepted value for H_0, one finds

$$\varrho_{\text{eff}} \simeq 5\varrho_0 \ ,$$

where ϱ_0 is the (estimated) present density of visible matter. So, our model can fit the observations if the contribution of the distant galaxies (in principle divergent, as remarked above) is four times the contribution of local visible matter. This is the first effect in which the contribution of distant objects might replace the contribution of local dark matter.

4.   Application to the velocity dispersion in clusters of galaxies, and to the rotation curves in spiral galaxies

To describe the effect faraway matter has on the velocity dispersion in clusters of galaxies and on the rotation curves of spiral galaxies, we have to compute the "gravitational force" due to the distant galaxies.

We recall that the gravitational field affects the motion of a test particle inasmuch as the motion satisfies Lagrange equations with a Lagrangian that involves the metric tensor, namely,

$$L = g_{\mu\nu} \frac{\mathrm{d}x_\mu}{\mathrm{d}\tau} \frac{\mathrm{d}x_\nu}{\mathrm{d}\tau} \ .$$

So, the equations of motion of a test particle, with position vector $\mathbf{x} = (x_1, x_2, x_3)$ have, for small velocities, the form

$$\ddot{x}_k = -\partial_k h_{00} - \frac{1}{2c}\partial_t h_{0k} + \text{smaller terms} \stackrel{\text{def}}{=} f_k \ .$$

One can check that

- $\partial_k h_{00}$ corresponds to the Newtonian force $\simeq 1/r^2$

- $\frac{1}{2c}\partial_t h_{0k}$ corresponds to the far field $\simeq a/r$, a being again the acceleration of the source.

The analogy with electromagnetic theory is thus complete.

From (1), it is possible to estimate the force per unit mass, $\mathbf{f}$, acting on a test particle. The most important term is that given by the far field. Notice that the acceleration of a galaxy can be obtained by differentiating Hubble's law, so that not only velocities, but also accelerations increase linearly with the distance. In such a way one finally gets

$$\mathbf{f} = \frac{4GH_0^2}{c^2} \sum_{j=1}^N M_j \frac{\mathbf{q}_j}{|\mathbf{q}_j|} \ . \tag{3}$$

Notice that, again, $\mathbf{f}$ is a random vector because such are the position vectors $\mathbf{q}_j$. So, for what concerns the mean, assuming an isotropic probability distribution of the galaxies one has

$$\langle \, \mathbf{f} \, \rangle = 0 \ ,$$

as expected. Now, although having a vanishing mean, $\mathbf{f}$ is not a vanishing quantity at all. In fact, for a random variable X with vanishing mean, one knows that its typical value is given by its standard deviation σ_X, in our case σ_f. At this point, in order to compute the variance σ_f^2 of $\mathbf{f}$, a crucial role is played by the probability distribution of the galaxies. Two cases may be considered.

1. The positions of the galaxies can be assumed to be independent and identically distributed (as for gases). Then $\mathbf{f}$ is the sum of N independent identically distributed random variables, so that its variance σ_f^2 is simply the sum of the variance of each term. In this way one gets

$$\sigma_f \simeq cH_0/\sqrt{N} \simeq 0 \ ,$$

   i.e., the force due to the faraway matter is negligible.

2. Some authors (Sylos Labini *et al.* 1998) have proposed that the distribution of the galaxies is a fractal, with dimension $D \simeq 2$. This means that the galaxy positions are correlated, so that the computation of σ_f^2 is no more immediate. In any case, σ_f^2 can be estimated numerically by generating samples from a distribution of points with fractal dimension 2. We generated such a distribution using the recursive relation

$$\mathbf{q}_{j+1} = \mathbf{q}_j + \mathbf{z}$$

   where $\mathbf{z}$ is a random vector with a Gaussian distribution. A numerical estimate of the standard deviation σ_f gives

$$\sigma_f \simeq 0.2 \, cH_0 \ ,$$

   Thus, in the fractal case the contribution of faraway matter is no more negligible. In fact, it is even of the order of magnitude of that ascribed to dark matter (Milgrom 1983, Milgrom & Beckenstein 1987), both in clusters of galaxies and in connection with the rotation curves of spiral galaxies, as will be discussed below.

So we have estimated the size of the force (per unit mass) due to distant matter, under the fractal assumption. However, one has to recall that the relevant force which effectively acts on an object within a system is the tidal one, namely, $\mathbf{f} - \mathbf{f}^*$, where $\mathbf{f}$ is the force acting on the object, and $\mathbf{f}^*$ that acting at the center of mass of the considered system. Indeed, in the presence of a locally constant force field, according to the equivalence principle (think of Einstein's lift example), the locally constant field can be eliminated by a suitable change of coordinates. Now, the relevant force $\mathbf{f} - \mathbf{f}^*$ turns out to be completely different in the two cases, smooth or nonsmooth.

Indeed, let us estimate the variance σ^2 of $\mathbf{f} - \mathbf{f}^*$ in the two cases. In the smooth case one can estimate $\mathbf{f} - \mathbf{f}^*$ by Taylor expansion, getting

$$\sigma^2 \simeq H_0^2 L^2$$

where L is the linear dimension of the system (the cluster of galaxies, or the galaxy). This contribution is found to be totally negligible for the case of the

"Coma" cluster, and also for the galaxies we studied. On the other hand, for the variance of $\mathbf{f} - \mathbf{f}^*$ in general one has

$$\sigma^2 = 2\,\sigma_f^2 - 2C(\mathbf{f}, \mathbf{f}^*) \ ,$$

where

$$C(\mathbf{f}, \mathbf{f}^*) = \langle \mathbf{f} \cdot \mathbf{f}^* \rangle$$

is the correlation of the two considered quantities,, $\langle\rangle$ denoting mean value. Thus, under the nonsmoothness assumption, which means that the correlation of the tow forces $\mathbf{f}$ and $\mathbf{f}^*$ is negligible, the size of the tidal force is just equal to $\sqrt{2}$ the size of the force itself.

Having determined the size of the tidal force under the decorrelation assumption, there remains now the problem of its global effect on the system considered. The most significant cases are those in which the force field acts as a pressure or as a tension, i.e., the cases in which the force field is locally predominantly centripetal or centrifugal, respectively. Actually one expects that, in the Universe, local structures be formed where the far field conspires to produce pressure. Obviously this remark raises a consistency problem for a possible future more complete theory, in which one abandons the simplifying assumption considered in the present "theory", in which the motions of the sources were taken to be assigned (according to the phenomenological Hubble law). At the present level of approximation, we just consider the choice of the direction (pressure or tension) as a free element of the theory, to be determined form observations. Obviously, one expects that in the large majority of cases, a pressure will be found.

We can now come to a description of the results for the clusters of galaxies and for the rotation curves of spiral galaxies.

In order to estimate the contribution to the virial of a cluster of n galaxies (Carati *et al.* 2008), one starts from the virial theorem, and for the variance σ_v^2 of the velocity on gets

$$n\sigma_v^2 \stackrel{\text{def}}{=} \overline{\sum_i v_i^2} = \overline{\sum_i (\mathbf{f}_i - \mathbf{f}^*) \cdot \mathbf{x}_i} \ ,$$

where $\mathbf{f}_i$ is the force (per unit mass) due to the faraway matter acting on the i–th galaxy of the cluster located at $\mathbf{x}_i$, while $\mathbf{f}^*$ is the value of the force field at the center of mass of the cluster.

Assuming that the forces acting on the different galaxies of the cluster be uncorrelated, and moreover that the force field acts as a pressure, which helps stabilizing the cluster[3], one gets

$$\sigma_v^2 \simeq 0.07 \ cH_0L \ . \tag{4}$$

For the Coma cluster (L of the order of 1 Mpc) this formula gives $\sigma_v = 900$ km s^{-1} against an observed value of 700 km s^{-1}. So the contribution of faraway matter can explain the measured value of the variance σ_v^2, without any need of dark matter.

---

[3] One should recall that the gravitational force due the visible matter of the cluster is utterly unable to keep the cluster together.
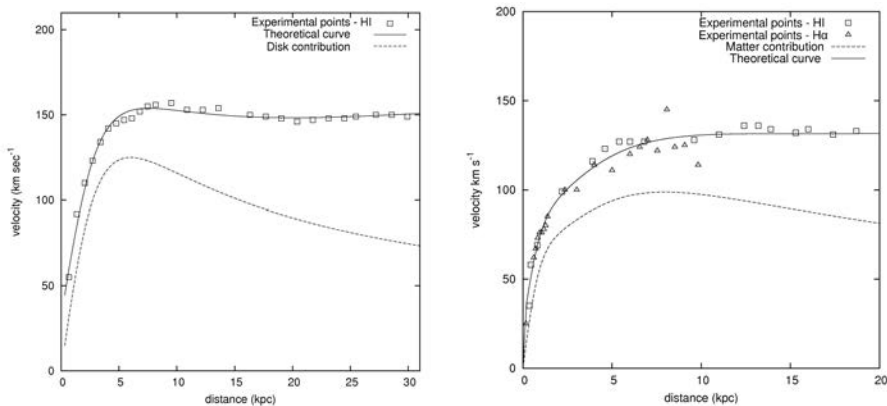
Figure 1. The rotation curves for the galaxy NGC 3198 (left) and NGC 2403 (right). Solid line is the theoretical curve with the contribution of faraway matter taken into account, dashed line refers to the contribution of the local matter.
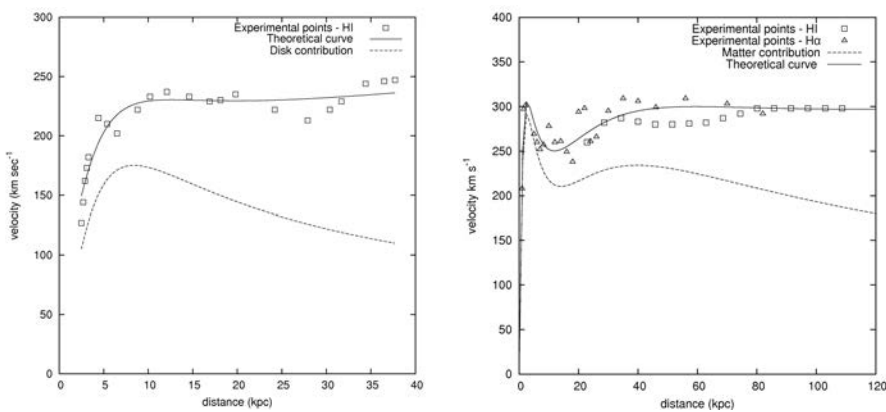

Figure 2. The rotation curves for the galaxy NGC 4725 (left) and UGC 2885 (right). Solid line is the theoretical curve with the contribution of faraway matter taken into account, dashed line refers to the contribution of the local matter.
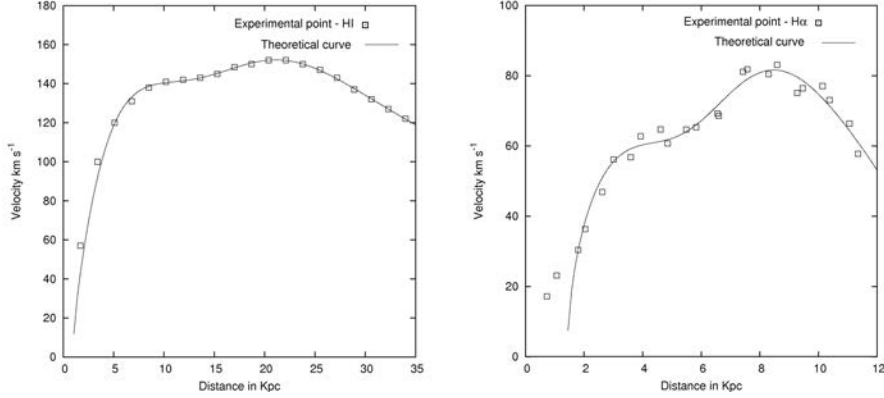
Figure 3.   The rotation curves for the galaxy NGC 864 (left) and AGC 400848 (right). Solid line is the theoretical curve with the contribution of faraway matter taken into account. The decrease of the rotation curve is faster then keplerian.

Table 1.   Value of the correlation l, for four galaxies.

Galaxy	Mass	Mass/Luminosity	Correlation length
NGC 3198	$4.0\ 10^{10}\ M_{\odot}$	$4.6\ M_{\odot}/L_{\odot}$	0.6 kpc
NGC 2403	$3.5\ 10^{10}\ M_{\odot}$	$4.4\ M_{\odot}/L_{\odot}$	0.8 kpc
UGC 2885	$1.0\ 10^{12}\ M_{\odot}$	$2.1\ M_{\odot}/L_{\odot}$	1.7 kpc
NGC 4725	$1.1\ 10^{11}\ M_{\odot}$	$2.1\ M_{\odot}/L_{\odot}$	3.1 kpc

For what concerns the rotation curves in spiral galaxies (Carati 2011), at variance with the case of the clusters of galaxies one cannot forget the contribution of the local visible matter, which is the larger one. This force can be taken in the nonrelativistic approximation, i.e., expressed in terms of a potential V^{loc}, which depends on the distribution of the local matter.

In order to describe the effects of the force due to the faraway matter in this case, the treatment requires to assume a decorrelation property, i.e., that the correlation decreases exponentially on a certain scale l, which plays the role of a free parameter to be determined by fits with the observations.

In this case, one gets for the speed v of rotation the expression

$$\frac{3}{2}\frac{v^2}{r} = -\partial_r V^{loc}(r) - \partial_r V^{eff} \ , \tag{5}$$

where $\partial_r V^{eff}$ comes from the contribution of the faraway galaxies. This term has to be understood not as a derivative of a potential, but as a random term with vanishing mean, the standard deviation of which can be estimated. One finds

$$\partial_r V^{eff} \simeq \pm 0.2 H_0 c \sqrt{\frac{l}{r}} \ . \tag{6}$$

We still have a free choice for the sign of this term, which entails either a pressure, which helps keeping the galaxy together, or a tension, which tends to break it apart. One can conjecture that the positive sign has to appear more often. Actually, in the literature there are reported observations for a small percentage of galaxies, in which the rotation curves decrease faster than expected from the Newtonian action of the local visible matter, which means either that the galaxy is expanding, or that there is a force acting on the system as a tension.

One can obtain the order of magnitude of the parameter l, by fitting the observation with our formula. We report in Table 1 the value of l obtained by fitting the rotation curves of very different galaxies. One can check that the order of magnitude is always the same, $l \simeq 1$ kpc.

5. Conclusions

So we have shown how the the gravitational action of faraway matter may explain the two classical phenomena for which local dark matter was first introduced. Other phenomena exist which are explained in terms of dark matter, and we hope to tackle them in the future. Obviously, the present "theory" introduces some hypotheses, such as fractal structure of the Universe or decorrelation properties of the gravitational field, as usual with scientific theories.

References

Bekenstein, J. D. 2004, Phys. Rev. D, 70, 083509

Carati, A. 2011, Gravitational effects of the faraway matter on the rotation curves of spiral galaxies, arXiv: 1111.5793

Carati, A., & Galgani, L. 2003, Nuovo Cimento, 118 B, 839

Carati, A., & Galgani, L. 2004, Int. Journ. Mod. Phys. B, 18, 549

Carati, A., Cacciatori, S.L., & Galgani, L. 2008, Europhys. Letts., 83, 59002

Einstein, A. 1922, The meaning of relativity, Princeton

Marino, M., Carati, A., & Galgani, L. 2007, Annals of Physics, 322, 799

Milgrom. M. 1983, ApJ, 270, 365

Milgrom, M., & Beckenstein, J. 1987, in UAI Symposium 117, Dark Matter in the Universe, ed. Kormedy, J., & Knapp, G.R., Dordrecht, 319

Romero, G.E., & Pérez, D., 2011, Int. J. Mod. Phys. D 20, 2831

Sylos Labini, F., Montuori, M., & Pietronero, L. 1998, Phys. Rep., 293, 61

Wheeler, J.A., & Feynman, R. 1945, Rev. Mod. Phys., 17, 157

Part II

Oral contributions

ORAL COMMUNICATION

Renormalization tools to study the loss of stability in the Area Preserving Maps

A. Olvera

IIMAS, UNAM

Abstract. The renormalization method for area preserving maps was introduced by R. S. MacKay in 1982 to study critical invariant circles. Similar ideas appear in the study of critical phenomena in physics (MacKay 1983). The renormalization scenario for the breakdown of golden invarian circles in the twist map is described in this paper. In the first part we describe the renormalization methods introduced by MacKay. In the second part, the obstruction criterion (Olvera & Simó 1987) is described and we show that the renormalization group and the obstruction criterion can work together (De la Llave & Olvera 2006).

1. Introduction

The global stability of the area preserving twist map is connected to the existence of non trivial invariant curves. From the beginning of the last century, the study of the dynamics of twist maps was related to the detection of the breakdown of the invariant circles. For a one parameter family of area preserving twist maps, the Twist theorem shows the existence of invariant circles for small values of the parameter. Nevertheless, the main interest has been to determine the value of the parameter when the last invariant circle disappears. In the last 40 years, many theories were developed to determine the breakdown of the last invariant circle. Many researchers like Greene (1979) , MacKay (1983), Shenker & Kadanoff (1982), Aubry (1983), Mather (1982), etc. developed new ideas to obtain the critical value of the parameter which defines the loss of global stability. Shenker & Kadanoff (1982) found an interesting scaling behavior in the neighborhood of the breakdown of invariant circles (Shenker & Kadanoff 1982), MacKay introduced the renormalization theory in order to explain amazing scaling behavior (MacKay 1983). The residue method, introduced by Grenne, gives an empirical method to determine the critical value of the parameter in a very accurate way (Greene 1979). In the past years, there were many attempts to formalize the residue method but none of those ideas were successful (Delshams & De la Llave 2000 and Falcolini & De la Llave 1992).

The renormalization approach was introduced by Shenker & Kadanoff (1982). They found that there exist self-similarity phenomena in the rupture of invariant circles. In that case, scaling exponents which are associated with the self-similarity behavior are a universal property because it is shared by a wide class of area preserving maps.

The dynamics in the vicinity of the invariant circles can be understood in terms of the set of periodic orbits which are close to the invariant circle. An important role is played by the invariant manifolds of the hyperbolic periodic orbits in the existence and rupture of invariant circles. The topological structure of these manifolds define conditions to allow or prevent the existence of invariant circles. Olvera & Simó (1987) introduced the obstruction criterion to determine the critical value of the parameter. The main idea of this method is to show that the heteroclinic intersection of invariant manifolds of periodic orbits, which are close to the invariant circle, gives a topological obstruction to the continuity of the invariant circle. In this form, the method provides a lower bound for the parameter which determines the non-existence of invariant circles between these periodic orbits. The method can be applied to any pair of periodic orbits which approximate the invariant circle. In that way, it is possible to define a sequence of lower bounds whose limit corresponds to the critical value of the parameter. The precision of this method is similar to the Greene's residue method, nevertheless the obstruction criterion gives formal bounds for the non existence of invariant circles.

The aim of this paper is to show how the obstruction criterion is related to the renormalization group. In the space of twist maps, which depend on one parameter and contain a golden invariant circle (invariant circle with rotation number equal to the golden mean), we can define a manifold which contains those maps which have heteroclinic tangencies of the invariant manifold of the approximants of the golden invariant circle. The main point is to show that this manifold has a transversal intersection with the unstable manifold of the non trivial point of the renormalization group. In particular, de la Llave & Olvera (2006) showed that the upper part of the unstable manifold of the non trivial fixed point has hyperbolic dynamics.

2.   Hamiltonian systems and twist maps

Hamiltonian systems with two degrees of freedom are singular mechanical systems because their invariant tori are topological barriers which separate phase space. This property is no longer preserved for a system with more degrees of freedom. KAM theory allows to determine the existence of invariant tori for Hamiltonian systems which are close to an integrable system. It is common to write non integrable Hamiltonian systems as $H = H_0 + \lambda H_1$, where H_0 is an integrable Hamiltonian system and H_1 is a Hamiltonian perturbation. KAM theory shows that the measure of invariant tori of the system H goes to the total measure if λ goes to zero. Nevertheless, we cannot use KAM theory if we are interested in finding the value of λ for which the last invariant torus breaks down.

The Hamiltonian system can be reduced to a discrete dynamical system using a Poincaré section (Chirikov 1979). In the case of a Hamiltonian system with two degrees of freedom, the corresponding discrete dynamical system is an area preserving twist map in an open set of the phase space. As with Hamiltonian systems, we can represent the twist map as the sum of an integrable twist map and a perturbation function. The size of the perturbation function depends on

one parameter λ, it is written as a combination of the set of parameters which define the dynamical system. In this form, the discrete dynamical system is a one parameter family of area preserving twist maps. The phase space of the twist map is the cylinder $\mathbb{R} \times \mathbb{S}$ which has a radial variable $r \in \mathbb{R}$ and an angular variable $\theta \in \mathbb{S}$. The integrable twist map is then written as:

$$
\begin{aligned}
r_{n+1} &= r_n \ , \\
\theta_{n+1} &= \theta_n + \varrho(r_{n+1}) \ ,
\end{aligned}
\qquad n = 1, 2, \ldots \qquad (1)
$$

where $\varrho(r)$ is a smooth increasing function. The perturbed twist map, which is known as the whisker map (Chirikov 1979), is defined in this form,

$$
F_w : \left\{
\begin{aligned}
r_{n+1} &= r_n + f(r_n, \theta_n, \lambda) \ , \\
\theta_{n+1} &= \theta_n + \varrho(r_{n+1}) + g(r_{n+1}, \theta_n, \lambda) \mod 1 \ .
\end{aligned}
\right.
\qquad n = 1, 2, \ldots \quad (2)
$$

The perturbation functions f and g depend on the perturbation parameter λ. These functions go to zero when $\lambda \to 0$. f and g are smooth functions such that the whisker map preserves area.

An important feature of the twist map is the existence of periodic orbits, the sequence $\{r_i, \theta_i\}$ is a periodic orbit with period N if $F_w^N(r_0, \theta_0) = (r_0, \theta_0)$ and $(r_i, \theta_i) \neq (r_0, \theta_0)$ for $i = 1, \ldots, N-1$. We can assign a rotation number ϱ to any periodic orbit using the lift of the twist map (Katok & Hasselblatt 1995), which corresponds to extending the angular variable to the real line. In that form, the rotation number ϱ has the following definition:

$$
\varrho = \lim_{i \to \infty} \frac{\theta_i - \theta_0}{i} \ , \qquad (3)
$$

where $\theta \in \mathbb{R}$.

The whisker map has infinitely many fixed points known as resonances. We can expand around any of these resonances in order to get a simpler map; this map is called standard map (Chirikov 1979),

$$
F_\lambda : \left\{
\begin{aligned}
r_{n+1} &= r_n + \lambda \sin(2\pi \theta_n) \ , \\
\theta_{n+1} &= \theta_n + r_{n+1} \pmod 1 \ ,
\end{aligned}
\right.
\qquad n = 1, 2, \ldots \qquad (4)
$$

The whisker map can be interpreted as an infinite sequence of standard maps where the value of the parameter λ is changed in a continuous way.

The standard map has infinitely many periodic orbits, the rotation number of any periodic orbit is a rational number. These orbits are ordered, in the radial direction of the cylinder, by the rotation number. The Birkhoff fixed point theorem (Birkhoff 1950), shows that there exists a pair of periodic orbits for any rational rotation number ϱ which belongs to some interval $I_q \subset \mathbb{Q}$ and the linear stability of these orbits shows them to be hyperbolic and elliptic.

The invariant tori of the hamiltonian flow become invariant circles around the cylinder of the corresponding twist map. The standard map is foliated by invariant circles when $\lambda = 0$, which corresponds to the integrable case. In that case, there is an invariant circle for each $\varrho \in I_r \subset \mathbb{R}$. When $\lambda > 0$, the invariant circles with rational rotation numbers disappear. The Twist theorem shows the existence of invariant circles, with Diophantine rotation numbers, for small values of λ (Moser 1973). The shape of the invariant circles was studied by Birkhoff who showed that the shape of these circles coincides with a graph of a function (Birkhoff 1950).

In the eighties, the Aubry-Mather theory showed a variational description of the twist maps (Katok & Hasselblatt 1995). The periodic orbits, which belong to the Birkhoff type (for this kind of orbit, the coordinate θ in the lift of the map, is an increasing function), correspond to an extremal solution of a variational equation. The minimal type solutions correspond to periodic orbits with hyperbolic linear stability and the min-max type solutions are related to periodic orbits with elliptic linear stability. A very important result is the following: for any sequence of Birkhoff periodic orbits, with rotation numbers converging to an irrational number ϱ, the limit of the sequence exists and it corresponds to a minimal (or min-max) type solution. For the minimal case, the solution corresponds to an invariant circle with rotation number ϱ. On the other hand, the min-max case corresponds to an invariant circle which has infinitely many discontinuities; this kind of orbit is known as a cantorus.

3. Breakdown of noble invariant circles

KAM theory is a useful tool to determine the existence of invariant circles with small values of the parameter in twist maps; however to determine the breakdown of the last invariant circle, we have to use a different mathematical tool. The Aubry-Mather theory shows that any invariant circle can be approximated by a sequence of periodic orbits (Katok & Hasselblatt 1995), such that their rotation numbers converge to the rotation number of the invariant circle. Numerical experiments have shown that the invariant circles with noble rotation numbers, are the most robust circles of the twist maps (Reichl 1992). The continued fraction expansion of the noble numbers is a periodic (or eventually periodic) sequence of one natural number, in particular, the expansion of the golden mean number, ϱ_g, is given by an infinite sequence of ones. The approximations of ϱ_g are the sequence $\{\varrho_n\}|_{n=1}^{\infty}$, where $\varrho_n = \mathcal{F}_{n+1}/\mathcal{F}_n$ and $\mathcal{F}_n$ is the n-th Fibonacci number. From the numerical experiments with the standard map, we can conclude that the last invariant circle has a rotation number equal to ϱ_g (Greene 1979).

Following the results of Senker and Kadanoff in the seventies, J. Greene found an amazing result for the breakdown of the golden mean invariant circle in the standard map. Greene's method computes the critical value of the parameter of (4). In this method we have to compute the residue of the periodic orbits which approximate the invariant circle. The residue of a periodic orbit with $\varrho = M/N$, is given by

$$Re = \left[2 - \mathrm{Tr}(\mathrm{Jac}(F^N(x)))\right]/4 \quad , \tag{5}$$

where Jac is the Jacobian matrix of the map $F^N(x)$. Greene found that the residues of the elliptic periodic orbits, whose rotation numbers approach ϱ_g, converge to a constant value when λ achieves the value that corresponds to the breakdown of the last invariant circle. Greene's method is an heuristic procedure to find, in a very accurate form, the value of the parameter for which an invariant circle disappears. The method is the following: let $\{\varrho_n\}|_{n=1}^{\infty}$ be a sequence of rational rotation numbers which converge to ϱ, the rotation number of the invariant circle. Let $\{Re_n\}$ be the sequence of residues of the elliptic periodic orbits with rotation number ϱ_n for a fixed value of the parameter. If the limit of the residues goes to zero, then there exists an invariant circle with rotation number ϱ. If the limit diverges to infinity, the invariant circle has disappeared. When the limit goes to a real number in $(0, \infty)$, the value of the parameter corresponds to the breakdown of the invariant circle. In particular, Greene found that the residues converge to the value 0.2500888 for the break down of the golden invariant circle in the standard map, the corresponding value of the parameter is 0.97636... (De la Llave & Olvera 2006). In that case, the residues converge following a power law.

In the eighties, formal procedures to find the breakdown of the invariant circles were developed. Some of these methods use the fact that in the neighborhood of any invariant circles, there are no periodic orbits which are not of the Birkhoff type. Boyland and Hall developed a formal procedure using this idea (Boyland & Hall 1987). The Mackey's cone method follows a similar idea of using the non Birkhoff periodic orbits (MacKay & Percival 1985 and Leage & MacKay 1986). Mather used the variational theory of minimal orbits of the twist maps to show the existence and destruction of invariant circles (Mather 1982).

4.   Renormalization group

Greene's residue method has been used by a large group of people because this method is very accurate and simple to implement in a numerical form. Nevertheless, Greene's method is an heuristic procedure because it uses local information to determine the global behavior of the dynamical system. In the past years, there were many attempts to show, in a formal way, a justification of the method (Delshams & De la Llave 2000). Shenker and Kadanoff found a remarkable result related to the existence of universal scalings in the distribution of the iterates of the standard map on the critical invariant circle with rotation number ϱ_g (Shenker & Kadanoff 1982). From this result, MacKay introduced the renormalization group to give an explanation of this universal scaling behavior in the standard map (MacKay 1983). The idea was to justify the residue method using renormalization theory.

4.1.   Integrable case

The first step to show how the renormalization group works in the standard map, is to do the integrable case (Reichl 1992). When $\lambda = 0$, the standard map becomes an integrable map. The periodic orbits are located in straight lines, which are parallel to the angular axis. For any periodic orbit with rotation

number $\frac{p}{q}$, we can choose an initial point x_0 with coordinates $r = \frac{p}{q}$ and $\theta = 0$; we denote with x_s the point on the periodic orbit, which is closest to x_0. We can consider the set of periodic orbits which approximate the invariant circle with rotation number ϱ_g. The rotation numbers of the approximants of this invariant circle, are the ratio of the consecutive Fibonacci numbers $\mathcal{F}_n$: $\varrho_n = \frac{\mathcal{F}_{n-1}}{\mathcal{F}_n}$. If we denote by $x(n)_s$ the closest point to $(r, \theta) = (\varrho_n, 0)$ on the periodic orbit with rotation number ϱ_n, whose initial point is $(\varrho_n, 0)$, then there exists a matrix B_n which transforms the point $x(n)_s$ to $x(n-1)_s$, where B_n is a linear transformation defined in this form:

$$B_n = \begin{pmatrix} -\frac{W_n - W_{n-1}}{W_{n+1} - W_n} & 0 \\ 0 & -\frac{1}{W_{n-1}} \end{pmatrix} \quad \text{where} \quad W_n = \frac{\mathcal{F}_{n-1}}{\mathcal{F}_n} \quad . \tag{6}$$

If $B_n x(n)_s = x(n-1)_s$, where

$$x(n)_s = \begin{pmatrix} W_{n+1} - W_n \\ (-1)^n \frac{1}{\mathcal{F}_n} \end{pmatrix} \quad \text{then} \quad B_0 B_1 B_2 \dots B_n x(n)_s = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad . \tag{7}$$

Let B^* be the limit of the transformation B_n:

$$\lim_{n \to \infty} B_n = B^* = \begin{pmatrix} -\varrho_g^2 & 0 \\ 0 & -\varrho_g \end{pmatrix} \quad . \tag{8}$$

The points $x(n)_s$ converge to zero in the form $x(n)_s = C n^{-\varrho_g^2}$. The elements in the diagonal of the matrix B^* are the scaling exponents of the sequence of transformations B_n.

The form of the integrable map, in the neighborhood of the golden invariant circle, can be written as $G^M R^N$, where

$$G^M \begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} r \\ \theta + Mr \end{pmatrix} \quad , \quad R^N \begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} r \\ \theta - N \end{pmatrix} \quad , \tag{9}$$

where G^M is the M-th iteration of the map G and R is the map which shifts the angular variable to the interval $[0, 1)$.

MacKay introduced two operators called commuting pairs, T and U (MacKay 1983). These operators are defined as:

$$
\begin{aligned}
T_{i-1} &= G^{M_{i-1}} R^{N_{i-1}} \quad , \\
U_{i-1} &= G^{M_{i-2}} R^{N_{i-2}} \quad .
\end{aligned}
\tag{10}
$$

For the case of the golden invariant circle, M_i and N_i are the Fibonacci numbers, that is, $M_i = \mathcal{F}_{i+1}$, $N_i = \mathcal{F}_i$ and $\frac{N_i}{M_i}$ are the rational approximants of ϱ_g. The rescaling rules for the commuting pairs are these:

$$
\begin{aligned}
T_{i-1} &= B_{i-1} T_{i-1} U_{i-1} B_{i-1}^{-1} \quad , \\
U_{i-1} &= B_{i-2} T_{i-1} B_{i-2}^{-1} \quad .
\end{aligned}
\tag{11}
$$

When i goes to infinity, the commuting pairs reach their fixed point, such that:

$$
\begin{aligned}
T^* &= B^* T^* U^* (B^*)^{-1} \quad, \\
U^* &= B^* T^* (B^*)^{-1} \quad.
\end{aligned}
\tag{12}
$$

The composition of the fixed points of the commuting pairs give raise to the universal map which is known as the Universal Trivial Fixed Point:

$$
T^* U^* \begin{pmatrix} r \\ \theta \end{pmatrix} = \begin{pmatrix} r \\ \theta + (\frac{1}{\varrho_g} + 1)r - \frac{1}{\varrho_g} \end{pmatrix} \quad.
\tag{13}
$$

The map (13) gives the generic dynamics around the golden invariant circle of any integrable twist map; the rotation numbers of the golden circles are given in the form $\varrho_g c = [a_1, a_2, \ldots, a_n, (1)^\infty]$.

4.2. Critical case

Shenker and Kadanoff (Shenker & Kadanoff 1982) studied the dynamics close to the critical invariant circle with rotation number ϱ_g. They point out some crucial properties of the critical case. For the standard map, the critical value of the parameter is $\lambda_c = 0.9716354\ldots$ (this value was obtained using the Greene's method). This value is called critical because if $\lambda < \lambda_c$ there exists a golden invariant circle, and if $\lambda > \lambda_c$ the circle becomes a cantorus.

Consider an elliptic periodic orbit which is the i-th approximant of the golden invariant circle. Let λ_i be the value of the parameter such that the residue (5) of this orbit achieves the value $Re^* = 0.250088$. Then, for the critical case, we have the following scaling factors:

1. The set of $\{\lambda_i\}$ has the following property,

$$
\delta = \lim_{i \to \infty} \left(\frac{\lambda_i - \lambda_{i-1}}{\lambda_{i+1} - \lambda_i} \right) = 1.6280\ldots
\tag{14}
$$

2. Consider the i-th approximant of the golden invariat circle which corresponds to an elliptic periodic orbit with rotation number $\varrho_n = \mathcal{F}_{i-1}/\mathcal{F}_i$. Let $x(i)_s$ be the point of the periodic orbit which is closest to the vertical axis (where $\theta = 0$) for the value of the parameter λ_i. If r_i is the radial coordinate of $x(i)_s$ and θ_i the corresponding angular coordinate, then

$$
-\lim_{i \to \infty} \left| \frac{r_i - r_{i-1}}{r_{i+1} - r_i} \right| = \beta = -3.066\ldots
\tag{15}
$$

and

$$
-\lim_{i \to \infty} \left| \frac{\theta_i - \theta_{i-1}}{\theta_{i+1} - \theta_i} \right| = \alpha = -1.4148\ldots
\tag{16}
$$

The scaling factors α and β have a geometrical interpretation: For the standard map with λ_c, in the neighborhood of the critical golden circle, the corresponding phase space is shown in figure 1. We can observe that the graph is selfsimilar, the phase space which is inside of the rectangle is similar to the complete phase space. The corresponding scales are α in the vertical direction and β in the horizontal direction.
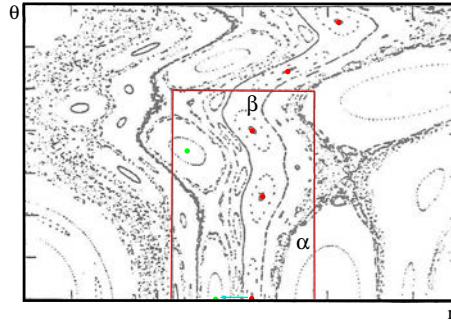

Figure 1.   Phase space of (4) map with value of the parameter λ_c. We notice that the phase space has the auto-similarity property.

Based on these geometric scalings, MacKay defined the transformation B^* in terms of the scaling factors α and β, such that,

$$B^* = \lim_{i \to \infty} B_i = \begin{pmatrix} \beta & 0 \\ 0 & \alpha \end{pmatrix} \quad . \tag{17}$$

The transformations B_i are defined in the same way as in the integrable case. In a similar way, Mackey introduced the commuting pairs for the critical case. If F_λ is the standard map (4) and $N_i/M_i = \mathcal{F}_i/\mathcal{F}_{i+1}$, the commuting pairs T and U are defined as:

$$
\begin{aligned}
T_{i-1} &= F_{\lambda_{i-1}}^{M_{i-1}} R^{M_{i-1}} \quad , \\
U_{i-1} &= F_{\lambda_{i-2}}^{M_{i-2}} R^{M_{i-2}} \quad ,
\end{aligned}
\tag{18}
$$

where F_λ^M is the M-th iteration of the map F_λ.

The properties (11) and (12) also hold for the critical case. The fixed point maps T^* and U^* are related with the universal map $F_{\lambda_g}^*$ in this form:

$$F_{\lambda_g}^* = \lim_{i \to \infty} B_i F_{\lambda_i}^{M_i} R^{N_i} B_i^{-1} \quad . \tag{19}$$

In renormalization language, the fixed point (U^*, T^*) is called the universal critical fixed point, which belongs to the infinite dimensional space of the twist maps. This fixed point has an unstable manifold of only one dimension $\mathcal{W}_*^u$, with eigenvalue $\delta = 1.6280\ldots$. The dominating attraction rate on its stable

manifold $\mathcal{W}_*^s$ is $\delta' = -0.6108\ldots$. The universal map (19) is related with the fixed point by expression (18).

The infinite dimensional space of area preserving twist maps, which has a golden circle, is separated by the stable invariant manifold $\mathcal{W}_*^s$ of the non-trivial fixed point (U^*, T^*). This is because that unstable manifold $\mathcal{W}_*^u$ is of codimension one. One of the directions of $\mathcal{W}_*^u$ connects with the trivial fixed point (13). In that form, any one-parameter family which belongs to this space is represented as a parametric curve. Therefore, the conjecture is that any family has one transversal intersection with this stable manifold $\mathcal{W}_*^s$ (see figure 2). The action of the renormalization group over any of these families, corresponds to a remap of the curve into a new curve which is closer to the invariant unstable manifold $\mathcal{W}_*^u$ of the critical fixed point. After an infinite number of iterations, any one-parameter family converges to $\mathcal{W}_*^u$. The maps which belong to the part of the curve which is below $\mathcal{W}_*^u$, converge to the trivial fixed point (13). The interpretation is that all members of the family, which are in this part of the curve, have an invariant circle with rotation number ϱ_g. On the other hand, the upper part of the curve is composed of maps which have a cantorus with rotation number ϱ_g. The elements of the family which are located at the intersection point between the curve and $\mathcal{W}_*^u$ converge to the non trivial fixed point. This means that this map has a critical invariant circle.
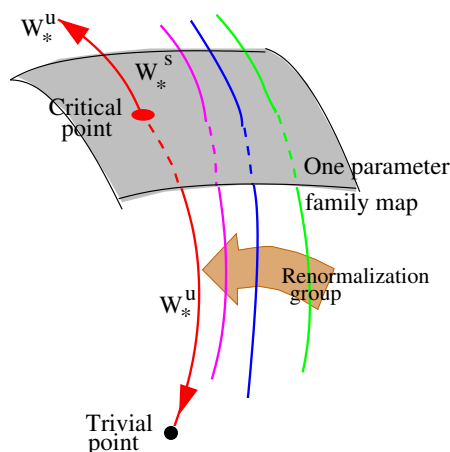

Figure 2. Schematic representation of the action of the renormalization group in the space of one-parameter families of area preserving twist maps which contain a golden circle

Figure 2 shows a scenario which was conjectured by MacKay about the action of the renormalization group over the space of one-parameter families of twist maps (MacKay 1983). The behavior of the critical point was studied in a numerical way. Nevertheless, there is no formal proof of the existence of the critical point. We point out that H. Koch showed the existence of the critical point for the renormalization problem of the continuous Hamiltonian system (Koch 2002). Another remark is that A. Stirnemann has developed, in a different form, the construction of the renormalization operator (Stirnemann 1993).

Instead of using the MacKay's commuting pairs, in this case, the renormalization group is defined with only one operator.

5. Invariant manifolds and the obstruction criterion

In 1984, Simó and Olvera developed an alternative way to show the non existence of invariant circles in the twist maps which they called the obstruction criterion (Olvera & Simó 1987). Using the fact that we can approach an invariant circle by periodic orbits (of Birkhoff type) whose rotation numbers converge to the rotation number of the invariant circle, the obstruction criterion takes into account the behavior of the invariant manifolds of the hyperbolic periodic points. In the pioneering work of Chirikov, he defined the resonance of a periodic orbit as the region between the stable and unstable invariant manifold of any hyperbolic point. In that form, the resonance of a hyperbolic periodic orbit looks like a sequence of pendulum's separatrices (Chirikov 1979). Using this scheme, Chirikov defined the overlap criterion to show the destruction of invariant curves, that is, if there is overlapping of resonances, then the invariant circles cannot exist in the overlap region. The obstruction criterion follows an idea similar to the Chirikov overlap criterion but is more precise in the definition of the overlap of resonances.

Let X_0 and X_1 be a set of hyperbolic periodic orbits of (2) of Birkhoff type. The corresponding rotation numbers of these orbits are ϱ_0 and ϱ_1 respectively. Every hyperbolic point which belongs to X_0 has a stable and unstable invariant manifold, $\mathcal{W}_0^s$ and $\mathcal{W}_0^u$. In a similar way, an hyperbolic periodic point of X_1 has its corresponding invariant manifolds, $\mathcal{W}_1^s$ and $\mathcal{W}_1^u$. It is well known that $\mathcal{W}_0^s$ and $\mathcal{W}_0^u$ intersect each other an infinite number of times. These manifolds look like a sequences of pendulum's separatrices (Moser 1973). The first tongue of the invariant manifold is defined as the initial part of the manifold, which departs from the hyperbolic point until the first turning point of the manifold. The main point of the obstruction criterion is that if the invariant manifolds of X_0 and X_1 have a heteroclinic intersection, then there cannot exist an invariant circle whose rotation number, ϱ, belongs to the interval (ϱ_0, ϱ_1). In that sense, the heteroclinic intersection forms an obstruction to the continuity of the invariant circle. When we consider a one-parameter family of area-preserving twist maps, like the (4) map, the obstruction criterion is defined as the heteroclinic tangency of the first tongue of $\mathcal{W}_0^s$ with the first tongue of $\mathcal{W}_1^u$. This tangency takes place when the parameter reaches the value λ_0, as seen in figure 3.

The obstruction criterion gives a formal lower limit for the non existence of invariant circles with rotation number ϱ_g. The periodic orbits which approximate the invariant circle have rational rotation number ϱ_n, such that,

$$\lim_{n\to\infty} \varrho_n = \varrho_g \quad \text{where} \quad \varrho_n < \varrho_g < \varrho_{n+1} \qquad \text{or} \quad \varrho_{n+1} < \varrho_g < \varrho_n \quad . \qquad (20)$$

We choose the sequence $\{\lambda_n\}_{n=0}^{\infty}$ which corresponds to the value of the parameter for which the periodic orbits, with rotation numbers ϱ_n and ϱ_{n+1}, have heteroclinic tangency in the first tongue of their invariant manifolds. The sequence has the property that $0 < \lambda_{n+1} < \lambda_n$; the limit of this sequence is
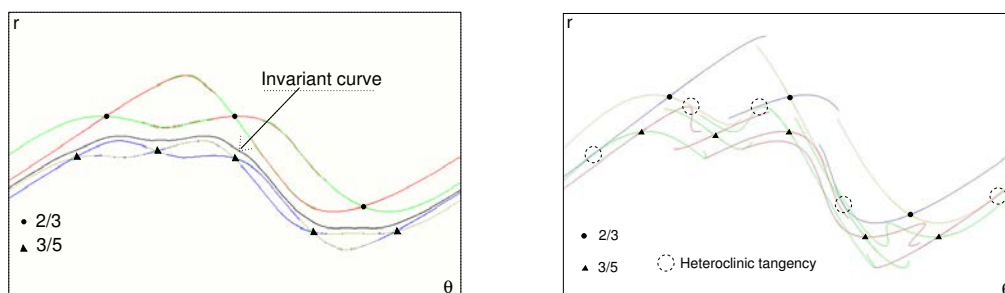
Figure 3. The bold circles represent the periodic orbit with rotation number $\varrho = 2/3$, the bold triangles lie on the orbit with $\varrho = 3/5$. The left figure shows the invariant circle with rotation number ϱ_g, the invariant manifolds of the periodic orbits does not have heteroclinic intersection. The right figure shows the heteroclinic tangencies of the invariant manifolds, the tangencies are shown in the dashed circles. In that case, the invariant circle has disappeared.

then λ_T. The obstruction criterion claims that λ_T is a lower bound for the non-existence of invariant circles with rotation number ϱ_g; this limit is optimal. The value λ_T agrees with the value of the parameter that we can obtain with Greene's method. We have to remark that the obstruction criterion is formal because for each λ_n we have a topological obstruction for the continuity of the invariant circle.

The rate of convergence of the sequence $\{\lambda_n\}_{n=0}^{\infty}$ has the following property,

$$\delta_n = \frac{\lambda_{n-1} - \lambda_{n-2}}{\lambda_n - \lambda_{n-1}} \ . \tag{21}$$

The limit of δ_n is $\delta \sim 1/\varrho_g$, and the rate of convergence is (Olvera & Simó 1987 and De la Llave & Olvera 2006),

$$\lambda_n = \lambda_T + C \, \delta^{-n} \ . \tag{22}$$

The obstruction criterion is also equivalent to other formal topological critera of non-existence of invariant circles, for example, the method of the existence of non-Birkhoff orbits given by Boyland and Hall (Olvera & Simó 1993).

We can modify the obstruction criterion in order to have better numerical accuracy of the computation of λ_T. Instead of computing the heteroclinic tangency of the approximants of the invariant circle, we can compute the area defined by the first lobe of the heteroclinic intersection of the invariant manifolds. Let A_n be the area of the lobe which is given by the intersection of $\mathcal{W}_n^s$ and $\mathcal{W}_{n-1}^u$, as seen in figure 4. We can fix the initial value A_1 and also define the ratio of $A_n/A_{n-1} = \alpha\beta$, where α and β are the horizontal and vertical scaling of the renormalization procedure given by MacKay's Commuting Pair. In that way, we can define, in a unique way, the value of the parameter, $\bar{\lambda}_n$, such that the lobe's area, of the intersection of $\mathcal{W}_n^s$ and $\mathcal{W}_{n-1}^u$ reaches the value A_n. The sequence $\{\bar{\lambda}_n\}$ converges to $\bar{\lambda}_A$ and the rate of convergence follows equation (21). The limit $\bar{\lambda}_A$ coincides with the other limit λ_T obtained with the computation of the tangencies of the invariant manifolds.
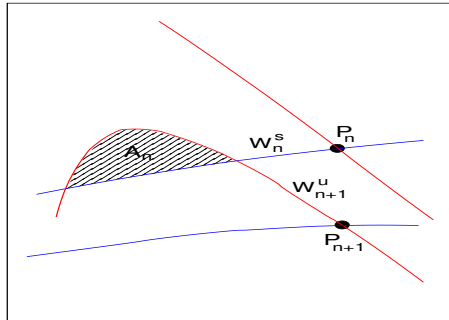
Figure 4.   In this graph, P_n and P_{n+1} are hyperbolic periodic points, where $\mathcal{W}_n^s$ and $\mathcal{W}_{n+1}^u$ are their corresponding invariant manifolds. The first tongues of manifolds $\mathcal{W}_n^s$ and $\mathcal{W}_{n-1}^u$ have a heteroclinic intersection. The region enclosed by the manifolds defines a lobe, the area of the lobe is A_n.

6.   Renormalization group and the obstruction criterion

In the last sections, we have revisited the renormalization group. The global stability of twist maps which have a golden invariant circle can be understood from the way that the renormalization group transforms the map. If the golden circle exists, then the renormalization group sends the map to the trivial point, which corresponds to the integrable case. When the golden circle is in its breakingdown point, the renormalization group sends the map to the critical fixed point of the renormalization group. When the golden circle has broken, the renormalization group sends the map to the upper part of the invariant stable manifold of the critical point. It is not clear what kind of dynamics we can find in the neighborhood of the cantorus set. The lack of a formal description of the dynamics close to the critical point of the renormalization group, does not permit to understand the behavior of the orbits which are close to the cantorus with golden rotation number.

In 2006, de la Llave and Olvera showed that the renormalization group and the obstruction criterion can work together (De la Llave & Olvera 2006). In this way, the dynamics close to the cantorus is given by the web of the invariant manifolds of the periodic orbits which have their heteroclinic intersections in the neigborhood of the cantorus. If we consider the following conjectures as true:

1. In the space of twist maps which have an invariant golden circle (or cantorus), the renormalization group has a critical point, which has an invariant stable manifold of codimension one, and an unstable manifold of dimension one. One of the branches of the stable manifold connects the critical fixed point with the trivial fixed point of the renormalization group.

2. If the map F_λ has a periodic orbit with rotation number $\mathcal{F}_{n-1}/\mathcal{F}_n$ and the spectrum of this orbit is $DF_\lambda^{\mathcal{F}_n}$, then the renormalized map $\mathcal{N}(F_\lambda)$ has an orbit $\mathcal{F}_{n-2}/\mathcal{F}_{n-1}$ and the spectrum of $DF_\lambda^{\mathcal{F}_n}$ is the same as the spectrum of $D\mathcal{N}(F_\lambda)^{\mathcal{F}_{n-1}}$.

3. The map F_λ has a golden circle if and only if the map $\mathcal{N}(F_\lambda)$ does.

4. If the map F_λ has two hyperbolic periodic orbits with rotation numbers $\mathcal{F}_{n-1}/\mathcal{F}_n$ and $\mathcal{F}_{n-2}/\mathcal{F}_{n-1}$, respectively, whose local stable and unstable manifolds intersect, then the corresponding orbits of the renormalized map have local stable and unstable manifolds that intersect.

5. Maps F, for which $\mathcal{N}^n(F)$ converges to the trivial fixed point have a smooth invariant circle. If $\mathcal{N}^n(F)$ converges to the critical fixed point then there is an invariant critical circle, but presumably not smooth.

Then the following holds:

1. The stable manifold of the critical fixed point would be part of the boundary between the existence of smooth invariant curves and hyperbolic orbits with golden mean rotation number.

2. The obstruction criterion for the non-existence of golden mean invariant circles is sharp. That is, either the map has a golden mean invariant circle or the obstrucion criterion holds.

3. The residue criterion given by Greene would be valid.

4. If there is no invariant circle, there are uniformly hyperbolic sets with golden mean rotation number. This set is called a cantorus and it has hyperbolic stability. The hyperbolicity of the cantorus is inherited by the heteroclinic intersection of the invariant manifold of the hyperbolic periodic orbits which are in the neighborhood of the cantorus.

The obstruction criterion is a formal method to show the non-existence of invariant circles. Furthermore, this criterion is consistent with the renormalization group, therefore we can consider this criterion as a renormalization procedure. In this way, the numerical procedures, which are used in the obstruction method, look like a renormalization process. The computation of heteroclinic tangencies is consistent with the renormalization procedure because the rates of convergence have the scaling factors which they are expected to have in the renormalization theory. On the other hand, the obstruction criterion is also consistent with Greene's method. Therefore, the numerical procedure is very accurate and the value of the critical parameter that is provided by the obstruction method is optimal.

Renormalization group has become a standard tool for describing universal properties of different routes to chaos; these routes are related with the destruction of invariant circles with algebraic rotation numbers ϱ. The continued fraction expansion of these numbers is eventually periodic. The universal scaling exponents of each route are related with the periodic sequence of ϱ. Invariant circles with noble rotation number has a simple combinatoric, that is, the period of the sequence is one. The *Principle of Approximate Combination of Scaling Exponents* (PACSE) organizes the scaling exponents for different transition to chaos (De la Llave, Olvera & Petrov 2011). If the combinatorics of a transition is a composition of two simpler combinatorics, then the scaling exponents of the combined combinatorics is approximately equal to the product of the scaling

exponents, both in the parameter space an in the configuration space, corresponding to each of these two combinatorics. PACSE states quantitatively as precise asymptotics of the scaling exponents for combined combinatorics.

7.   Conclusions

The development of a formal theory that describes the global stability of twist maps in the plane is not yet complete. The renormalization group is a suitable theory to sort out the complex dynamics on the border of chaos. The existence of the critical fixed point has been proved with the help of a computer assistant tools (Arioti & Kock 2010). Renormalization group for Hamiltonian systems with three or higher degree of freedom have been studied by H. Kock. Nevertheless, this theory is only for a Hamiltonian flows (Koch 2008).

In order to improve the understanding of the renormalization group in the neighborhood of the critical fixed point, the obstruction method allows to make a link between numerical procedures and the renormalization method. The idea is to formalize the numerical steps of the obstruction method and then provide an analytical tool which could show the existence of these fixed points of the renormalization group, perhaps with the help of a computer assisted proof.

References

Arioti, G. & Koch, H. 2010, Comm. Math. Phys. 295, 415

Arnold, V. I. 1978, *Mathematical Methods of the Classical Mechanics*, Springer-Verlag, New York.

Aubry, S. & Le Daeron, P. 1983, Physica D, 8, 381

Birkhoff, G. D. 1950, *Colleccted Works, Vol. 1*, AMS, New York.

Boyland, P. L. & Hall, G. R. 1987, Topology 26, 21

Chirikov, B. V. 1979, Phys. Report 52, 262

De la Llave, R. & Olvera, A. 2006, Nonlinearity 19, 1907

De la Llave, R., Olvera, A. & Petrov, N. 2011, J. of Statistical Physics 143, 889

Delshams, A. & De la Llave, R. 2000, SIAM J. Math. Anal. 31- 6, 1235.

Falcolini, C. & De la Llave, R. 1992, J. of Stat. Phys., 67, 609

Greene, J. M. 1979, J. of Math. Phys. 20, 1183

Katok, A. & Hasselblatt, B. 1995, *Introduction to the Modern Theory of Dynamical Systems*, Cambridge University Press.

Koch, H. 2002, Discrete Contin. Dyn. Syst. 8-3, 633

Koch, H. 2008, Ergodic Theory and Dynamical Systems 28, 1559

Leage, I. & MacKay, R. S. 1986, Phys. Letters A 118, 2741

MacKay, R. S. 1983, Phys. D 7, 283

MacKay, R. S. & Percival, I. C. 1985, Comm. Math. Phys. 98, 469

Mather, J. M. 1982, Topology 21, 457

Moser, J. 1973, *A renormalisation approach to invariant circles in area-preserving maps*, Princeton Univ. Press.

Olvera, A. & Simó 1987, Physica D 26, 181.

Olvera, A. & Simó, C. 1993, Int. J. Bifurcation and Chaos 3, 165

Reichl, L. E. 1992, *The transition to chaos in conservative classical systems: Quantum manifestation*, Institute of Nonlinear Science, Springer Verlag.

Shenker, S. J. & Kadanoff, L. P. 1982, J. Statist. Phys. 27-4, 631

Stirnemann, A. 1993, Commun. Math. Phys. 152, 369

306

ORAL COMMUNICATION

Dynamical origin of V-type asteroids outside the Vesta family

F. Roig[1], H. Folonier[2], C. Beaugé[3], A. O. Ribeiro[1]

(1) Observatório Nacional, Rio de Janeiro, Brazil
(2) Instituto de Astronomia, Geofísica e Ciências Atmosféricas, São Paulo, Brazil
(3) Observatorio Astronómico de Córdoba, UNC, Argentina

Abstract.   We review some recent results on the long term dynamical evolution of V-type asteroids that point to their origin as fugitives from the Vesta family. Three scenarios are explored: (i) interplay of weak mean motion and non linear secular resonances in the inner Belt with the Yarkovsky effect, (ii) crossing of the 3:1 mean motion resonance with Jupiter, and (iii) evolution by planetary encounters and resonance stickiness. These mechanisms may explain a large fraction of the V-type asteroids that are observed outside the Vesta family, but there are some particular cases that would need other explanations.

1.   Introduction

Asteroid (4) Vesta, located at about 2.35 AU in the inner Main Belt of asteroids, is the only one among the largest minor bodies in the Solar System (diameter $D \sim 500$ km) to show a basaltic crust. This crust is considered to be the consequence of a process of heating and melting of the asteroid's interior that led to a mineralogical differentiation forming a core of heavy minerals (iron) and a mantle of lighter minerals (olivine and pyroxene). Vesta has also associated one of the largest asteroid families in the Main Belt: the Vesta family. This family is formed by the small ($D < 5$ km) fragments ejected during a catastrophic collision that excavated a large crater on Vesta's surface. Therefore, it is expected that the family members show a basaltic composition too (Mothé-Diniz *et al.* 2005).

In the visible wavelength interval (0.5 - 1.0 μm), the reflectance spectrum of a basaltic asteroid is characterized by a steep slope downwards 0.85 μm and a deep absorption band upwards 0.85 μm with a minimum centered at about 1 μm. Asteroids showing this kind of spectrum are classified as V-type in the usual taxonomic systems (e.g. Lazzaro *et al.* 2004). It is worth mentioning that even if all basaltic asteroids are V-type, not all the V-type asteroids will be basaltic because the actual mineralogy of the asteroid's surface can be characterized only by looking at the whole visible plus near infrared reflectance spectrum (0.5 - 2.5 μm; e.g. de Sanctis *et al.* 2011).

Besides the V-type asteroids belonging to the Vesta family (usually referred to as "vestoids"), there is a significant amount of V-type asteroids that are not members of the Vesta family. Most of them are found in the inner Main Belt

but so far away from the family edges that cannot be explained by the typical ejection velocities produced during the cratering event that formed the family (Roig & Gil-Hutton 2006; Carruba *et al.* 2007b; Duffard 2009). There are also some recently discovered V-type asteroids beyond 2.5 AU in the middle and outer Main Belt (Moskovitz *et al.* 2008; Duffard & Roig 2009), as well as a considerable number among the population of Near Earth Asteroids (NEAs).

There are different possibilities to explain the existence of these bodies:

- Even being classified as V-type according to their visible reflectance spectra, they are not actually basaltic.

- They came from a yet unknown reservoir of basaltic asteroids other than the Vesta family, or they formed from the catastrophic disruption of another differentiated body whose existence can only be hypothesized.

- They formed as such by pure accretion of basalt (Lazzaro 2009).

- They are former members of the Vesta family that reached their present orbits by dynamical evolution over long timescales.

In this paper, we review some recent results that address this later possibility as the most plausible scenario for the origin of V-type asteroids outside the Vesta family. These results show that the long term dynamical evolution driven by the web of resonances in the Main Belt, combined with non conservative effects produced by the thermal re-emission of the asteroid's surface (Yarkovsky effect) and with close encounters with planets and massive asteroids may explain a significant fraction of V-type asteroids as being fugitives from the Vesta family.

2. V-types in the inner Main Belt

In Fig. 1, left panels, the Vesta family is represented by the small dots while the triangles represent the known V-type asteroids outside the family in the inner Main Belt. It is worth mentioning that these asteroids have been classified as V-type from their reflectance spectra only. A much larger sample can be obtained, especially for $a > 2.5$ AU, if adopting a classification scheme based on reflectance colors from large photometric catalogs (e.g. Carvano *et al.* 2010).

In Nesvorný *et al.* (2008), we performed a numerical simulation of the long term evolution of the Vesta family members. We started with a configuration of 6600 test particles with their initial orbits cloned from the 600 largest family members and let them evolve over 2 Gy using a symplectic N-body integrator. The simulation took into account the full gravitational perturbations by all planets from Venus to Neptune, as well as the systematic drift in semi-major axis caused by the so-called Yarkovsky effect (Bottke *et al.* 2002). This drift depends on the thermal properties of the asteroid surface, and for each clone it was estimated by the formula:

$$\dot{a}_Y = 2.5 \times 10^{-4} \frac{1}{D} \cos \varepsilon \qquad \text{AU/My} \qquad (1)$$

where the diameter D (in km) was chosen within the typical range of vestoids' sizes and the spin obliquity ε was chosen at random. Figure 1 shows the initial

(left panels) and final (right panels) states of the simulation. It is worth noting that each dot in the left panels of Fig. 1 actually corresponds to the initial orbit of 11 test particles in the simulation: 1 real asteroid plus 10 clones that had assigned 10 different values of the Yarkovsky drift[1]. The different drift rates may be attributed to variations of either the size or the spin obliquity.
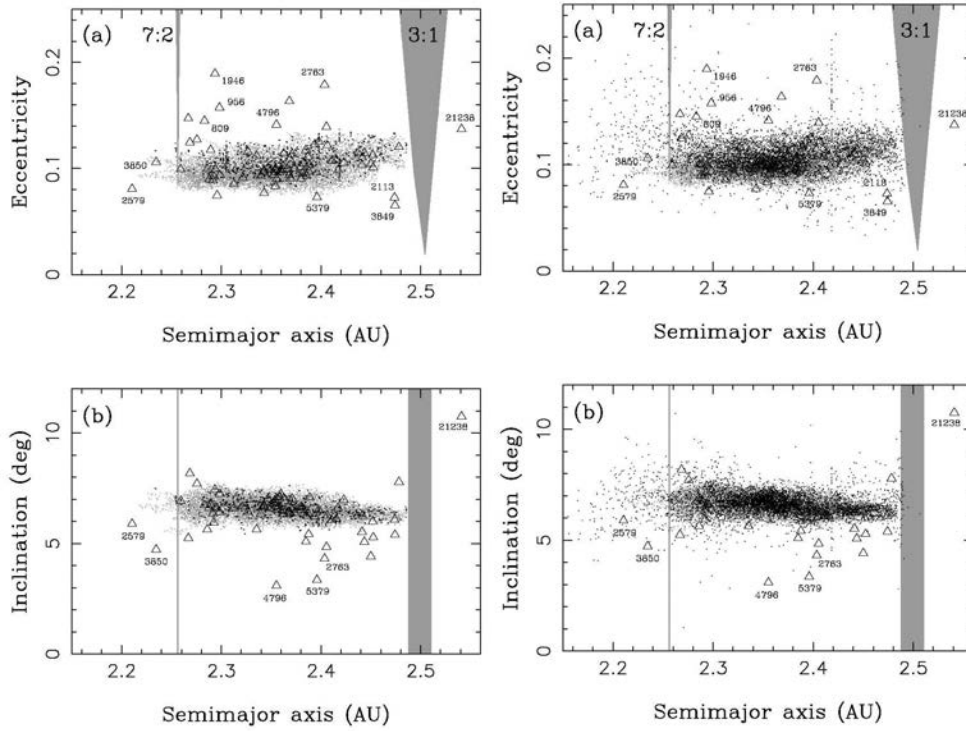

Figure 1. Initial (left) and final (right) states of the test particles in the Nesvorný *et al.* (2008) simulation (small dots). The initial state corresponds to the current distribution of the 600 largest family members. The open triangles give the present location of the know V-type asteroids that are not members of the Vesta family. The gray areas represent the approximate boundaries of the main Jovian mean motion resonances in the region.

After 2 Gy (right panels) we observe a significant spread of the original family with many test orbits that fell well beyond the outskirts of the family. These orbits mostly evolved to their final positions through interaction of the Yarkovsky effect with several mean motion and non linear secular resonances found in the region. Particularly notorious is the vertical stream of test particles at ~ 2.42 AU which is related to the 1:2 mean motion resonance with Mars (Gallardo *et al.* 2011).

Our simulation produced a large number Vesta family fugitives with smaller semi-major axes ($2.2 < a < 2.4$ AU) and higher eccentricities ($0.05 < e < 0.2$) than the average family members. This result is compatible with many of the

---

[1]This is why the distribution of initial states is less dense than the distribution of final states.

observed V-type asteroids outside the Vesta family. In some cases, like asteroids (809) Lundia and (956) Elisa, the current orbits can be accurately reproduced by slow diffusion along non linear secular resonance over time spans larger than 1.2 Gy (Carruba *et al.* 2005), which puts a lower limit to the age of the Vesta family.

On the other hand, our simulation was unable to produce an amount of fugitives compatible with the V-type asteroids having higher semi-major axes ($2.35 < a < 2.5$) and smaller inclinations ($2^{\mathrm{o}} < I < 6^{\mathrm{o}}$) than the average family members. It is worth noting that this low inclination group includes some of the largest ($D \sim 5$ km) V-types in the inner Belt which are expected to be less affected by the Yarkovsky effect. If these bodies are to be former members of the Vesta family, it is necessary to find other mechanisms to explain their current orbits.

An alternative explored by Carruba *et al.* (2007a) involves the effect of close encounters between small Vesta family members and (4) Vesta itself. Since Vesta is one of the most massive asteroids in the Main Belt, the accumulation over long timescales of the gravitational pulls suffered during successive close approaches to Vesta may cause significant variations in the orbit of a small vestoid. Although this mechanism proved to be size independent (contrary to the Yarkovsky effect), the resulting orbital changes are not too big and only the V-type asteroids just outside the edges of the Vesta family were well reproduced by Carruba *et al.* (2007a) simulations. Therefore, the presence of large V-type asteroids with low inclinations in the inner Belt remains an open problem.

3. V-types in the middle Main Belt

3.1. Captures and crossings in the 3:1 resonance

The simulation by Nesvorný *et al.* (2008) did not produce fugitives beyond 2.5 AU because the authors only simulated large objects and therefore few test particles were able to reach the border of the 3:1 mean motion resonance with Jupiter over the simulation time span. In Roig *et al.* (2008), we performed a more detailed simulation of the evolution of Vesta family members that interact with this resonance. We integrated the orbits of 2000 test particles, cloned from real Vesta family members initially very close to the 3:1 resonance, over 150 My. As in Nesvorný *et al.* (2008), the simulation includes planetary perturbations and the Yarkovsky effect tuned to produce a positive drift in semi-major axis, thus forcing the test particles to reach the resonance border.

As expected, most of test particles in the simulation were captured into the 3:1 resonance and subsequently followed a fast chaotic evolution that drove them into close encounters with the terrestrial planets, being eventually removed from the Main Belt. On the other hand, we also found a small fraction of test particles that were able to cross the 3:1 resonance, reaching stable orbits in the middle Main Belt. An example of this evolution is shown in Fig. 2 for two different rates of the Yarkovsky drift. Our simulations indicated that the probability of these crossings is actually very small and roughly scales with the Yarkovsky drift as:

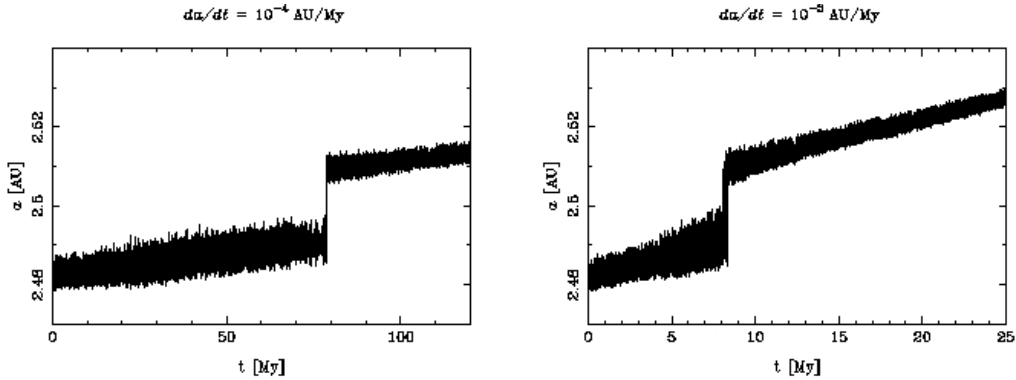$$p_{cross} = 0.003 \frac{1}{D} \cos \varepsilon \qquad (2)$$

Figure 2.   Example of two test particles in the Roig *et al.* (2008) simulation that became temporarily captured in the 3:1 resonance and were able to cross to the middle Belt. Two different values of the Yarkovsky drift were assumed roughly corresponding to 0.1-1.0 km sized bodies.

Aiming to better understand this crossing mechanism, we performed a systematic study of the capture probability into the 3:1 resonance by means of a pseudo symplectic mapping approach (Folonier 2011; Folonier *et al.* 2012, in preparation). The mapping is based in the construction of a generating function B for an implicit canonical transformation representing a perturbed twist mapping:

$$B(\boldsymbol{\theta}_n, \boldsymbol{J}_{n+1}) = \boldsymbol{\theta}_n \boldsymbol{.} \boldsymbol{J}_{n+1} + \tau F(\boldsymbol{\theta}_n, \boldsymbol{J}_{n+1})$$

where

$$F(\boldsymbol{\theta}, \boldsymbol{J}) = F_0(\boldsymbol{J}) + F_1(\boldsymbol{\theta}, \boldsymbol{J})$$

is the Hamiltonian of the non planar, elliptic, restricted three body problem around the 3:1 resonance averaged over the short (synodic) periods, $(\boldsymbol{\theta}_n, \boldsymbol{J}_n)$ are the action angle variables of the unperturbed (F_0) problem at the n-th step of the mapping, and τ is the mapping time step. The mapping equations are given by:

$$
\begin{aligned}
\boldsymbol{J}_{n+1} &= \boldsymbol{J}_n - \tau \frac{\partial F}{\partial \boldsymbol{\theta}_n} + \tau \boldsymbol{J}_{n+1} \frac{\dot{a}_Y}{2a} \\
\boldsymbol{\theta}_{n+1} &= \boldsymbol{\theta}_n + \tau \frac{\partial F}{\partial \boldsymbol{J}_{n+1}}
\end{aligned}
\tag{3}
$$

where the Yarkovsky drift in semi-major axis, $\dot{a}_Y$, was included to cause an additional perturbation in the actions (since a depends only on the actions). It is worth noting that this mapping is not strictly symplectic due to the Yarkovsky perturbation; that is why we refer to it as pseudo symplectic. Besides the Yarkovsky effect, we also included in the mapping the time variations of Jupiter's eccentricity and inclination due to the secular and long term planetary perturbations (Roig & Ferraz-Mello 1999). These were introduced applying the usual

formulas of the secular linear planetary theory:

$$e_{\mathrm{J}} \exp i\varpi_{\mathrm{J}} = \sum_k G_k \exp(g_k t + \gamma_k)$$

$$\sin \frac{I_{\mathrm{J}}}{2} \exp i\Omega_{\mathrm{J}} = \sum_k S_k \exp(s_k t + \sigma_k) \tag{4}$$

The main advantage of using this mapping approach model is that the different perturbations can be arbitrarily turned on and off, thus allowing to address the actual role of each perturbation in the evolution of the test particles. The mapping also allows to perform a large amount of simulations using different values of the model parameters with a low computational cost.

The initial conditions for our simulations were chosen to be non resonant orbits close to the 3:1 resonance border, with $a < 2.5$ AU, $0 < e < 0.4$, $I = 6^{\underline{o}}$, and $\dot{a}_Y > 0$. Each orbit was cloned several times choosing random values of initial angles in the interval $0^{\underline{o}} < \boldsymbol{\theta} < 360^{\underline{o}}$. The simulations were carried out until all the initial conditions were either captured into the resonance or crossed it. Our results indicated that the capture probability p_{cap} of each orbit is a function of $\dot{a}_Y, e, \boldsymbol{\theta}$, and for subsequent analysis we computed the mean probability, $\bar{p}_{cap}(\dot{a}_Y, e,)$ by averaging over all the possible initial angles.

The simulations showed that orbits with the same initial eccentricity may cross the resonance or become captured depending on the rate of the Yarkovsky drift. At the fastest drifts ($\dot{a}_Y > 1$ AU/My) the captures occur at discrete windows of eccentricity alternating with windows of crossings, and $\bar{p}_{cap}$ shows an oscillating profile with e. It is worth noting that such fast drifts roughly correspond to cm-size bodies and therefore they are unrealistic for real asteroids. On the other hand, at lower drifts ($\dot{a}_Y < 1$ AU/My) the structure of discrete windows is no longer evident and $\bar{p}_{cap}$ resembles the classical analytic result obtained in the adiabatic limit, i.e. 100% of captures at low eccentricities and an exponentially decreasing probability at larger eccentricities (e.g. Henrard 1982; Gomes 1995). The above results are illustrated in Fig. 3, that shows the final semi-major axes (left panels) of the orbits and their mean capture probabilities (right panels) against their initial eccentricities, for two different values of $\dot{a}_Y$.

Having determined $\bar{p}_{cap}$ as a function of e, the average capture probability for the Vesta family members was obtained by integrating over the eccentricity limits of the family:

$$\langle \bar{p}_{cap} \rangle_{\mathrm{Vesta}} = \frac{1}{0.07} \int_{0.07}^{0.14} \bar{p}_{cap}(\dot{a}_Y, e) \, de \tag{5}$$

Table 1 shows this average probability obtained with a mapping model that includes the secular variations of Jupiter's orbit (Eqs. 4), assuming different values of the Yarkovsky drift. We may note that these results are compatible with the crossing probability given by Eq. 2. Actually, we verified that the addition of the secular perturbations is crucial to reproduce Eq. 2, since the solely elliptic model gave much smaller capture probabilities.

By combining the crossing probability p_{cross} with the estimated flux of vestoids that can reach the border of the 3:1 resonance over the age of the Vesta family, we conclude that only the smallest V-types ($D < 2$ km) have a chance
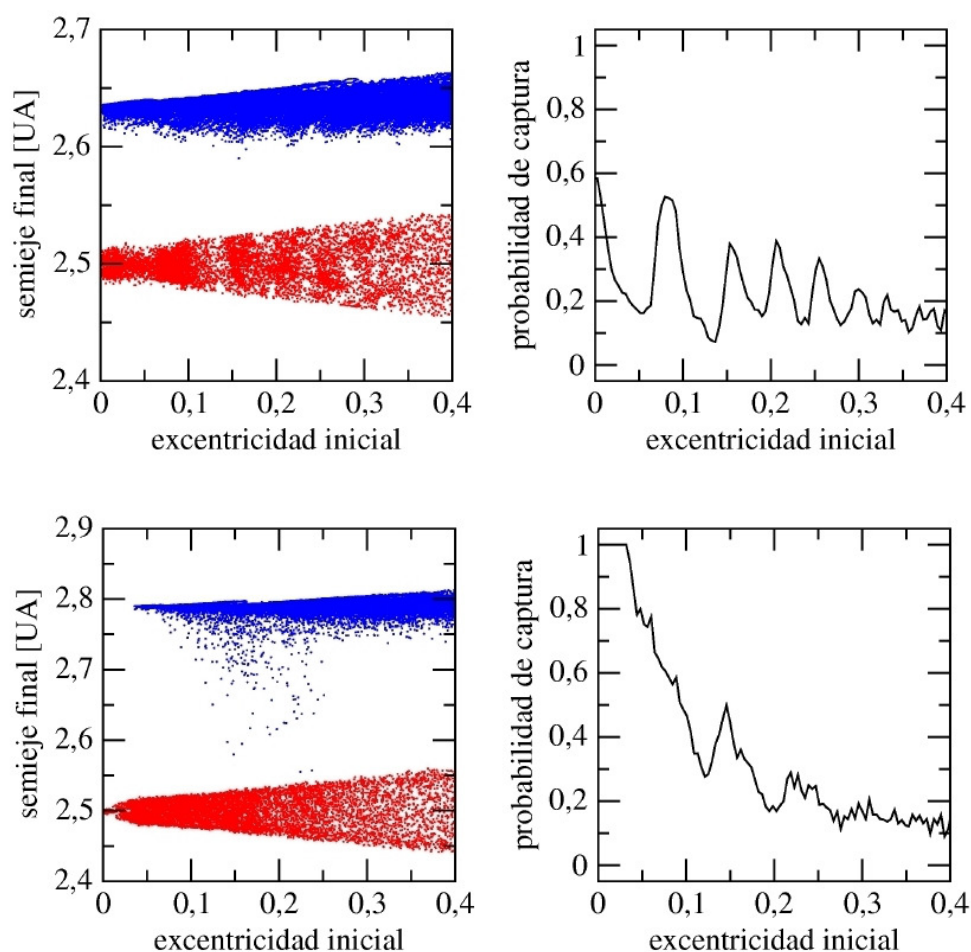
Figure 3. Final semi-major axes (left) and mean capture probabilities (right) as function of the initial eccentricities of test orbits simulated with our mapping model. The top panels correspond to $\dot{a}_Y = 10$ AU/My and the bottom panels to $\dot{a}_Y = 0.1$ AU/My. Red dots are orbits captured into the resonance and blue ones are those that crossed the resonance. The simulations correspond to the elliptic model, with Jupiter in a fixed orbit. The differences between the non adiabatic (top) and the adiabatic (bottom) cases are evident.

of being transported from the inner to the middle Belt over Gy timescales. Although this result cannot explain the largest V-type asteroid found in the middle Belt, that has a diameter of 5 km, recent studies (de Sanctis *et al.* 2011; Hammergren *et al.* 2011) indicate the existence of several smaller basaltic asteroids beyond 2.5 AU whose origin might be compatible with the resonance crossing scenario.

Table 1.   Average capture probabilities of Vesta family members (Eq. 5) in the secular elliptic mapping model for some Yarkovsky drifts. These values can be compared to Eqs. 1 and 2 taking into account that $p_{cross} = 1 - \langle \bar{p}_{cap} \rangle$.

$\dot{a}_Y$ [AU/My]	$\langle \bar{p}_{cap} \rangle_{\text{Vesta}}$
10	0.28
1	0.45
0.1	0.62
0.01	0.67
0.005	0.85
0.0025	0.92
0.001	0.97

3.2.   Planetary encounters and resonance stickiness

A second possibility to explain the transport of V-type asteroids from the inner to the middle Belt involves the temporary capture of former vestoids into a planet crossing regime with Mars (and possibly also with the Earth), and the further detaching from the planet crossing regime due to resonance stickiness. In Ribeiro (2010) and Ribeiro & Roig (2011), we explored this possibility through the numerical integration of Mars crossing orbits under the influence of the Yarkovsky effect. Our working hypothesis was that a fraction of about 10% of the fugitives from the Vesta family actually reaches the Mars crossing region (Nesvorný *et al.* 2008) and suffer strong changes of the semi-major axis due to close encounters with Mars. Some of these bodies temporarily acquire values of $a > 2.5$ AU and at this stage they might be captured in a mean motion resonance by becoming stuck to the resonant separatrix.

We simulated the evolution of test particles with initial conditions corresponding to both real and fictitious Mars crossing orbits over 120 My (which is the typical lifetime of the Mars crossers). The model included planetary perturbations from Venus to Neptune, as well as the Yarkovsky effect, and we used a symplectic N-body integrator that manages close encounters between the test particles and the planets. The evolution of each test particle was monitored until it eventually reached a semi-major axis > 2.5 AU. We assumed that a particle was transported to the middle Belt if it continuously remained with $a > 2.5$ AU for a time interval $\Delta T \geq 10^5$ years. In order to qualitatively analyze the behavior of the transported particles, we built maps of residence times. Figure 4 shows a typical map obtained for a set of fictitious Mars crossing test particles with initial perihelion distances in the range $1.33 < q < 1.44$ AU. Captures in mean motion resonances due to resonance stickiness appear in the map as stalactites and are related to the 5:2, the 7:3 and, to a less extent, the 8:3 resonances with Jupiter. Nevertheless, all these captures were temporary and none of the simulations in Ribeiro (2010) were able to produce a long lasting orbit in the middle Belt. Therefore, we concluded that this mechanism would not be efficient enough to explain the V-type asteroids in the middle Belt as fugitives from the Vesta family.
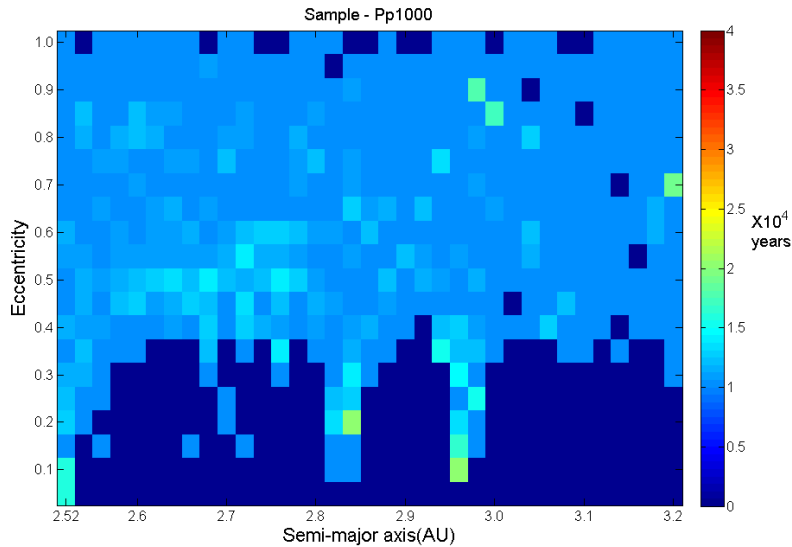
Figure 4.   Map of residence times for a set of fictitious Mars crossing orbits. The map was computed by subdividing the space e vs. a beyond 2.5 AU in 20×70 cells, and by determining the corresponding time interval that each orbit continuously remained in each cell during the simulation. Then, for each cell we computed the average time interval over all the orbits, and this gives the residence time of any particle in that cell.

4.   Conclusions

In this article, we have reviewed some recent work that intend to address the origin of basaltic asteroids outside the Vesta family by several mechanisms of chaotic transport in the inner and middle Asteroid Main Belt. These mechanisms involve the interplay of resonances and planetary encounters with the Yarkovsky effect. Our conclusions can be summarized as follows:

- Most non vestoid V-type asteroids in the inner Belt can be explained as fugitives from the Vesta family that reached their present orbits over Gy timescales of chaotic evolution. The exception is a population of large ($D > 4$ km) V-type asteroids located at lower orbital inclinations than the Vesta family members, whose present orbits cannot be explained by long term chaotic diffusion.

- Small ($D < 2$ km) members of the Vesta family are able to cross the 3:1 resonance with Jupiter, pushed by the Yarkovsky drift in semi-major axis, and may arrive to stable orbits in the middle Belt. However, this mechanism is not efficient to explain the presence of larger V-types beyond 2.5 AU.

- Close encounter with terrestrial planets may temporarily drive V-type asteroids from the inner to the middle and outer Belt that become captured in mean motion resonance by stickiness effect. Nevertheless, this mech-

anism does not produce long lasting orbits and would not explain the presence of V-types with $a > 2.5$ AU.

It is worth noting that the explanation of V-types asteroids outside the Vesta family is not only limited by dynamical constraints, as shown here, but also by the mineralogical compatibility with Vesta's crust. There is at least one V-type asteroid, (1459) Magnya, located in the outer Belt (Lazzaro *et al.* 2000), that has proven to be incompatible with Vesta's mineralogy. According to Michtchenko *et al.* (2002), this asteroid could have been originated during the disruption of another differentiated asteroid and the corresponding family would have been dynamically dispersed over long timescales. A similar approach was also proposed by Carruba *et al.* (2007b) to explain the large V-type asteroid (21238) 1995 WV7 in the middle Belt. Therefore, in the end, the hypothesis of a source of basaltic material in the Main Belt other than Vesta still remains an open question.

References

Bottke, W. F. Jr., Vokrouhlický, D., Rubincam, D. P., & Broz, M. 2002. In: Asteroids III, 395

Carruba, V., Michtchenko, T. A. & Lazzaro, D. 2007a, A&A 473, 967

Carruba, V., Michtchenko, T., Roig, F., Ferraz-Mello, S., & Nesvorný, D. 2005, A&A 441, 819

Carruba, V. Roig, F., Michtchenko, T. A., Ferraz-Mello, S., & Nesvorný, D V. 2007b, A&A 465, 315

Carvano, J. M., Hasselmann, P. H., Lazzaro, D., & Mothé-Diniz, T. 2010, A&A 510, A43.

de Sanctis, M. C., Migliorini, A., Luzia Jasmin, F., Lazzaro, D., Filacchione, G., Marchi, S., Ammannito, E., & Capria, M. T. 2011, A&A 533, A77

Duffard, R. 2009, EM&P 105, 221

Duffard, R. & Roig, F. 2009, P&SS 57, 229

Folonier, H. 2011, Undergraduate's Thesis, Universidad Nacional de Córdoba, Argentina.

Gallardo, T., Venturini, J., Roig, F., & Gil-Hutton, R., 2011, Icarus 214, 632

Gomes, R. S. 1995, CeMDA 61, 97

Hammergren, M., Gyuk, G., Solontoi, M., & Pucket, A. 2011. In: EPSC-DPS Joint Meeting 2011, 668

Henrard, J. 1982. Celest. Mech. 27, 3

Lazzaro, D. 2009, RMxAA Conf. Ser. 35, 1

Lazzaro, D. Michtchenko, T., Carvano, J. M., Binzel, R. P., Bus, S. J., Burbine, T. H., Mothé-Diniz, T., Florczak, M., Angeli, C. A., & Harris, Alan W. 2000, Sci. 288, 2033

Lazzaro, D., Angeli, C. A., Carvano, J. M., Mothé-Diniz, T., Duffard, R., & Florczak, M. 2004, Icarus 172, 179

Michtchenko, T. A., Lazzaro, D., Ferraz-Mello, S., & Roig, F. 2002, Icarus 158, 343

Moskovitz, N. A., Burbine, T. H.. Mayne, R., Bus, S. J., Willman, M., Jedicke, R., & Gaidos, E. 2008, Icarus 198, 77

Mothé-Diniz, T., Roig, F. & Carvano, J. M. 2005, Icarus 174, 54

Nesvorný, D., Roig, F., Gladman, B., Lazzaro, D., Carruba, V., Mothé-Diniz, T., & 2008, Icarus 193, 85

Ribeiro, A. O. 2010, Master's Thesis, Observatório Nacional, Brazil.

Ribeiro, A. O. & Roig, F. 2011, J. Phys. Conf. Ser. 285, 012024.

Roig, F. & Ferraz-Mello, S. 1999, P&SS 47, 653

Roig, F. & Gil-Hutton, R. 2006, Icarus 183, 411

Roig, F., Nesvorný, D., Gil-Hutton, R.,& Lazzaro, D. 2008, Icarus 194, 125

ORAL COMMUNICATION

Diffusion measurements in a 3DoF Hamiltonian flow.

M. F. Mestre [1,2], P.M. Cincotta [1,2], C.M. Giordano[1,2]

(1) Instituto de Astrofísica de La Plata (CONICET)
(2) Grupo de Caos en Sistemas Hamiltonianos
Facultad de Ciencias Astronómicas y Geofísicas - UNLP
Paseo del Bosque, B1900FWA La Plata, Argentina

Abstract. We present measurements of diffusion phenomenae in the phase space of a three degree of freedom *quasi*–integrable Hamiltonian flow. By tracking the evolution of the variance of ensembles of test particles, we characterize the diffusion as basically anomalous. We find a time range in which it can be considered as a subdiffusion process and other in which it can be considered as a normal diffusion process. In the former case we fit Hurst exponents and in the latter case we compute diffusion coefficients based on the rate of growth of the variances.

1.   Introduction

In this contribution we will study numerically a diffusion process that takes place in the stochastic layer of a non–linear resonance of a particular *quasi*–integrable three degree of freedom (3DoF) Hamiltonian flow. We will focus in a resonant region of multiplicity one; i.e. isolated resonance region.

The experiments are based on evolving ensembles of test particles and measuring some statistical quantities. For the values of total energy and resonant action that will be considered here, we will have examples of subdiffusion in both *overlapping* and *non–overlapping* local regimes, depending on the value of the perturbation parameter. We will also identify time ranges in which there is a diffusion behaviour approximately normal that will allow us to compute standard diffusion coefficients.

The structure of the article is as follows.

In subsection 1.1. we present a one degree of freedom (1DoF) quartic oscillator and give its solution and its action–angle variables. In subsection 1.2. we present the 3DoF *quasi*–integrable Hamiltonian flow whose diffusion properties we will study throughout this work. The integrable part of this Hamiltonian consists in three independent 1DoF quartic oscillators. In subsection 1.3. we specify a linear canonical transformation that allows to measure diffusion in a new set of *quasi*–integrals which are geometrically easier to analyse. In subsection 1.4. we define some statistical quantities based on ensemble averages. In subsection 1.5. we explain briefly the numerical integrators used herein.

In section 2. we present the geometric situation of the chosen resonant action, the guiding resonance and its neighbourhood. In section 3. we describe some numerical experiments done with ensembles of test particles, analysing the

functional form of the time evolution of some statistical quantities. We compute Hurst exponents and diffusion coefficients for different values of the perturbation parameter. Finally, in section 4. we give a summary of the obtained results.

1.1.   The quartic oscillator

The *quartic oscillator* is given by the Hamiltonian

$$\tilde{H}(v, x) = \frac{v^2}{2} + \frac{x^4}{4}, \tag{1}$$

where v and x are real variables. The associated system of differential equations is given by:

$$\frac{dv}{dt} = -x^3, \qquad \frac{dx}{dt} = v. \tag{2}$$

Let h be the total energy and let a be the oscillation amplitude. Then,

$$h = \tilde{H}(0, a) = \frac{a^4}{4}$$

and Eqs. (2) can be rewritten as

$$\left(\frac{dx}{dt}\right)^2 = \frac{1}{2}(a^4 - x^4),$$

whose solution can be expressed in terms of the *Jacobi elliptic cosine* (cn) *of modulus* $k = 1/\sqrt{2}$:

$$x(t) = a \, \text{cn}(at, k) = a \, \text{cn}(at, \, 1/\sqrt{2}).$$

Using the Fourier series development of the Jacobi elliptic cosine given by Gradshtyn & Ryzhik (1980), we have that (Chirikov 1979):

$$x(t) = a\frac{\sqrt{2}\pi}{K_0} \sum_{n=1}^{\infty} \frac{1}{\cosh\left((n - 1/2)\pi\right)} \cos\left((2n - 1)\frac{\pi a t}{2K_0}\right),$$

where $K_0 \equiv K(1/\sqrt{2})$ denotes the *complete elliptic integral of the first kind.*
Introducing the following constants:

$$\beta \equiv \frac{\pi}{2K_0} \approx 0.847213084793979, \qquad \alpha_n \equiv \frac{1}{\cosh\left((n - 1/2)\pi\right)} \quad \text{and} \quad \omega \equiv \beta a, \tag{3}$$

we have that

$$x(t) = 2^{3/2}\omega \sum_{n=1}^{\infty} \alpha_n \cos\left((2n - 1)\omega t\right).$$

This equation implies that the quantity ω is the fundamental frequency of the motion.

The α_n coefficients satisfy:

$$\frac{\alpha_{n+1}}{\alpha_n} \approx \frac{1}{23} \quad \text{for n} \geq 2 \quad \text{and} \quad \alpha_1 \approx 0.4.$$

Taking into account the previous relation between a and h and Eq. (3), we obtain the dependence between the latter and ω:

$$\omega = \sqrt{2}\beta h^{1/4}. \tag{4}$$

Due to the fact that the Hamiltonian equations in action–angle variables (I, θ) imply that $\omega(I) = \frac{\partial H(I)}{\partial I}$, it is possible to integrate Eq. (4), obtaining the relation between the action and the energy of the system.

$$h = AI^{4/3}, \quad \text{or equivalently} \quad I = \left(\frac{h}{A}\right)^{3/4},$$

where $A \equiv (3\beta/2\sqrt{2})^{4/3} \approx 0.867145326484821$.

Let us mention that the dependence of the frequency on the action is given by:

$$\omega(I) = \frac{4}{3}AI^{1/3} \tag{5}$$

The cartesian coordinate can be expressed in terms of the action–angle variables of the system, in the fashion:

$$x(I, \theta) = (3\beta I)^{1/3}\text{cn}\left(\frac{\theta}{\beta}, \frac{1}{\sqrt{2}}\right)$$

$$v(I, \theta) = \varrho\sqrt{2\left(AI^{4/3} - \frac{1}{4}[x(I, \theta)]^4\right)}, \tag{6}$$

where ϱ stands for the sign of v and its dependence on the angle is given by:

$$\varrho \equiv \begin{cases} 1 & \text{if } 0 \leq \theta < \pi, \\ -1 & \text{if } \pi \leq \theta < 2\pi. \end{cases} \tag{7}$$

Besides, the inverse transformation[1] of Eqs. (6)(7) is given by:

$$I(v, x) = \left[\frac{1}{A}\left(\frac{1}{2}v^2 + \frac{1}{4}x^4\right)\right]^{3/4}, \tag{8}$$

$$\theta(v, x) = \begin{cases} \beta \text{ cn}^{-1}\left(\frac{x}{[3\beta\ I(v,x)]^{1/3}}, \frac{1}{\sqrt{2}}\right) & \text{if } v \geq 0, \\ 2\pi - \beta \text{ cn}^{-1}\left(\frac{x}{[3\beta\ I(v,x)]^{1/3}}, \frac{1}{\sqrt{2}}\right) & \text{if } v < 0. \end{cases} \tag{9}$$

---

[1]The angle is not well defined when the action is zero.

1.2.   The 3DoF Hamiltonian system

Now we will introduce the 3DoF Hamiltonian flow whose diffusion properties we would like to investigate. The system consists of three quartic oscillators coupled by a non–integrable perturbation. In cartesian variables, $\boldsymbol{x} = (x_1, x_2, x_3)$ and $\boldsymbol{v} = (v_1, v_2, v_3)$, the Hamiltonian is given by:

$$\tilde{H}(\boldsymbol{v}, \boldsymbol{x}) = \tilde{H}_0(\boldsymbol{v}, \boldsymbol{x}) + \varepsilon \tilde{V}(\boldsymbol{x}), \tag{10}$$

with

$$\tilde{H}_0(\boldsymbol{v}, \boldsymbol{x}) = \frac{1}{2}(v_1^2 + v_2^2 + v_3^2) + \frac{1}{4}(x_1^4 + x_2^4 + x_3^4),$$

$$\tilde{V}(\boldsymbol{x}) = x_1^2(x_2 + x_3); \tag{11}$$

where ε is a *perturbation parameter* that controls the strength of the coupling. For a null value of ε one obtains the integrable Hamiltonian of three independent quartic oscillators. By means of Eqs. (6)(7) the cartesian variables can be expressed in terms of their action–angle counterparts:

$$x_j(I_j, \theta_j) = (3\beta I_j)^{1/3}\mathrm{cn}\left(\frac{\theta_j}{\beta}, \frac{1}{\sqrt{2}}\right), \qquad \text{for} \quad j = 1, 2, 3$$

Moreover, according to Eq. (5) the frequency vector is given by:

$$\boldsymbol{\omega}(\boldsymbol{I}) \equiv \left(\omega_1(I_1), \omega_2(I_2), \omega_3(I_3)\right) = \frac{4}{3}A(I_1^{1/3}, I_2^{1/3}, I_3^{1/3}).$$

On the other hand, for $\varepsilon \neq 0$, only one globally conserved quantity exists, the total energy $\tilde{H}(\boldsymbol{v}, \boldsymbol{x}) = \tilde{H}(\boldsymbol{v}(0), \boldsymbol{x}(0)) \equiv h$. In terms of the action–angle variables (of the unperturbed Hamiltonian) the full Hamiltonian, given by Eqs. (10)(11), can be rewritten as:

$$H(\boldsymbol{I}, \boldsymbol{\theta}) = H_0(\boldsymbol{I}) + \varepsilon V(\boldsymbol{I}, \boldsymbol{\theta}), \tag{12}$$

with

$$H_0(\boldsymbol{I}) \ = A(I_1^{4/3} + I_2^{4/3} + I_3^{4/3}), \tag{13}$$

$$V(\boldsymbol{I}, \boldsymbol{\theta}) = 3\beta I_1^{2/3}\mathrm{cn}^2\left(\frac{\theta_1}{\beta}, \frac{1}{\sqrt{2}}\right)\left[I_2^{1/3}\mathrm{cn}\left(\frac{\theta_2}{\beta}, \frac{1}{\sqrt{2}}\right) + I_3^{1/3}\mathrm{cn}\left(\frac{\theta_3}{\beta}, \frac{1}{\sqrt{2}}\right)\right].$$

This Hamiltonian system has been previously studied by Cincotta *et al.* (2003), Giordano and Cincotta (2004) and Mestre *et al.* (2009). Herein we adopt a fixed value of the total energy:

$$h \equiv 0.485254297422903 \quad \approx 1/4\beta^4, \tag{14}$$

which corresponds to a *characteristic period*[2] of the system very close to 2π.

---

[2]It refers to the period of the x_2, x_3-axial periodic orbits, which remain stable for every ε.

1.3.   A natural base for the action space

Chirikov (1979) introduces a linear change of variables that exploits the resonant geometry in the neighbourhood of the initial conditions that belong to the stochastic layer of an isolated resonance. In this part, we present a 3DoF version of Chirikov's change of variables.

Let $\boldsymbol{m}_g$ be a harmonic vector whose contribution in the Fourier series of V is not zero. Then, Eqs. (12)(13) can be recast as:

$$H(\boldsymbol{I}, \boldsymbol{\theta}) = H_0(\boldsymbol{I}) + \varepsilon V_{\boldsymbol{m}_g}(\boldsymbol{I}) \cos(\boldsymbol{m}_g \cdot \theta) + \varepsilon V(\boldsymbol{I}, \boldsymbol{\theta})$$

$$V(\boldsymbol{I}, \boldsymbol{\theta}) = \sum_{\boldsymbol{m} \neq \boldsymbol{m}_g} V_{\boldsymbol{m}}(\boldsymbol{I}) \cos(\boldsymbol{m} \cdot \theta).$$

Now we assume that the term associated to $\boldsymbol{m}_g$ turns on a non–linear resonance of H_0, along which we want to measure the diffusion. This resonance is usually called the *guiding resonance*. Let $\boldsymbol{I}^r$ be the resonant action in whose neighbourhood we want to perform these measurements. Besides, let $\boldsymbol{\omega}^r \equiv \boldsymbol{\omega}(\boldsymbol{I}^r)$ be the concomitant frequency and let $\boldsymbol{n}_g$ be a vector normal to the resonant surface, called $\mathcal{S}_g$. This last vector can be computed by

$$\boldsymbol{n}_g = \left[\frac{\partial}{\partial \boldsymbol{I}} (\boldsymbol{m} \cdot \boldsymbol{\omega}(\boldsymbol{I})) \right]_{|\boldsymbol{I}=\boldsymbol{I}^r} \tag{15}$$

Moreover, we define the vectors:

$$\begin{aligned}
\boldsymbol{\mu}_1 &= \boldsymbol{m}_g, \\
\boldsymbol{\mu}_2 &= \boldsymbol{\omega}^r / |\boldsymbol{\omega}^r| \\
\boldsymbol{\mu}_3 &= (\boldsymbol{n}_g \wedge \boldsymbol{\omega}^r) / |\boldsymbol{n}_g \wedge \boldsymbol{\omega}^r|,
\end{aligned} \tag{16}$$

and the matrix $\Upsilon \in \mathbb{R}^{3 \times 3}$, whose $i-$th row is the vector $\boldsymbol{\mu}_i$. These vectors are linearly independent *iff* $\boldsymbol{m}_g$ is not perpendicular to $\boldsymbol{n}_g$. A way to ensure this geometrical condition is to assume that the unperturbed Hamitonian satisfies a convexity condition which can be stated as follows in terms of the *Hessian* matrix, C, of H_0:

$$[C(\boldsymbol{I})\boldsymbol{u}] \cdot \boldsymbol{u} \neq \boldsymbol{0} \qquad \forall \boldsymbol{u} \in \mathbb{R}^3 \setminus \boldsymbol{0}.$$

Then, our Hamiltonian is convex because it satisfies:

$$[C(\boldsymbol{I})\boldsymbol{u}] \cdot \boldsymbol{u} = \frac{4}{9} A \left(\frac{u_1^2}{I_1^{2/3}} + \frac{u_2^2}{I_2^{2/3}} + \frac{u_3^2}{I_3^{2/3}} \right) > 0 \qquad \forall \boldsymbol{u} \in \mathbb{R}^3 \setminus \boldsymbol{0}.$$

Let F be a generatrix function given by:

$$F(\boldsymbol{p}, \boldsymbol{\theta}) \equiv \sum_{j=1}^{3} \left(I_j^r + \sum_{k=1}^{3} p_k \Upsilon_{kj} \right) \theta_j.$$

Then, the canonical transformation

$$\psi_i = \frac{\partial F}{\partial p_i} = \sum_k \Upsilon_{ik}\theta_k,$$

$$I_i = \frac{\partial F}{\partial \theta_i} = I_i^r + \sum_k p_k \Upsilon_{ki},$$

can be rewritten explicitly as:

$$\psi_i = \frac{\partial F}{\partial p_i} = \sum_k \Upsilon_{ik}\theta_k,$$

$$p_i = \sum_k \Upsilon_{ik}^{-T}(I_k - I_k^r),$$

(17)

where $\Upsilon^{-T} \equiv (\Upsilon^T)^{-1}$ denotes the inverse matrix of Υ's transpose. This transformation consists of a translation of the origin to the resonant action, followed by a change of base in the phase space. The restriction of the new base to the action space, that will be called *Chirikov's base*, is made up of the $\boldsymbol{\mu}_i$ vectors. As $\boldsymbol{\mu}_1$ is equal to the resonant harmonic, its conjugate angle, $\psi_1 \equiv \boldsymbol{m}_g \cdot \boldsymbol{\theta}$, is the resonant phase of the simple pendulum that models (approximately) the dynamics in this single resonance region, $\boldsymbol{\mu}_2$ is a versor normal to the unperturbed energy surface, $\mathcal{I}_0$, and $\boldsymbol{\mu}_3$ is a versor that is orthogonal to both n_g and $\boldsymbol{\mu}_2$. In other words, it is tangent to the curve formed by the intersection between the surfaces $\mathcal{S}_g$ and $\mathcal{I}_0$, at the point $\boldsymbol{I}^r$.

Eq. (17) can be rewritten in matrix shape as follows:

$$\boldsymbol{\psi} = \Upsilon\boldsymbol{\theta},$$

$$\boldsymbol{p} = \Upsilon^{-T}(\boldsymbol{I} - \boldsymbol{I}^r).$$

Although the variables $\boldsymbol{p}$ and $\boldsymbol{\psi}$ are not necessarily action–angle variables of H_0, it does happen that the components of $\boldsymbol{p}$ are conserved quantities under the unperturbed flow. According to the definition of Υ, the columns of Υ^T correspond to Chirikov's base and, therefore, Υ^T is the matrix that transforms from this new base towards the canonical one. For this reason, p_k $(k = 1, 2, 3)$ are the components of the vector $\boldsymbol{I} - \boldsymbol{I}^r$ in Chirikov's base, whose geometrical importance is as follows (see Cincotta 2002):

- p_1 measures the displacement of $\boldsymbol{I}$ respect to $\boldsymbol{I}^r$, in the direction of the resonant vector.

- p_2 measures the displacement in the direction perpendicular to $\mathcal{I}_0$.

- p_3 measures the displacement along the resonance; i.e. in the direction in which Arnold diffusion (see below) occurs.

1.4.   Statistical quantities and diffusion

We will measure some statistical quantities of the integrals of the unperturbed flow, $\boldsymbol{p}$, defined in the previous subsection. The statistical quantities will be ensemble averages computed numerically over a finite number, N_p, of test particles. This averaging operation will be symbolized by $\langle \cdot \rangle$. The initial conditions of these test particles will be *quasi*–identically chosen in the action space, all of them having the same total energy.

Let $p_j(t, i)$ be the value of the j–th component of the vector $\boldsymbol{p}$, at the time t, associated to the i–th particle. Thus, the (instantaneous) *mean* value of this component is given by:

$$\mu_j(t) \equiv \langle p_j(t) \rangle \equiv \frac{1}{N_p} \sum_{i=1}^{N_p} p_j(t, i),$$

and the corresponding *variance* is:

$$\sigma_j^2(t) \equiv \langle (p_j(t) - \mu_j(t))^2 \rangle \equiv \frac{1}{N_p} \sum_{i=1}^{N_p} (p_j(t, i) - \mu_j(t))^2. \tag{18}$$

Other average quantity that will be used is the *mean square displacement* (MSD):

$$\gamma_j^2(t) \equiv \langle (p_j(t) - p_j(0))^2 \rangle \equiv \frac{1}{N_p} \sum_{i=1}^{N_p} (p_j(t, i) - p_j(0, i))^2, \tag{19}$$

There is not a universal definition of the term *diffusion*. The characterization of this concept depends on many factors: the research area, the system under study, the mathematical tools and the variables used, for example. When working with non–integrable deterministic systems, the term *chaotic diffusion* is usually adopted. In this article, we will say that an ensemble *diffuses* whenever the evolution of its variance performs a non–negligible secular growth in time, allowing this quantity to have oscillations.

In Nekhoroshev (1977), Chirikov (1979), Guzzo *et al.* (2005) and Efthymiopoulos (2008) the concept of *Arnold diffusion* is understood as the one that takes place along the *Arnold web*[3] of a system that satisfies simultaneously the hypothesis of both KAM (Arnold 1989) and Nekhoroshev (Nekhoroshev 1977, see Morbidelli 2006 for a relaxed version) theorems. In this sense, Lega *et al.* (2003), Guzzo *et al.* (2005) and Froeschlé *et al.* (2005) show evidence of Arnold diffusion.

The *standard deviation* is the square root of the variance. This two quantities are sensitive to the degree of diffusion of the ensemble with respect to the geometrical centre ($\boldsymbol{\mu}$) but are not influenced by it. On the other hand, the MSD depends both on the diffusion of the ensemble and on the dynamics of the geometrical centre.

We will consider that a dynamical variable performs a *normal diffusion* if its variance is a linear function of time. Other close definition, generally adopted in the literature, is that *normal diffusion* implies that the MSD behaves linearly with time. One of the paradigmatic systems which present normal diffusion belongs to the class of *stochastic processes*[4], is called *Wiener process* or *Brownian motion* (BM) and was studied by Einstein (1956) among others.

---

[3] Arnold web is the intersection of all the perturbed resonant surfaces with the isoenergetic manifold given by $H(\boldsymbol{I}, \boldsymbol{\theta}) = h$.

[4] The definition of *stochastic process* can be found at Gardiner (2004) or Arnold (1973)

In the case of normal diffusion, it is possible to define a *diffusion coefficient*, denoted $\mathcal{D}_\sigma$, which in this work will be strictly associated to σ_3^2, by the following formula:

$$\mathcal{D}_\sigma \equiv \frac{\sigma_3^2(t) - \sigma_3^2(t_0)}{2(t - t_0)}. \tag{20}$$

Due to inhomogeneities of phase space, $\mathcal{D}_\sigma$ is significant as long as most of the ensemble particles are inside a region small enough so that the diffusion properties are similar all over it. Sometimes it is possible to obtain locally normal behaviours, by limiting the size of the time interval, $|t - t_0|$, in order that the majority of the ensemble stays close to its initial region of action space. This approach, of associating a diffusion coefficient to a neighbourhood of a given initial action, has been adopted in many works (Novaković *et al.* 2010, Todorović *et al.* 2008, Lega *et al.* 2003, Froeschlé *et al.* 2005, Bazzani & Mais 1998, Bazzani *et al.* 1997, Siboni *et al.* 1994, Bountis & Kollmann (1994), for instance). In particular, Lega *et al.* 2003 compute numerically a kind of diffusion coefficient, $\mathcal{D}_\gamma$, related with the MSD, for different values of ε. They detect that the diffusion coefficient goes to zero faster than any power law: $\mathcal{D}_\gamma(\varepsilon) \propto \varepsilon^a$, for any constant a. They also mention that this is compatible with an exponential law: $\mathcal{D}_\gamma(\varepsilon) \propto \exp(-1/\varepsilon^b)$, for some constant b.

Normal diffusion has been widely applied to model dynamical problems in Astronomy, specially in Solar System dynamics.

Diffusion can also be *anomalous*. Mandelbrot & van Ness (1968) introduce, for the first time, the *fractional Brownian motion* (FBM), in terms of a particular *stochastic integral*[5] of the BM.

Any FBM has a standard deviation whose time dependence is given by:

$$\sigma_\eta(t) \propto t^\eta, \tag{21}$$

where η is called the *Hurst exponent*.

This uniparametric family of stochastic processes can be classified in three groups, depending on the value of the exponent:

- *subdiffusive* if $0 < \eta < \frac{1}{2}$,

- *normal* (BM) if $\eta = \frac{1}{2}$,

- *superdiffusive* if $\frac{1}{2} < \eta < 1$.

Cordeiro & Mendes de Souza (2005) and Cordeiro (2006) model diffusion properties, in Solar System resonance domains, with an approach based on FBM.

There are many other types of anomalous diffusion processes. Chechkin *et al.* (2008) presents an introduction to *Lévy flights*. Metzler *et al.* (2007) analyses this process together with *Lévy walks*, subdiffusion processes and fractional *Fokker-Planck* equations. Zaslavsky (2002) revises many fractional stochastic models and their connection with dynamical models, phase space topology and other characteristics of chaos, as the *Poincaré recurrences* and the *sticky domains*.

---

[5]The definition of *stochastic integral* can be found at Gardiner (2004) or Arnold (1973)

In this work we will present numerical evidence of subdiffusion, that occurs during a rather large time interval. Nevertheless, the full diffusion behaviour along a single resonance of this Hamiltonian system is much more complex than the one of a FBM.

1.5. The numerical integrators

The numerical integrator used in this article to compute the trayectories with which the statistical quantities where performed is a symplectic integrator elaborated by E. Teloy, Freiburg, as mentioned in Schlier & Seiter (2000).

This subroutine consists of five integrators, two of 6th order and three of 8th order, called S6a, S6b, S8a, S8b and S8c, respectively.

The kernel of the source code, that performs the integration of a single time step, is called *testsymp*[6]. This code contains the coefficients for the five integration schemes with 35 decimal figures, which is enough for a quadruple precision integration.

After some comparison between the five schemes, the one which better preserved energy was the *S8b*. Regarding measurement of statistical (or average) quantities any of the integrators turned out to be appropriate. The same happened when doing chromatic maps with a chaos indicator called *Smaller Alignment Index* (SALI), introduced by Skokos (2001) , where all of the schemes displayed similar resonant structures. The results presented in this article where obtained with the *S8b* method.

We made a test experiment with two different integration time steps: $\Delta t_{int} = 10^{-2}$ and $\Delta t_{int} = 10^{-1}$, obtaining agreement in the resonant structure of the SALI maps and in the general trend of statistical quantities. Then, for the sake of efficiency, we decided to use in all the experiments of this work, an integration time step of size $\Delta t_{int} = 10^{-1}$.

In this work, we also use the integrator DOPRI8 (Prince & Dormand 1981), in order to build maps of other chaos indicator called *Mean Exponential Growth Factor of Nearby Orbits* (MEGNO), introduced by Cincotta & Simó (2000). Strictly speaking, the indicator effectively used is a *cumulative running time average* of the MEGNO, denoted by $\bar{Y}$. All the integrations where done in cartesian variables.

2. The guiding resonance and its neighbourhood

Now we will characterize geometrically the phase space of the 3DoF Hamiltonian system. In particular, we will specify the guiding resonance where we want to measure diffusion. Such information will be displayed for the case of $\varepsilon = 0.02$.

According to Giordano & Cincotta (2004) the fraction of *chaotic orbits*[7] in $\mathcal{I}_0$, for this parameter value, is of a 38%, while for $\varepsilon = 0.04$ it is approximately of a 92%. Thus, the authors deduced that there is some threshold value of ε close to 0.03, above which chaotic motion prevails in phase space.

---

[6]It can be found at CPC Program Library, Queen's University of Belfast, N. Ireland.

[7]They computed $\bar{Y}$ for initial conditions distributed all over $\mathcal{I}_0$, considering as *chaotic orbits* to those with value $\bar{Y} \geq 2$.

Additionally, in Mestre *et al.* (2009), it was determined analytically, through the *overlap criterion*, that for some critical value of ε (denoted as ε_c) also close to 0.03, the total area covered by resonant regions reaches 90% of the whole area of $\mathcal{I}_0$.

Fig. 1 was built using chromatic levels of $\bar{Y}$ for initial conditions that fill $\mathcal{I}_0$. The $\bar{Y}$ values where binned in five intervals: $[1.993, 2.005]$ (red), $[2.005, 5]$ (green), $[5, 45]$ (blue), $[45, 150]$ (magenta) and $[150, 383]$ (cyan).

The integration of this orbits was done with DOPRI8 for a total time, t_T, of 3500 characteristic periods ($t_T = 3500 \times 2\pi$ time units). In the mentioned figure we can see the resonant structure in the action space, for a section of initial angles given by $\theta_i = \pi/2$ ($i = 1, 2, 3$). Such an angle choice corresponds to null initial values of x_i ($i = 1, 2, 3$), as it can be deduced from the canonical transformation given by Eq. (6) and from properties of the elliptic cosine. Due to this reason, it happens that $\tilde{V}(\boldsymbol{x}(0)) = \tilde{V}(\boldsymbol{0}) = 0$, implying that

$$\tilde{H}(\boldsymbol{v}(0), \boldsymbol{x}(0)) = \tilde{H}_0(\boldsymbol{v}(0), \boldsymbol{0}). \tag{22}$$

Then, the initial value of I_3 was solved from the equation $H_0(\boldsymbol{I}) = h$, using the expression:

$$I_3(0) = \left(\frac{h}{A} - I_1(0)^{4/3} - I_2(0)^{4/3} \right)^{3/4}. \tag{23}$$

In the given figure, we identify in cyan a large region of resonance overlap. Moreover, we can appreciate how the most regular initial conditions (red points) are placed in the complement of the Arnold web, being it filled by the rest of the colors associated to chaotic orbits.

Due to the fact that the SALI indicator tends exponentially to zero for chaotic orbits, we will use the decimal logarithm of this quantity and will use a cut–off criteria that consists of assigning the values $\log(\text{SALI}) = -10$ to every orbit with $\text{SALI} < 10^{-10}$.

In Fig. 2 we display a bidimensional map of this indicator, for a part of $\mathcal{I}_0$, projected onto the $[I_1, I_2]$ plane. To make this figure we have integrated a grid of 1000×1000 initial conditions that belong to the square $(I_1, I_2) \in [0.2, \, 0.4] \times [0.0, \, 0.2]$ and to the same section of initial angles than in the previous figure. For each initial condition, we have used $t_T = 10^4$ time units, which corresponds to more than 10^3 times the characteristic period of the system. The behaviour of the SALI and the color palette are such that the most chaotic orbits appear in black while the most regular ones appear in yellow.

In the given figure, we can see three wide resonances together with many other of smaller size. The nearly horizontal resonance that goes from $(I_1, I_2) \approx (0.2, \, 0.05)$ up to $(I_1, I_2) \approx (0.4, \, 0.1)$ corresponds to the resonant vector $\boldsymbol{m}_g = (2, -3, 0)$, whose direction is indicated with a green arrow. Over the chaotic layer (perturbed separatrix) of this resonance we will study diffusion, i.e. it will be the guiding resonance.

The widest resonance in the figure corresponds to the resonant vector $\boldsymbol{m} = (2, -1, -1)$ and intersects the guiding resonance approximately at $(0.22, \, 0.06)$. The third of the wide resonances correspondes to the resonant vector $\boldsymbol{m} = (2, 0, -2)$ and intersects the guiding resonance approximately at $(0.37, \, 0.10)$.

We can appreciate that for this initial angle values, the chaotic layer of the guiding resonances is thicker than the other neighbouring layers. The thickness
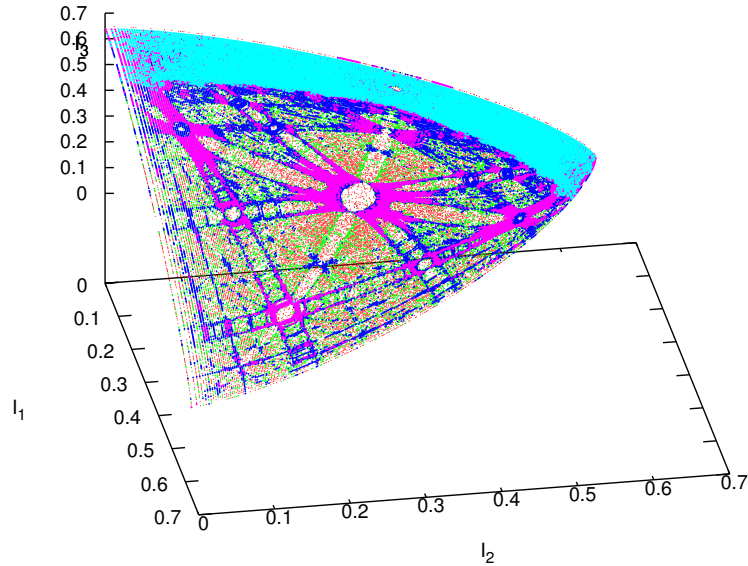
Figure 1.   Cumulative running time average of the MEGNO, $\bar{Y}(t_T)$ with $t_T = 3500 \times 2\pi$, for the unperturbed energy surface in action space, using $\varepsilon = 0.02$. The association between bins and colors si given by: $[1.993, 2.005]$– red, $[2.005, 5]$–green, $[5, 45]$–blue, $[45, 150]$–magenta and $[150, 383]$–cyan.

of the stochastic layer is not constant along the resonance. For example, the neighbourhood of the point $(I_1, I_2) = (0.22,\ 0.06)$, is a resonant region of multiplicity higher than one, so that there exists a type of (unavoidable) *overlap* phenomena. The rest of the resonances perturb the separatrix in much lower ways.

3.   Numerical experiments

In the following experiments we will study the diffusion for values of ε belonging to the following set:

$$\mathscr{E} = \{0.016,\ 0.018,\ 0.020,\ 0.022,\ 0.024,\ 0.026,\ 0.028,\ 0.030\}.$$

3.1.   The ensembles

The selection mechanism of the initial conditions of the ensemble is based in the similarity between the phase space structure in the neighbourhood of a multiplicity-one resonance with respect to the concomitant structure of a perturbed pendulum (Chirikov 1979). This similarity will be checked graphically below.
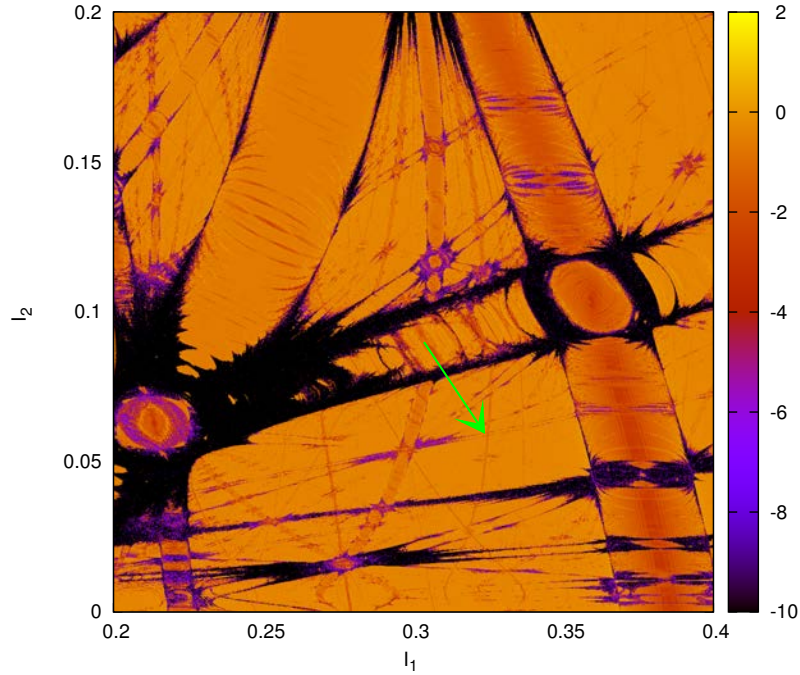
Figure 2. SALI chromatic scale in action space, for $\varepsilon = 0.02$. The color palette is such that the most chaotic orbits appear in black while the most regular ones appear in yellow. The green arrow is parallel to the vector $(2, -3)$.

With regards to the guiding resonance previously specified, and the natural change of variables given by Eq. (17), we define a particular resonant action:

$$\boldsymbol{I}^r = \begin{pmatrix} 0.303515780714549 \\ 0.089930601693199 \\ 0.420802747838120 \end{pmatrix}, \tag{24}$$

and consider the set of initial conditions that belong to a bidimensional portion of phase space parametrized by the following expression in mixed variables[8]:

$$\begin{cases} \psi_1 & \in & [0, \, 2\pi] \\ p_1 & \in & [-10^{-2}, \, 10^{-2}] \end{cases} \tag{25}$$

$$\begin{cases} \theta_1 & = & \pi/2 \\ \theta_2 & = & (2\theta_1 - \psi_1)/3, \qquad mod \quad 2\pi \\ \theta_3 & = & \pi/2 \\ \boldsymbol{I} & = & \boldsymbol{I}^r + p_1 \boldsymbol{m}_g. \end{cases} \tag{26}$$

Thus chosen, the energies of this initial conditions would differ in quantities of $\mathcal{O}(\varepsilon)$, i.e. they would certainly not belong to the same energy surface,

---

[8]The last condition of Eq. (26) is equivalent to initial values $p_2 = p_3 = 0$.

$H(\boldsymbol{I}, \boldsymbol{\theta}) = h$. Projecting the initial conditions given by Eqs. (25)(26) onto the energy surface can be done solving I_1 as a function of I_2, I_3, $\boldsymbol{\theta}$, h and ε. We proceed in the following way. Defining

$$b \equiv \frac{\varepsilon 3\beta}{A} \mathrm{cn}^2 \left(\frac{\theta_1}{\beta}, \frac{1}{\sqrt{2}} \right) \left[\mathrm{I}_2^{1/3} \mathrm{cn} \left(\frac{\theta_2}{\beta}, \frac{1}{\sqrt{2}} \right) + \mathrm{I}_3^{1/3} \mathrm{cn} \left(\frac{\theta_3}{\beta}, \frac{1}{\sqrt{2}} \right) \right],$$

$$c \equiv \frac{h}{A} - (I_2^{4/3} + I_3^{4/3}),$$

and considering the Hamiltonian in action–angle variables given by Eqs. (12)(13), the energy preservation can be expressed in this way:

$$I_1^{4/3} + b I_1^{2/3} = c.$$

This equation has a real non–negative solution:

$$I_1 = \left(\frac{\sqrt{b^2 + 4c} - b}{2} \right)^{3/2}$$

$$\equiv H^{-1}(I_2, I_3, \theta_1, \theta_2, \theta_3; \varepsilon, h).$$

provided that $\sqrt{b^2 + 4c} - b \geq 0$.

For the value of h and for the ε range used in this work, such inversion is possible for all the initial conditions given by Eq. (25).

Figs. 3 and 4 display chromatic maps of $\bar{Y}(3500 \times 2\pi)$, for $\varepsilon \in \mathscr{E}$, in terms of ψ_1 and p_1, which correspond, respectively, to the resonant phase and resonant moment[9].

Taking into account the behaviour of the MEGNO and the color palette used in these pictures, it results that the most chaotic orbits are those in yellow, orange and red colors, while the most regular ones are in dark blue and black colors.

The most chaotic orbits are placed in the stochastic layer associated to the guiding resonance. We observe that, although we are working with a system with an infinite hierarchy o resonances that perturb the guiding one, the resonant phase space structure resembles fairly well the phase space structure of a periodically perturbed pendulum, as is the case in the Standard Map.

The initial conditions of the ensemble were chosen using Eq. (26), with $\psi_1 \in [0, 0.7]$ and $p_1 \in [0, 10^{-5}]$. In this rectangle we stablished a grid of 200 initial conditions. For all the values of $\mathscr{E}$, these ensembles belong to the concomitant stochastic layers. In particular, they are placed close to $(\psi_1, p_1) = (0, 0)$, which corresponds to the unstable equilibrium point of the pendulum model. This initial ensembles are identified in the given figures with green points.

3.2. The measurements

For $\varepsilon \in \mathscr{E}$, we integrated the orbits of the ensembles. The task of the program, before each "printing time", is given by the following steps. First, it computes
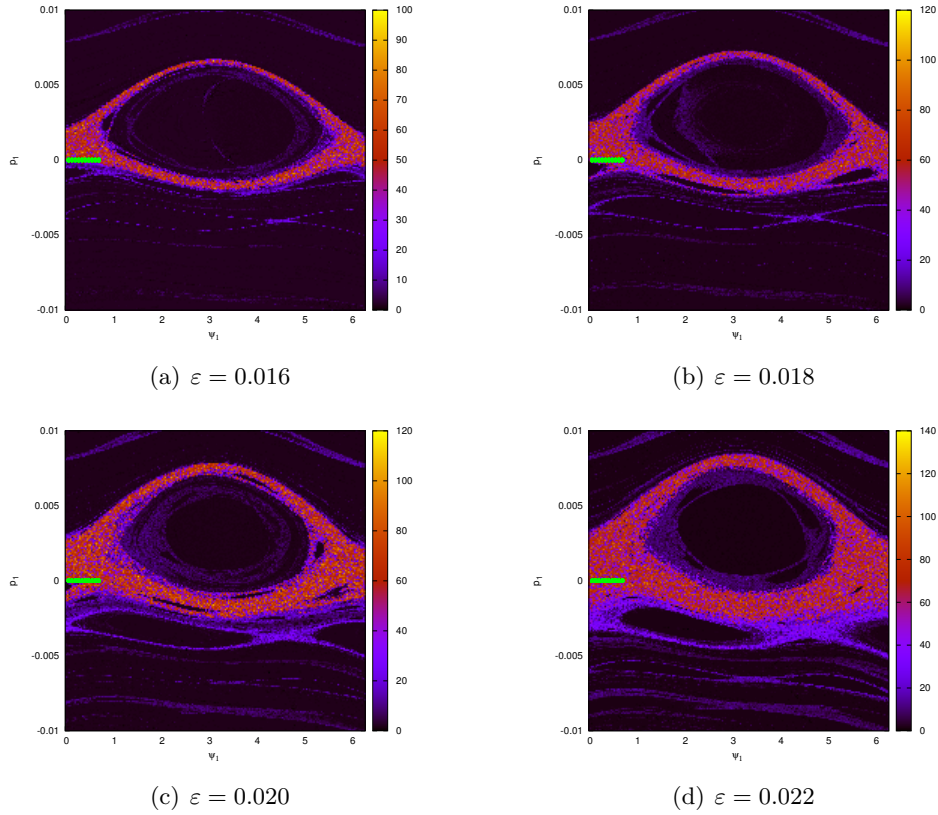
---

[9]As explained in subsection 1.3.

(a) $\varepsilon = 0.016$

(b) $\varepsilon = 0.018$

(c) $\varepsilon = 0.020$

(d) $\varepsilon = 0.022$

Figure 3. $\bar{Y}(3500 \times 2\pi)$, for $\varepsilon \in \{0.016,\ 0.018,\ 0.020,\ 0.022\}$, in a neigh-bourhood of $\boldsymbol{I}^r$ as given by Eq. (24), using pendulum variables (ψ_1, p_1). The green points identify the initial conditions of the ensembles.

the action variables using formula (8). Second, it transforms the actions to Chirikov's base (16), obtaining the values of the vector $\boldsymbol{p}$. Third, it computes the ensemble variances, σ_j^2 (18) and MSDs, γ_i^2 (19), for each one of the variables p_j ($j = 1, 2, 3$). As the base vector $\boldsymbol{\mu}_1$ has a norm $||\boldsymbol{m}_g|| = \sqrt{13} \neq 1$, while the other two base vectors are normalized by definition, the statistical quantities associated to p_1 have been normalized. In other words, from now on, with the notation σ_1, σ_1^2, γ_1 and γ_1^2, we will be really denoting the quantities $\sqrt{13}\sigma_1$, $13\sigma_1^2$, $\sqrt{13}\gamma_1$ and $13\gamma_1^2$, respectively. These are the quantities that we would have measured if we had taken a normalized version of $\boldsymbol{\mu}_1$.

Before analysing the results of the evolution of this statistical quantities for all the ε values, we will qualitatively exemplify the dynamics of this ensembles only for $\varepsilon = 0.016$ and $\varepsilon = 0.020$.

We start with the smallest of these parameter values, using a *double section* technique applied by Lega *et al.* (2003) among others. While integrating the test particles we have taken the surface of section $x_1 = 0$ ($v_1 > 0$) and asked whether the intersecting point also satisfies the condition $x_2^2 + x_3^2 \leq \delta^2$ ($v_2 > 0$, $v_3 > 0$),
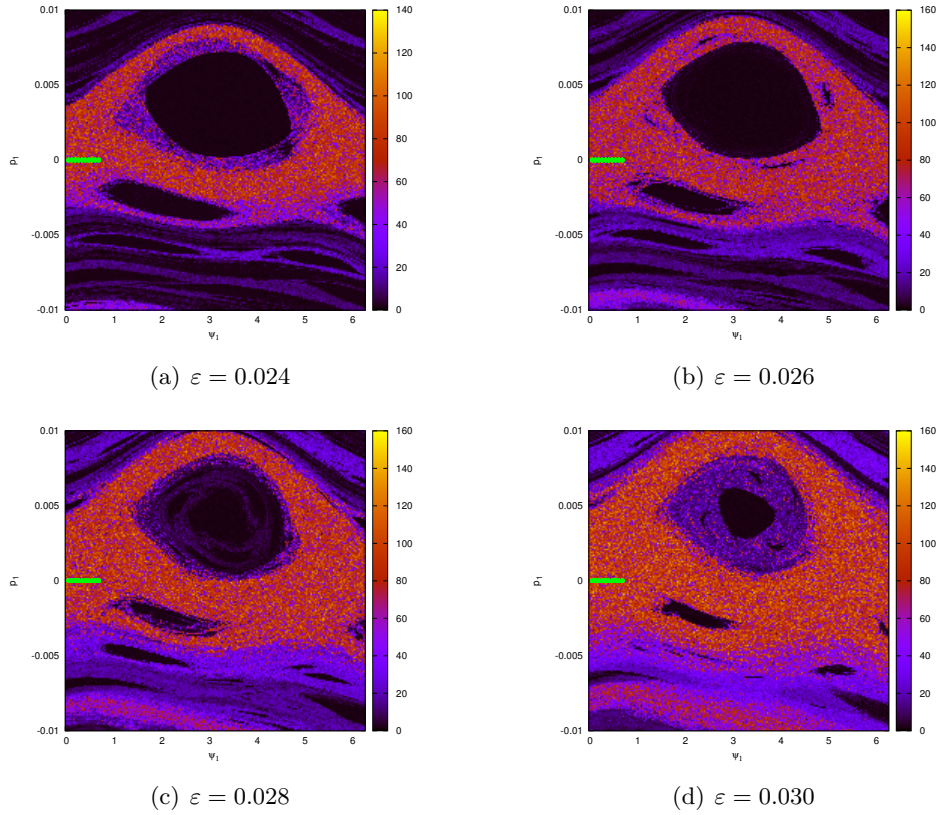
(a) $\varepsilon = 0.024$

(b) $\varepsilon = 0.026$

(c) $\varepsilon = 0.028$

(d) $\varepsilon = 0.030$

Figure 4. $\bar{Y}(3500 \times 2\pi)$, for $\varepsilon \in \{0.024, 0.026, 0.028, 0.030\}$, in a neighbourhood of $\boldsymbol{I}^r$ as given by Eq. (24), using pendulum variables (ψ_1, p_1). The green points identify the initial conditions of the ensembles.

with $\delta = 0.003$. In terms of angle variables the double section is equivalent to $\theta_1 = \pi/2$ with both θ_2 and θ_3 belonging to a certain neighbourhood (whose length decays to zero with δ) of $\pi/2$.

In Fig. 5 we plot all the intersections with the double section that have taken place since $t = 0$ up to five final times: $t = 10^5$, $5 \times 10^5, 10^6, 5 \times 10^6, 10^7$. There we can see how the ensemble expands along the stochastic layer of the guiding resonance. Besides, for $t \geq 5 \times 10^6$, some of the particles are located on "stochastic filaments" that go across the oscillation domain of the resonance.

The location of the initial conditions is shown with a white cross. The fact that the ensemble at the initial time seems to be in the regular zone inside the resonance, is a purely projective effect because in this experiment the initial angle values used in the SALI map are different than the ones used in the initial conditions of the ensemble.

Now we shift towards the $\varepsilon = 0.020$ case, making snapshots of the projections of the ensemble positions onto the $[I_1, I_2]$ plane, without applying any surface of section method. For this reason, this snapshots will show particles projected onto regular regions.
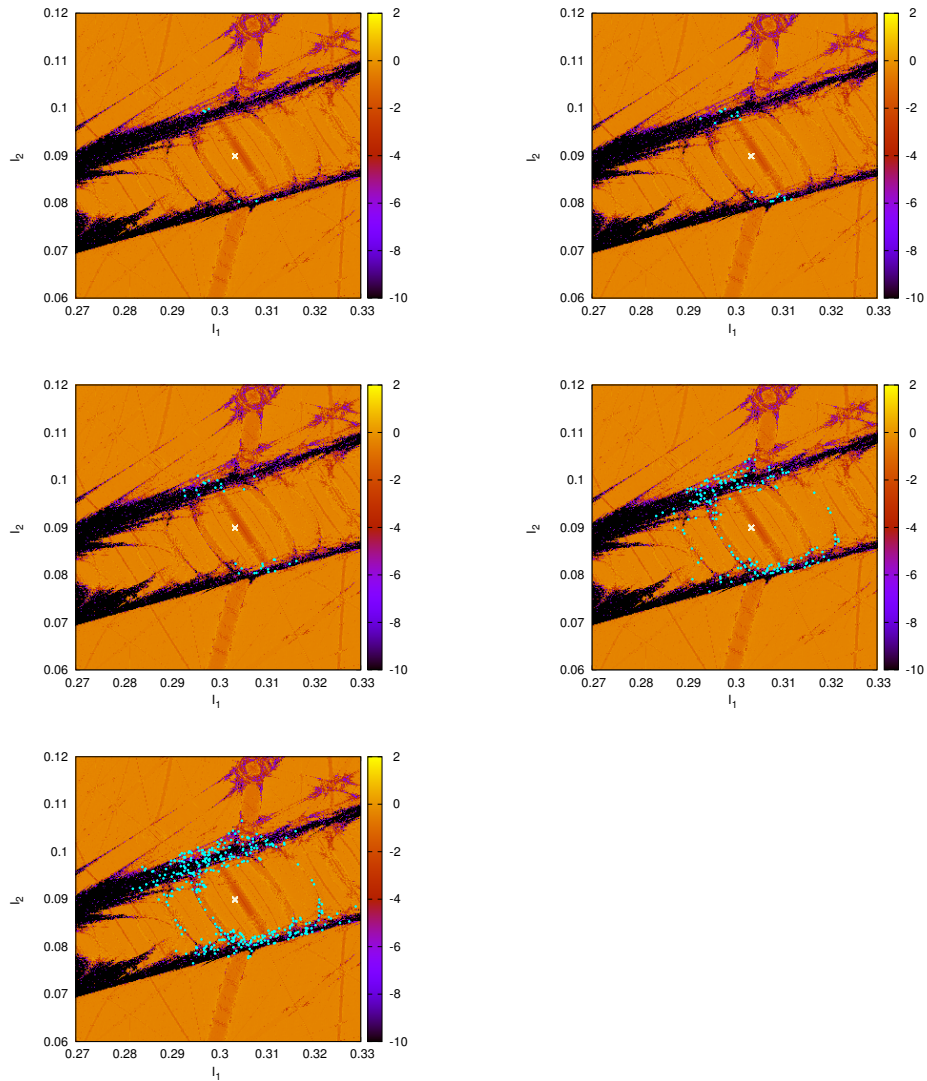
Figure 5.   Intersection of the trajectories with the double section defined by $x_1 = 0$, $(x_2^2 + x_3^2)^{1/2} \leq 0.003$ and $v_i > 0$ $(i = 1, 2, 3)$, projected onto the $[I_1, I_2]$ plane for times: $t \leq 10^5$ (top–left), $t \leq 5 \times 10^5$ (top–right), $t \leq 10^6$ (middle–left), $t \leq 5 \times 10^6$ (middle–right) and $t \leq 10^7$ (bottom). The data corresponds to $\varepsilon = 0.016$.

In Fig. 6–left. we observe the situation of the ensembles for times $t = n \times 10^2$, with $n = 1, \ldots, 5$, in red, green, blue, magenta and yellow colors, respectively. Each point corresponds to a test particle (that appear overlapped due to their high density) and again, the location of the initial conditions is shown with a white cross. We observe mainly two phenomenae. One of them is an oscillatory behaviour of the centre of mass of an order of magnitude similar to the size of
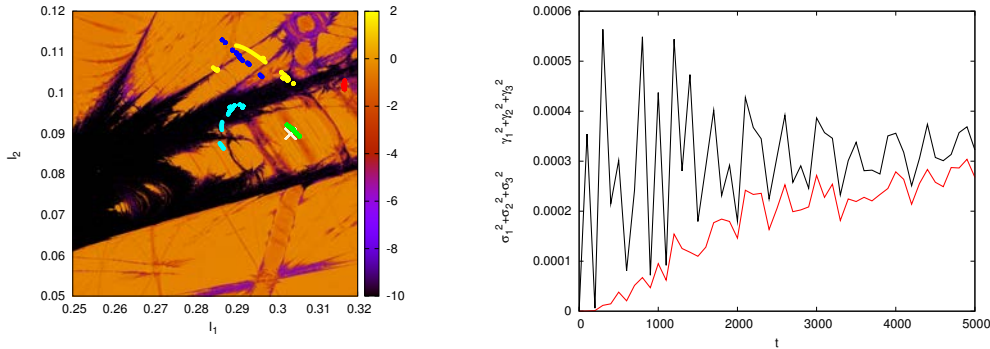
Figure 6. In the left hand side is displayed the location of the ensemble for times 0 (white cross), 10^2 (red), 2×10^2 (green), 3×10^2 (blue), 4×10^2 (cyan) and 5×10^2 (yellow) for $\varepsilon = 0.020$. In the right hand side is displayed $\sigma_*^2 \equiv \sigma_1^2 + \sigma_2^2 + \sigma_3^2$ (red) and $\gamma_*^2 \equiv \gamma_1^2 + \gamma_2^2 + \gamma_3^2$ (black), for $0 \leq t \leq 5 \times 10^3$ and the same ε value.

the resonance. The other one is a dispersion of the ensemble, related to the decorrelation of the resonant phases associated to the particles.

Both phenomenae can be quantified by observing at Fig. 6–right, where the evolution of the quantities $\sigma_*^2 \equiv \sigma_1^2 + \sigma_2^2 + \sigma_3^2$ (red) and $\gamma_*^2 \equiv \gamma_1^2 + \gamma_2^2 + \gamma_3^2$ (black), for the first 5×10^3 time units are shown.

The MSD presents oscillations. In particular, there is a peack of height approximately equal to 3.5×10^{-4} at $t = 100$. This implies that the distance of the ensemble with respect to the initial condition is of the order of $\gamma_* \approx (3.5 \times 10^{-4})^{1/2} \approx 0.019$, a fact that is verified graphically by the location of the red points in the snapshot.

On the other hand, the variance grows slower and with a more secular character than the MSD. Only after times $\gtrsim 2000$ the former quantity adquires values of the order of magnitude of the latter one. In the forthcoming diffusion measurements we will not consider this initial transient time.

The mentioned equivalence between these two quantities, for times outside the transient, can be appreciated in Fig. 7, where we give the positions of the particles for times of the form: $t = 3 \times 10^n$, with $n = 3, \ldots, 6$.

There it can be seen how the ensemble gradually expands. For times of the order of 10^3 (Fig. 7–top–left) the extension of the ensemble in the direction parallel to $\boldsymbol{\mu}_1$ is a bit larger that the one in the direction parallel to $\boldsymbol{\mu}_3$. Nevertheless, the situation is inverted when considering times of the order of 10^6 (Figs. 7–bottom–right), where there is evidence of a net macroscopic transport along the guiding resonance. A qualitatively similar situation was previously obtained for the $\varepsilon = 0.016$ case, in which the double section technique allowed a better visualization of the diffusion process in the action space. Fig. 8–left, shows the ensemble for 10^7 time units and gives evidence of the fact that there are particles that arrive close to the neighbouring resonances: $(2, -1, -1)$ and $(2, 0, -2)$. The $\sigma_*^2(t)$ y $\gamma_*^2(t)$ curves are displayed at Fig. 8–right for $0 \leq t \leq 10^7$. With the purpose of knowing the functional dependence of $\sigma_*^2(t)$, we assumed
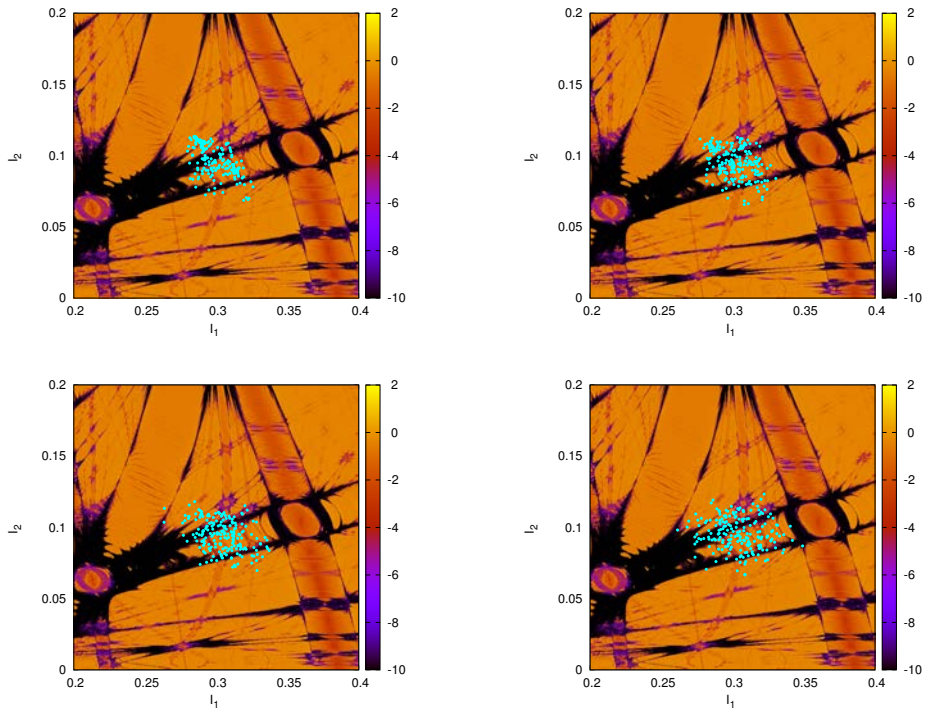
Figure 7.   Ensemble projection onto the $[I_1, I_2]$ plane for times 3×10^3 (top–left), 3×10^4 (top–right), 3×10^5 (bottom–left) and 3×10^6 (bottom–right). The data corresponds to $\varepsilon = 0.020$.
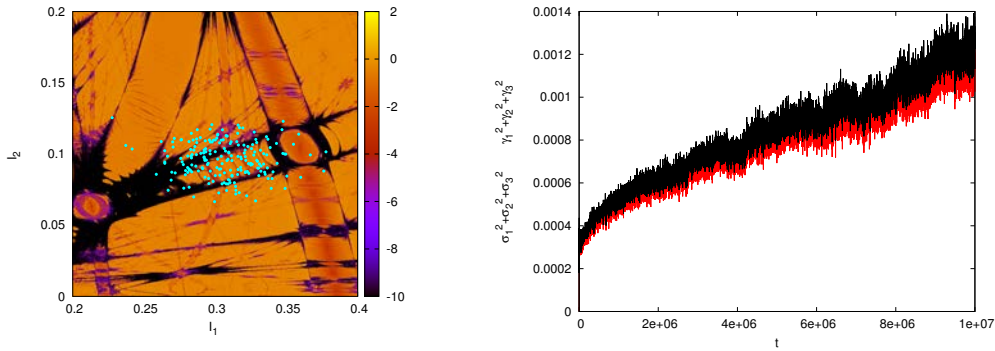

Figure 8.   The left image shows a snapshot of the ensemble for $t = 10^7$ and $\varepsilon = 0.020$. The right image displays the curves of $\sigma_*^2 \equiv \sigma_1^2 + \sigma_2^2 + \sigma_3^2$ (red) and $\gamma_*^2 \equiv \gamma_1^2 + \gamma_2^2 + \gamma_3^2$ (black), for $0 \leq t \leq 10^7$ and the same parameter value.

as *ansatz* a power law:

$$\sigma_*^2 = \kappa t^w, \tag{27}$$

which is equivalent to the concomitant dependence of the FBM (21), associating w with the Hurst exponent according to $w = 2\eta$.

Applying the decimal logarithm to both sides of Eq. (27), we obtain

$$\log(\sigma_*^2) = \log(\kappa) + w \log(t).$$

This implies a linear relation between the logarithm of time and the logarithm of variance, being its slope the exponent of the power law. A numerical fit of the variables $\log(\kappa)$ and w, for $t \in [10^5, 10^{6.5}]$, gave that $\kappa \approx 2.23 \times 10^{-5}$ and that $w \approx 0.225$.

The fact that w turned out to be considerably smaller than unity, is an experimental evidence that during this time interval the diffusion is anomalous, and in particular, it is a subdiffusion process.

The length of this interval, that includes around 3×10^6 time units, is not negligible. On the contrary, observing Fig. 7–bottom–right, it can be stated that this time interval is a significant part of the life of the ensemble in which it is strictly included inside the guiding resonance, before that the particles arrive to the resonance crossings.

Fig. 9–left shows $\sigma_*^2(t)$ in logarithmic scale, together with the linear fit, while Fig. 9–right shows that this fit diverges from the real measurement, for larger times.
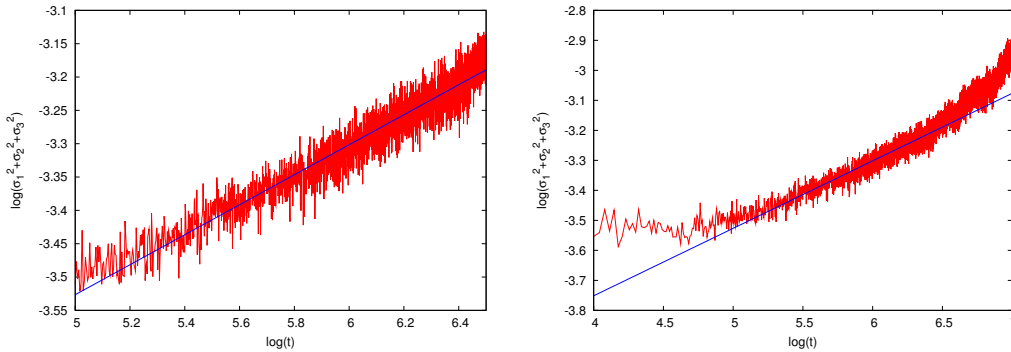

Figure 9.   The left image displays $\log(\sigma_*^2)$ as a function of $\log(t)$ (red) together with the linear fit of the slope $w \approx 0.225$ (blue), for $t \in [10^5, 10^{6.5}]$. The right image shows that this power law behaviour ($\sigma_*^2 \propto t^w$) stops being valid when considering a wider time interval.

Now we will consider the evolution of the variances associated to each direction individually. Fig. (10) displays the curves $\sigma_1^2(t)$, $\sigma_2^2(t)$ and $\sigma_3^2(t)$ in colors red, green and blue, respectively, for $t \in [10^5, 10^7]$. We observe an approximate stabilization in constant values of σ_1^2 and σ_2^2, while σ_3^2 shows sustained growth. The stabilization of σ_2^2 is due to the conservation of the total energy. The stabilization of σ_1^2 is due to the fact that the resonant moment (p_1) is bounded due to the (aproximate and local) conservation of the energy of the resonant normal form, i.e. conservation of energy of the pendulum.

This behaviour of the variances is the one that should take place in an Arnold diffusion process (subsection 1.4.), during the stage in which the ensemble
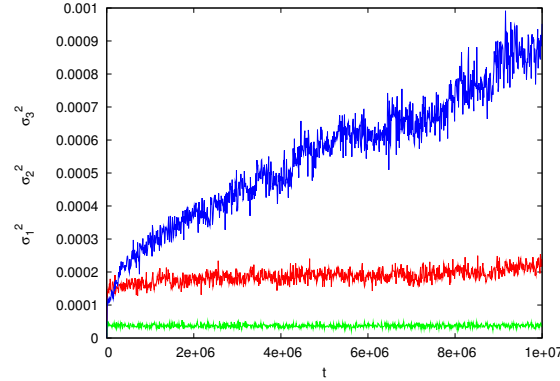
Figure 10. $\sigma_1^2(t)$, $\sigma_2^2(t)$ and $\sigma_3^2(t)$ respectively in red, green and blue colors, for $\varepsilon = 0.020$.

is mostly located inside the stochastic layer of the original guiding resonance. Once diffusion has proceeded considerably along other resonances, Chirikov's base associated to the original guiding resonance stops having geometrical sense. When particles change their guiding resonance, there is a change in $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_3$ as they both depend on $\boldsymbol{m}_g$; and there is a change in $\boldsymbol{\mu}_2$ due to the convexity of $\mathcal{I}_0$.

Fig. 11 displays the values of the three variances for each value of $\mathscr{E}$.

There it can be seen the change in the behaviour of variances as the intensity of the perturbation gradually increases. Figs. 11 (a,b,c) show that for $\varepsilon \in \{0.016,\ 0.018\}$ the behaviour is qualitatively similar to the one previously described for $\varepsilon = 0.020$.

Figs. 11 (d,e,f) show that for $\varepsilon \in \{0.022,\ 0.024,\ 0.026\}$, σ_1^2 does not stabilize, but it has a net growth. On the other hand, σ_2^2 still converges approximately to a constant value, while σ_3^2 is still the variance that grows largely in the analysed time interval.

Finally, in Figs. 11 (g,h) there is evidence of the fact that for $\varepsilon \in \{0.028,\ 0.030\}$, σ_2^2 presents a net growth and, besides, the magnitude of the growth of σ_1^2 is of the order of the one of σ_3^2. This behaviour does not agree with an Arnold diffusion at all. Therefore, for $\varepsilon = 0.028$, which is smaller than ε_c (section 2.), there is a diffusion situation geometrically more similar to an *overlapping regime* than to an Arnold diffusion.

Nevertheless, there is no contradiction in these facts because the quantity ε_c gives information of the global state of overlap of phase space, while in the local neighbourhood of our particular resonant action under study, the overlapping starts to manifest for smaller values of ε.

This argument can be experimentally proved by considering the two bottom images of Fig. 4, that show that for values of ε equal to 0.028 and 0.030, the resonant region is in an advanced state of overlap.

Once that we have analysed the way in which the variances (associated to the three independent directions of Chirikov's base) change with respect to ε, we will go ahead towards quantitative measurements of the evolution of only one of the variances: σ_3^2.
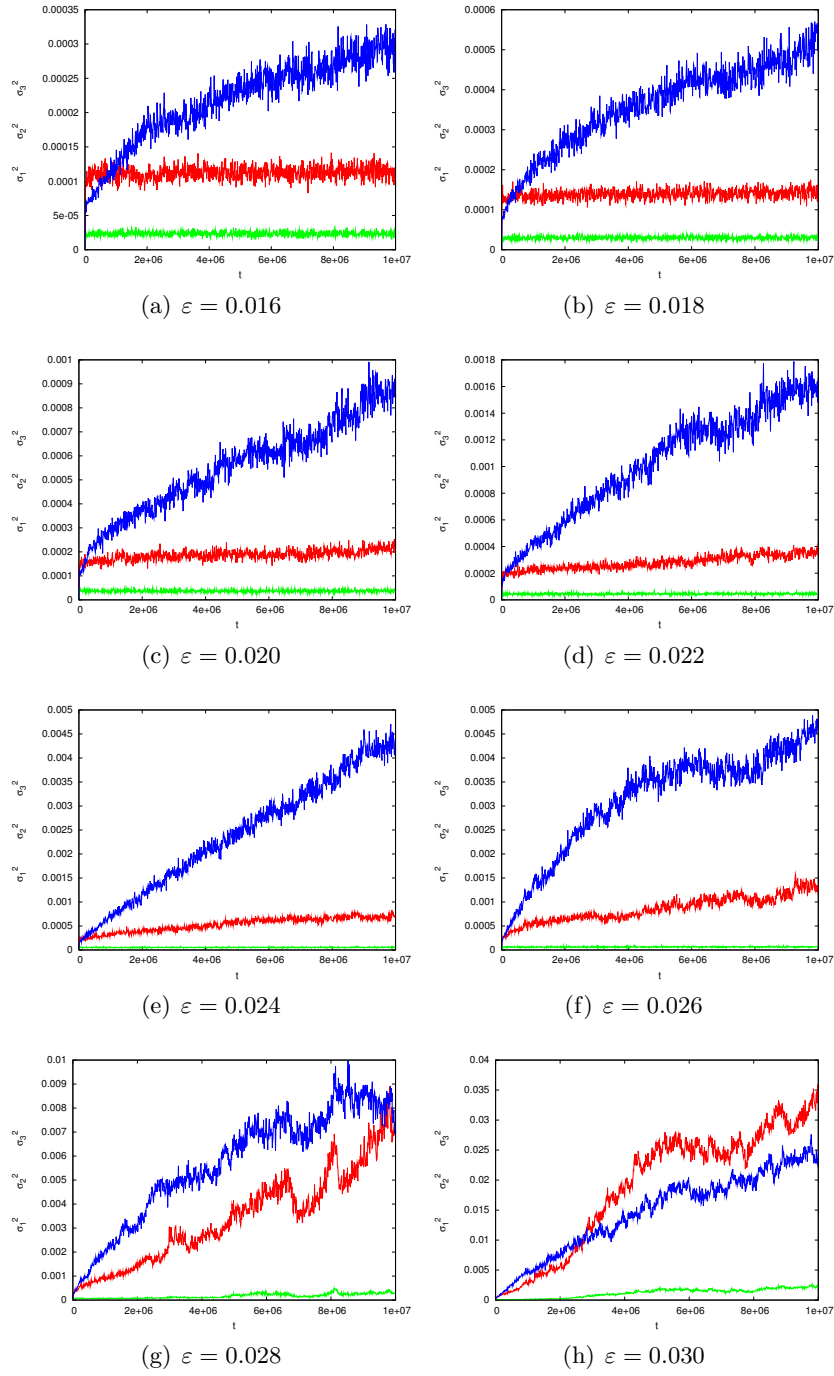
Figure 11. $\sigma_i^2(t)$ $(i = 1, 2, 3)$ respectively in red, green and blue colors, for $\varepsilon \in \mathscr{E}$.

Fig. 12–left shows $\sigma_3^2(t)$ in logarithmic scale for $\varepsilon \in \mathscr{E}$, in colors red, green, blue, magenta, cyan, yellow, black and orange, respectively. We observe that, in general, for each fixed time t, the larger the value of ε, the larger the value of $\sigma_3^2(t)$. For each curve, we have made a linear fit for $t \in [10^5, 10^{6.5}]$, proceeding in a similar way to what had previously been done with σ_*^2 and Eq. (27).

Table 1 displays the values of this numerical fits. In the fifth column the quantity w_* has been added, which corresponds to a fitted value of the exponent for a wider time interval, $t \in [10^5, 10^7]$, while the sixth column shows the concomitant percentage relative difference. It can be seen that for some values of ε, such a difference reaches values higher than 15%. This is an example of the fact that the behaviour of statistical quantities in systems with divided phase space can be highly dependent on the size of the time interval.

Fig. 12–right shows the agreement between the linear fit done in $10^5 \leq t \leq 10^{6.5}$ and the concomitant evolutions of variances. Analogously, Fig. 13–left displays the same information for the range $10^5 \leq t \leq 10^7$. Fig. 13–right shows the values of both exponents, w and w_*, as a function of ε in red and black colors, respectively.
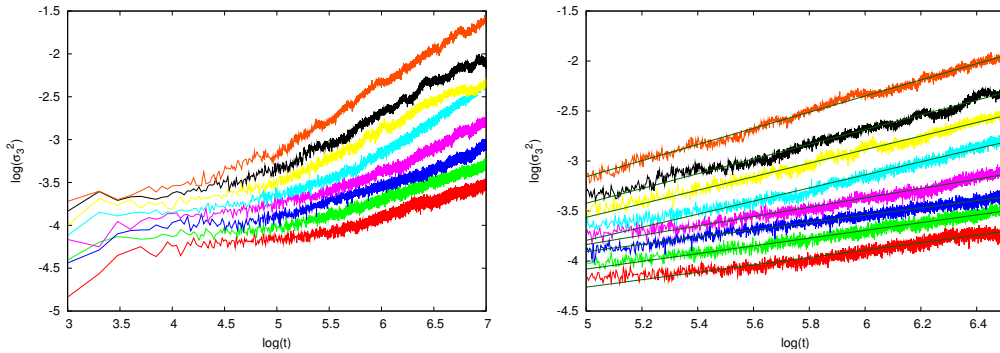

Figure 12. $\sigma_3^2(t)$ in logarithmic scale, for $\varepsilon \in \mathscr{E}$, in colors red, green, blue, magenta, cyan, yellow, black and orange, respectively. In the left hand side image the time range is $t \in [10^3, 10^7]$. In the right hand side image, the concomitant straight lines that fit approximately the data for $t \in [10^5, 10^{6.5}]$ have been added.

Due to the fact that every measured value of w (and of w_*) is less than unity, it can be concluded that for the two temporal ranges, and for the eight ε values analysed, the variance evolution corresponds to an anomalous diffusion process of the subdiffusive type.

Therefore, having also in mind the information obtained from Figs. 11 (a,b,c), it could be said that for $\varepsilon \in \{0.016, 0.018, 0.020\}$ and for the two time ranges used, an "Arnold subdiffusion" process might be taking place.

In spite of this, Fig. 14 indicates that when considering times sufficiently shorter than the ones recently used, the variances follow an approximately linear behaviour. In order to have an estimation of the average rate of change of the variance during the earliest 5×10^5 time units, we will compute the diffusion coefficient $\mathcal{D}_\sigma$ (20), choosing as initial time $t_0 = 10^4$ and as final time $t = 5 \times 10^5$.

ε	w	$\log(\kappa)$	κ	w_*	$\frac{|w-w_*|}{w} \times 100$
0.016	0.370366	-6.11277	7.713×10^{-7}	0.365061	~ 1
0.018	0.387073	-6.01603	9.638×10^{-7}	0.395128	~ 2
0.020	0.363961	-5.72554	1.881×10^{-6}	0.444984	~ 22
0.022	0.466993	-6.17054	6.752×10^{-7}	0.564104	~ 20
0.024	0.654695	-7.06677	8.575×10^{-8}	0.755714	~ 15
0.026	0.676915	-6.94744	1.129×10^{-7}	0.562896	~ 16
0.028	0.736347	-7.10383	7.874×10^{-8}	0.655036	~ 11
0.030	0.807722	-7.19636	6.363×10^{-8}	0.756279	~ 6

Table 1.   Results of the numerical fit of the power law, similar to the ansatz given by Eq. (27), with the substitution of σ_* with σ_3, for $t \in [10^5, 10^{6.5}]$. In the fifth column the quantity w_* has been added, which corresponds to a fitted value of the exponent for a wider time interval, $t \in [10^5, 10^7]$, while the sixth column shows the concomitant percentage relative difference.
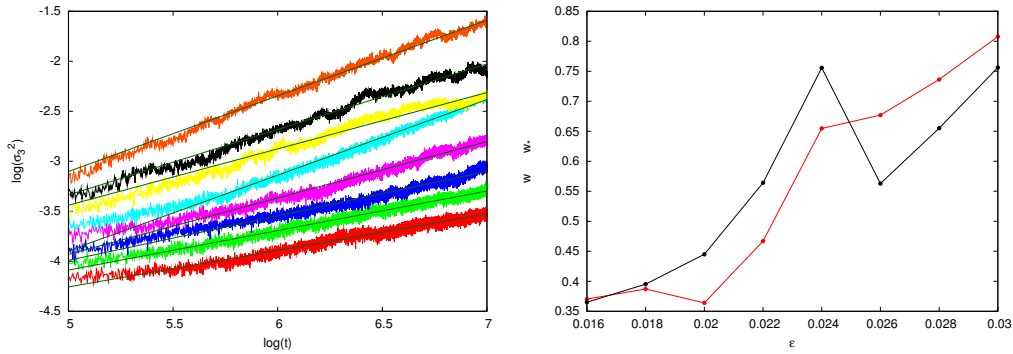

Figure 13.   The left hand side image shows the curves $\sigma_3^2(t)$ in logarithmic scale, for $\varepsilon \in \mathscr{E}$, in colors red, green, blue, magenta, cyan, yellow, black and orange, respectively, together with the concomitant straight lines that fit approximately the data, for $t \in [10^5, 10^7]$. The right hand side image shows the $w(\varepsilon)$ and $w_*(\varepsilon)$ curves in red and black colors, respectively.

This information is provided by Table 2 and by Fig. 15, as a function of ε. Looking at this figure it is not possible to determine the functional relation between the diffusion coefficient and the perturbative parameter. Due to the smallness of the ε–range used, the data could be compatible with many functional forms. Nevertheless, it is noticeable that $\mathcal{D}_\sigma$ increases when ε increases, as expected.

4.   Conclusion

In this work we have studied, through numerical experiments, a diffusion process that takes place along the stochastic layer of a guiding resonance of a particular *quasi*–integrable 3DoF Hamiltonian flow.
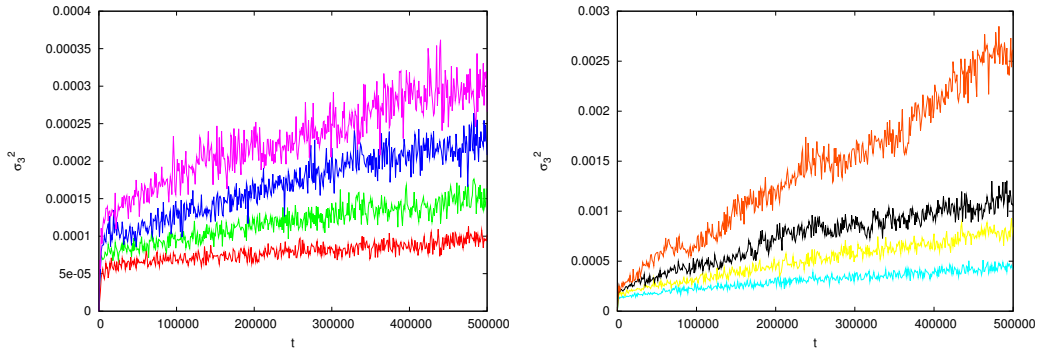
Figure 14. $\sigma_3^2(t)$ for $\varepsilon \in \mathscr{E}$ and $t \leq 5 \times 10^5$. Each color corresponds to an ε value, following the same criterion of Fig. 12.

ε	$\sigma_3^2(t_0)$	$\sigma_3^2(t)$	$\mathcal{D}_\sigma$
0.016	0.571×10^{-4}	0.102×10^{-3}	4.59×10^{-11}
0.018	0.739×10^{-4}	0.162×10^{-3}	9.01×10^{-11}
0.020	0.107×10^{-3}	0.208×10^{-3}	1.02×10^{-10}
0.022	0.136×10^{-3}	0.265×10^{-3}	1.31×10^{-10}
0.024	0.140×10^{-3}	0.414×10^{-3}	2.78×10^{-10}
0.026	0.153×10^{-3}	0.839×10^{-3}	6.99×10^{-10}
0.028	0.237×10^{-3}	0.105×10^{-2}	8.30×10^{-10}
0.030	0.288×10^{-3}	0.261×10^{-2}	2.37×10^{-9}

Table 2. $\mathcal{D}_\sigma(\varepsilon)$, using $t_0 = 10^4$ and $t = 5 \times 10^5$.
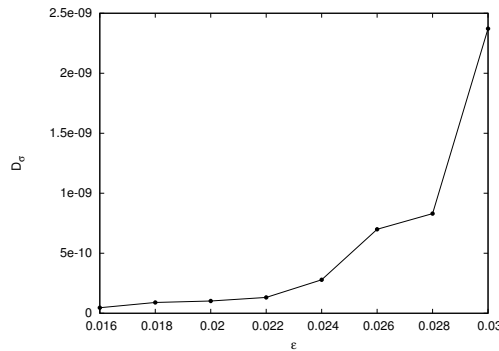

Figure 15. $\mathcal{D}_\sigma(\varepsilon)$, using $t_0 = 10^4$ and $t = 5 \times 10^5$.

We have considered eight values of the perturbation parameter (ε). For each one of these values we have used an ensemble of test particles whose initial conditions were chosen inside the stochastic layer of the perturbed guiding resonance.

We have measured the evolution of the variances and mean square displacements, in the three directions of Chirikov's base, determining that for $\varepsilon \leq 0.02$

the results geometrically agree with an Arnold diffusion. Such a process is characterized by the fact that $\sigma_1^2(t)$ and $\sigma_2^2(t)$ stabilize around constant values, while $\sigma_3^2(t)$ performs a nearly secular growth in the time interval under analys: $t \in [0, 10^7]$.

For higher ε values, we have found that $\sigma_1^2(t)$ has an average growth rate that increases with ε, in agreement with the fact that the degree of influence of the overlap with neighbouring resonances increases with ε.

For $t \lesssim 10^7$ we have observed an anomalous behaviour of $\sigma_3^2(t)$ for all the eight ε values. Through numerical fits of $\sigma_3^2(t)$, of the power law type, we have found that all the exponents are lower than one, In other words, all the Hurst exponents are lower than one half. This is experimental evidence of the existence of a subdiffusion process that *effectively* takes place in this given time interval.

On the other hand, for $t \lesssim 5 \times 10^5$ we have observed that these variances behave approximately linearly (normal diffusion). This allowed us to compute diffusion coefficients, $\mathcal{D}_\sigma$, associated to the rate of growth of the variances. We have found that $\mathcal{D}_\sigma$ increases with ε.

References

Arnold, L. 1973, Stochastic Differential Equations: theory and applications

Arnold, V. I. 1989, Mathematical Methods of Classical Mechanics

Bazzani, A., Beccaceci, L., Bigliardi, L., & Turchetti, G. 1997, American Institute of Physics Conference Series, 395, 109

Bazzani, A., & Mais, H. 1998, Nonlinear and stochastic beam dynamics in accelerators - A chalenge to theoretical and computational physics, 74

Bountis, T., & Kollmann, M. 1994, Physica D, 71, 122

Chechkin, A. V., Metzler, R., Klafter, J., & Gonchar, V. Y. 2008, Anomalous Transport, 129

Chirikov, B. V. 1979, Phys. Rep, 52, 263

Cincotta, P. M. 2002, New Astron. Rev, 46, 13

Cincotta, P. M., Giordano, C. M., & Simó, C. 2003, Physica D, 182, 151

Cincotta, P. M., & Simó, C. 2000, A&AS, 147, 205

Cordeiro, R. R. 2006, AJ, 132, 2114

Cordeiro, R. R., & Mendes de Souza, L. A. 2005, A&A, 439, 375

Efthymiopoulos, C. 2008, Celest. Mech. Dyn. Astron, 102, 49

Einstein, A. 1956, Annalen der Physik (Leipzig), 17, 549

Froeschlé, C., Guzzo, M., & Lega, E. 2005, Celest. Mech. Dyn. Astron, 92, 243

Gardiner, P. R. 2004, Handbook of Stochastic Methods

Giordano, C. M. & Cincotta, P. M. 2004, A&A423, 745

Gradshtyn, I. S., & Ryzhik, I. M. 1980, Table of Integrals, Series and Products

Guzzo, M., Lega, E., & Froeschlé, C. 2005, Disc. Cont. Dyn. Syst. B, 5, 687

Lega, E., Guzzo, M., & Froeschlé, C. 2003, Physica D, 182, 179

Mandelbrot, B. B. & van Ness, J. W. 1968, SIAM Review, 10, 422

Mestre, M., Cincotta, P. M., & Giordano, C. M. 2009, Int. J. Nonlinear Mech, 44, 180

Metzler, R., Chechkin, A. V., Klafter, J. 2007, ArXiv e-print

Morbidelli, A. 2006, Modern Celestian Mechanics – Aspects of Solar System Dynamics

Nekhoroshev, N. N. 1977, Russ. Math. Surv, 32, 1

Novaković, B., Tsiganis, K., & Knežević, Z. 2010, MNRAS, 402, 1263

Prince, P. J., & Dormand, J. R. 1981, J. Comput. Appl. Math, 7–1, 67

Schlier, C. & Seiter, A. 2000, Comput. Phys. Commun, 130, 176

Siboni, S., Turchetti, G., Vaienti, S. 1994, J. Stat. Phys, 75, 167

Skokos, C. 2001, J. Phys. A – Math. Gen, 34, 10029

Todorović, N., Lega, E., & Froeschlé, C. 2008, Celest. Mech. Dyn. Astron, 102, 13

Zaslavsky, G. M. 2002, Phys. Rep, 371, 461

ORAL COMMUNICATION

Chaos detection tools: The `LP-VIcode` and its applications

L. A. Darriba[1,2], N. P. Maffione[1,2], P. M. Cincotta[1,2], C. M. Giordano[1,2]

(1) Grupo de Caos en Sistemas Hamiltonianos - Facultad de Ciencias Astronómicas y Geofísicas - UNLP
(2) Instituto de Astrofísica de La Plata (CCT La Plata - CONICET, UNLP)

Abstract.   A very important topic in galactic dynamics is the detection of instabilities of a given system and the possible appearance of chaos. Such a chaotic bahaviour can be detected and studied by means of variational chaos indicators (CIs). The CIs are based on the study of the evolution of initial deviation vectors, which makes these techniques specially sensitive to indicate the presence of chaos. Notwithstanding their special sensitiveness to identify chaos, the CIs are still good alternatives to determine also the resonance web.

On the other hand, the so-called spectral analysis methods are based on the study of some quantity (e.g. the frequency) on a single orbit, which turns these techniques very efficient for the determination of the resonant structure of the system.

The analysis of the interaction among chaotic and regular components as well as the determination of the resonant structure of the Hamiltonian leads to a deeper understanding of the system's dynamics. Despite the advantages of the simultaneous application of both types of techniques, many researchers keep applying only one of them.

Herein, we present an alpha version of a program coded in Fortran, the `LP-VIcode`. Although the code is in a developing stage, it can compute several CIs, and here we apply it together with the Frequency Modified Fourier Transform (FMFT) (Sidlichovský & Nesvorný 1996) to study the stationary space (Schwarzchild 1993) of an average realistic Hamiltonian model (Muzzio et al. 2005).

Using the `LP-VIcode`, in Maffione et al. (2011b) and Darriba et al. (submitted) the authors suggest an efficient package of CIs to study a general Hamiltonian. Here the research is extended to show that the complementary use of the `LP-VIcode` and the spectral analysis methods is highly recommended to study a realistic Hamiltonian model.

1. Introduction

The detection of chaotic behaviour in any dynamical system, such as galaxies or planetary systems, may be carried out by means of several techniques. The most commonly used are those based on spectral analysis and on the study of the evolution of the deviation vectors, the so-called variational chaos indicators (CIs hereafter). Among the CIs we can find several examples: the Lyapunov Indicators (LIs), the Mean Exponential Growth factor of Nearby Orbits (MEGNO), the Smaller Alignment Index (SALI) and its generalized version, the Generalized Alignment Index (GALI), the Fast Lyapunov Indicator (FLI), its first order variant, the Orthogonal Fast Lyapunov Indicator (OFLI) and its second order variant, the $OFLI_2^{TT}$, the Spectral Distance (D) and the Dynamical Spectras of Stretching Numbers (SSNs), the Relative Lyapunov Indicator (RLI) and the Average Power Law Exponent (APLE), among others (see our text and references below). All of them have their own advantages and disadvantages, making them particularly suitable for different situations.

It could be very interesting and fruitful to have the possibility of easily computing *any* CI. This is the main goal of the first part of this work where we present an alpha version of the `LP-VIcode` (the acronym for La Plata-Variational Indicators code). The aim of the code, as its name suggests, is to easily compute several CIs and, for instance, in Maffione et al. (2011b) (hereafter M11) and Darriba et al. (submitted) (hereafter D12), the authors use it to make a comparative evaluation among them in order to analize the main advantages and drawbacks of each indicator. On the other hand, in the second part of this work we present the results of the application of the CIs implemented within the code, together with a spectral analysis method, to show that the complementary use of both types of chaos detection tools is strongly advisable.

The CIs can recover the resonant structure (see for instance, Kaneko & Konishi 1994; Cincotta et al. 2003; Froeschlé et al. 2006; Lukes-Gerakopoulos et al. 2008) but as they are based on the concept of local exponential divergence, they are specially sensitive to indicate the presence of chaos. The introduction of the Lyapunov Characteristic Exponents (LCEs) (see e.g. Skokos 2010 for a current thorough discussion) as well as its numerical implementation (Benettin et al. 1980; Skokos 2010) was a major contribution to the advance of chaos detection. The integration time is bounded, so we are able to reach just truncated approximations of the theoretical LCEs, i.e. the already mentioned LIs (see Benettin et al. 1976; Benettin et al. 1980; Froeschlé 1984; Tancredi et al. 2001 and Skokos 2010). A drawback of the computation of the LIs is their very slow speed of convergence. Nevertheless, since the introduction of the first definition of the LI, a large number of CIs have improved the LIs' slow speed of convergence holding many useful characteristics of it. The following CIs are already implemented in the `LP-VIcode` with the LI: the MEGNO (Cincotta & Simó 2000; Cincotta et al. 2003; Giordano & Cincotta 2004; Goździewski et al. 2005; Gayon & Bois 2008; Lemaître et al. 2009; Hince et al. 2010; Maffione et al. 2011a; Compère et al. 2011) and a quantity derived from it: the MEGNO's Slope Estimation of the largest LCE (SElLCE), the SALI (Skokos 2001; Skokos et al. 2004; Széll et al. 2004; Bountis & Skokos 2006; Carpintero

2008; Antonopoulos et al. 2010) and the GALI (Skokos et al. 2007; Skokos et al. 2008; Manos & Athanassoula 2011), the FLI (Froeschlé et al. 1997a; Froeschlé et al. 1997b; Froeschlé & Lega 1998; Froeschlé & Lega 2000; Lega & Froeschlé 2001; Guzzo et al. 2002; Froeschlé & Lega 2006; Paleari et al. 2008; Todorović et al. 2008; Lega et al. 2010), the OFLI (Fouchard et al. 2002) and the OFLI_2^{TT} (Barrio 2005; Barrio et al. 2009; Barrio et al. 2010); the D (Voglis et al. 1999) and the SSNs (Voglis & Contopoulos 1994; Contopoulos & Voglis 1996; Contopoulos & Voglis 1997; Contopoulos et al. 1997; Voglis et al. 1998). Finally, we have also implemented the RLI (see Sándor et al. 2000; Széll et al. 2004; Sándor et al. 2004 and Sándor et al. 2007) and the APLE (Lukes-Gerakopoulos et al. 2008). The RLI is not based on the evolution of the solution of the first variational equations as the rest of the variational indicators implemented, but on the evolution of two different but very close orbits. The APLE is based on the concept of Tsallis Entropy.

The other widespread techniques devoted to chaos detection are the analysis of some particular quantities (e.g. the frequency) of a single orbit. The main contributions in the area of chaos detection is due to Binney & Spergel (1982) and Laskar (1990) (see also Laskar et al. 1992; Papaphilippou & Laskar 1996; Papaphilippou & Laskar 1998). The Frequency Modified Fourier Transform (FMFT) outlined by Sidlichovský & Nesvorný (1996) is another example of such kind of technique. The FMFT is the spectral analysis method selected for this investigation.

As previously mentioned, in M11 the authors compare the CIs implemented in an early version of the `LP-VIcode` on symplectic mappings. In D12, the authors use a later version of the `LP-VIcode` (where the CIs library was increased). Therefore, they do not only extend the work on mappings to a simple Hamiltonian flow: the Hénon & Heiles (1964) potential, but also they increase the number of CIs considered in the comparison. Both works deal with a comparative evaluation of the following CIs: the LI, the MEGNO, the SALI and the GALIs (GALI_k with $k = 2, 3, 4$), the FLI and the OFLI, the D and the SSNs and the RLI on symplectic mappings and a Hamiltonian flow. Finally, they suggest an efficient set of CIs (or CI's function which they call CIsF) composed by the pair FLI/OFLI, the MEGNO and the GALI_{2N} to study a general N–degree of freedom (d.o.f.) Hamiltonian system. In a work in progress, we use the latest version of the `LP-VIcode`, and do some experiments in a somehow realistic model of a triaxial stellar Hamiltonian system (Muzzio et al. 2005; Cincotta et al. 2008). The `LP-VIcode` latest version library of CIs includes all the CIs mentioned in the earlier versions of the code, plus the SElLCE, the OFLI_2^{TT} and the APLE. We extend the previous comparative studies of the CIs and find that the CIsF to study a general Hamiltonian system can be improved considering the pair FLI/OFLI, the pair MEGNO/SElLCE and the OFLI_2^{TT} or the GALI_{2N}. The pair FLI/OFLI and the pair MEGNO/SElLCE are recommended to study big samples of orbits by means of just computing their final values. The OFLI_2^{TT} or the GALI_{2N} are suggested to study small regions of very complex dynamics or regions dominated by strong chaos (Skokos et al. 2007; Skokos et al. 2008; D12), respectively. However, here we are going to test the CIs against a spectral analysis method on two regions of the stationary and the $x_0 - z_0$ start spaces

(Schwarzchild, 1993) of the self–consistent triaxial stellar Hamiltonian model previously mentioned. In order to do so, we use one of the recommended CIs to study big samples of orbits, i.e. the MEGNO/SElLCE, and the LI (both techniques already implemented in the `LP-VIcode`) and the FMFT as the selected spectral analysis method.

This paper is organized as follows: in Section 2 we present the code and explain its main features. In Section 3 we apply the `LP-VIcode` to study a realistic model. In order to investigate the advantages and drawbacks of the selected CIs included in this version of the `LP-VIcode` and the FMFT, we apply both types of chaos detection techniques to study the same space and compare the results in Section 4.

2.   The `LP-VIcode`

The `LP-VIcode` (in its alpha version) computes several CIs. It was coded in `FORTRAN 77`, although it is intended to be recoded in `FORTRAN 90` on a later version.

Although the current version of the code is in a developing stage, it has already implemented twelve indicators, already named in Section 1. The record is: the LI, the RLI, the SALI, the $GALI_k$, the MEGNO, the SElLCE, the FLI, the OFLI, the $OFLI_2^{TT}$, the D, the SSNs and the APLE.

2.1.   The CIs implemented in the `LP-VIcode`

The Lyapunov Indicator (LI).   Consider a continuous dynamical system defined on a differentiable manifold $\mathcal{S}$, where $\vec{\Phi}^t(\vec{x}) = \vec{x}(t)$ characterizes the state of the system at time t, $\vec{x}(0) = \vec{x}_0$ being the state of the system at time $t = 0$. Therefore, the state of the system after two consecutive time steps t and t' will be given by the composition law: $\vec{\Phi}^{t+t'} = \vec{\Phi}^t \circ \vec{\Phi}^{t'}$.

The tangent space of $\vec{x}$ maps onto the tangent space of $\vec{\Phi}^t(\vec{x})$ according to the operator $d_{\vec{x}}\vec{\Phi}^t$ and following the rule $\vec{w}(t) = d_{\vec{x}}\vec{\Phi}^t(\vec{w}(0))$ where $\vec{w}(0)$ is an initial deviation vector. The action of such operator at consecutive time intervals satisfies the equation:

$$d_{\vec{x}}\vec{\Phi}^{t+t'} = d_{\vec{\Phi}^{t'}(\vec{x})}\vec{\Phi}^t \circ d_{\vec{x}}\vec{\Phi}^{t'}.$$

If we suppose that our manifold $\mathcal{S}$ has some norm denoted by $\| \cdot \|$, we can define the quantity:

$$\lambda_t(\vec{x}) = \frac{\|d_{\vec{x}}\vec{\Phi}^t\vec{w}\|}{\|\vec{w}\|}$$

called "growth factor" in the direction of $\vec{w}$.

Consider an N-dimensional Hamiltonian $H(\vec{p}, \vec{q})$, with $\vec{p}, \vec{q} \in \mathbb{R}^N$, which we consider it autonomous just for the sake of simplicity. Let us remember that

$$\vec{x} = (\vec{p}, \vec{q}) \in \mathbb{R}^{2N}, \vec{f}(\vec{x}) = (-\partial H/\partial \vec{q}, \partial H/\partial \vec{p}) \in \mathbb{R}^{2N},$$

and then, the equations of motion are

$$\dot{\vec{x}} = \vec{f}(\vec{x}). \tag{1}$$

Let $\gamma(\vec{x_0}; t)$ be an arc of the orbit in the flux given by equation (1) over a compact energy surface: $M_h \subset \mathbb{R}^{2N}, M_h = \{\vec{x} : H(\vec{p}, \vec{q}) = h\}$ with h a constant, then

$$\gamma(\vec{x_0}; t) = \vec{x}(t'; \vec{x_0}) : \vec{x_0} \in M_h, 0 \le t' < t.$$

We define the LCE χ:

$$\chi[\gamma(\vec{x_0}; t)] = \lim_{t \to \infty} \frac{1}{t} \ln \lambda_t[\gamma(\vec{x}_0; t)], \tag{2}$$

and for its numerical implementation in the `LP-VIcode` we take the finite time limit of Eq. (2);

$$LI = \lim_{t \to T} \frac{1}{t} \ln \lambda_t[\gamma(\vec{x}_0; t)],$$

with T a finite time.

The Relative Lyapunov Indicator (RLI). If we graph the fluctuations of the LI, they are not significative. Thus, in order to amplify those fluctuations Sándor et al. (2004) define the quantity:

$$\Delta LI(\vec{x_0}; t) = \|LI(\vec{x_0} + \vec{\Delta x}; t) - LI(\vec{x_0}; t)\|,$$

where $\vec{x_0}$ and $\vec{x_0} + \vec{\Delta x_0}$ are two very close initial conditions at time t, separated by a quantity $|\vec{\Delta x}|$, which is a free parameter. Then, the RLI is defined through the expression:

$$RLI(t) = <\Delta LI(\vec{x_0}) >_t = \frac{1}{t} \sum_{i=1}^{t/\delta_t} \Delta LI(\vec{x_0}, i \times \delta_t), \tag{3}$$

with i the number of steps of stepsize δ_t. We use the expression (3) in the `LP-VIcode` in order to compute the RLI.

The Dynamical Spectra of the Stretching Numbers (SSNs) and the Spectral Distance (D). The local stretching number s_i is defined as:

$$s_i = \frac{1}{\delta_t} \ln \frac{|d_{\vec{x}} \vec{\Phi}^{t+i \times \delta_t}(\vec{\omega}(0))|}{|d_{\vec{x}} \vec{\Phi}^{t+(i-1) \times \delta_t}(\vec{\omega}(0))|}, \tag{4}$$

where $d_{\vec{x}} \vec{\Phi}^{t+i \times \delta_t}(\vec{\omega}(0)) = \vec{\omega}(t+i \times \delta_t)$ is the deviation vector at time $t+i \times \delta_t$.

Then, the SSNs are given by the density probability of the values s given by the s_i, i.e.

$$S(s)ds = \frac{dN(s)}{N},\tag{5}$$

where N is the total number of s_i and $dN(s)$ is the number of s_i in the interval $(s, s + ds)$. Thus, the computing of the SSNs in the LP-VIcode is just the construction of these histograms.

Then, the D is computed as the difference of two histograms of a given orbit. That is

$$D^2 = \sum_s [S_1(s) - S_2(s)]^2 \times \Delta s,\tag{6}$$

where $S_j(s)$ is the normalized number of s_i associated to the initial deviation vector $\vec{\omega}_j(0)$, which has values in the interval $s, s + \Delta s$.
The implementation of both the SSNs and the D in the LP-VIcode is based on the work of Voglis et al. (1999) and summarized by Eqs. (4), (5) and (6).

The Mean Exponential Growth Factor of Nearby Orbits (MEGNO) and the MEGNO's Slope Estimation of the largest Lyapunov Characteristic Exponent (SElLCE). The concept of local mean exponential rate of divergence of nearby orbits becomes evident when we rewrite the value of the LCE (Eq. 2) in an integral fashion:

$$\chi[\gamma(\vec{x}_0; t)] = \lim_{t \to \infty} \frac{1}{t} \int_0^t \frac{\|\dot{d}_\gamma \vec{\Phi}^{t'} \vec{w}\|}{\|d_\gamma \vec{\Phi}^{t'} \vec{w}\|} dt'.$$

Then, Cincotta & Simó (2000) defined the value Y as

$$Y[\gamma(\vec{x}_0; t)] = \frac{2}{t} \int_0^t \frac{\|\dot{d}_\gamma \vec{\Phi}^{t'} \vec{w}\|}{\|d_\gamma \vec{\Phi}^{t'} \vec{w}\|} t' dt'.$$

Finally, they introduced the MEGNO define as the average of Y, i.e.:

$$\overline{Y}[\gamma_q(\vec{x}_0)] \equiv \frac{1}{t} \int_0^t Y[\gamma_q(\vec{x}_0; t')] dt'.\tag{7}$$

Having the value of the MEGNO (Eq. (7)), Cincotta et al. (2003) suggest a linear behaviour to enclose the MEGNO's performances for regular and chaotic orbits,

$$\overline{Y}[\gamma(\vec{x}_0; t)] \approx a_\gamma t + b_\gamma,\tag{8}$$

where $a_\gamma = \chi_\gamma/2$ and $b_\gamma \approx 0$ for irregular, chaotic motion, while $a_\gamma = 0$ and $b_\gamma \approx 2$ for quasiperiodic motion.
The SElLCE takes the last 80% of the time series samplings and makes a linear least square fit, in order to estimate the value of χ through the MEGNO behaviour given by Eq. (8).

The LP-VIcode computes the MEGNO following the Eq. (7), and makes the least square fit to recover the Eq. (8) and compute the SElLCE.

The Smaller Alignment Index (SALI) and the Generalized Alignment Index (GALI). In Skokos (2001), the authors introduce the SALI in the following way: first they define the parallel and antiparallel index

$$d_- = \|\vec{\omega}_1 - \vec{\omega}_2\|, \qquad d_+ = \|\vec{\omega}_1 + \vec{\omega}_2\|, \tag{9}$$

respectively. Then, they define the SALI at a time t as the lowest of these two indexes:

$$SALI(t) = \min(d_+, d_-). \tag{10}$$

In Skokos et al. (2007), the SALI is generalized, introducing the GALI as

$$GALI_k(t) = \|\hat{w}_1(t) \wedge \hat{w}_2(t) \wedge \cdots \wedge \hat{w}_k(t)\|,$$

where $\hat{w}_i(t) = \frac{\vec{w}_i(t)}{\|\vec{w}_i(t)\|}, i = 1, 2, \ldots, k$ is the normalized deviation vector.

As the computing of the GALI is very expensive in computational terms, Skokos et al. (2008) introduced a variation for the numerical computation of the $GALI_k$ by making use of the singular value decomposition (SVD routine) of matrices and found that

$$\log(GALI_k) = \sum_{i=1}^{k} \log(z_i), \tag{11}$$

where the z_i are singular values of a given matrix Z.

The `LP-VIcode` computes the SALI following Eq. (10) and the GALI using the SVD routine of Numerical Recipes[1] to calculate the indicator through Eq. (11).

The Fast Lyapunov Indicator (FLI) and the Orthogonal Fast Lyapunov Indicator (OFLI). The FLI is a quantity closely related to the LI, which can distinguish between chaotic and regular motion (Froeschlé et al. 1997a; Froeschlé et al 1997b) and even between resonant and non-resonant motion (Froeschlé & Lega 2000; Lega & Froeschlé 2001; Guzzo et al. 2002) using (just) the first part of the numerical computing of the largest LCE.

For an N-dimensional system, in the `LP-VIcode` we follow the time evolution of the $2N$ deviation vectors and take the (euclidean) norm of each one. Then we record every K time steps, the largest of the norms, i.e. at time t the FLI is computed as

$$FLI(t) = \sup_t \left[\|\vec{w}_1(t)\|, \|\vec{w}_2(t)\|, \ldots, \|\vec{w}_{2N}(t)\| \right]. \tag{12}$$

As to the OFLI (see Fouchard et al. 2002), it is similar to the FLI, but in this case we take the orthogonal component to the flow of each deviation vector of the basis, time to time. Then it is defined as

---

[1]See Skokos et al. (2008) for further details.

$$OFLI(t) = \sup_t \left[w_1(t)^\perp, w_2(t)^\perp, \ldots, w_{2N}(t)^\perp \right]. \tag{13}$$

Notice that, although in the `LP-VIcode` we initially use the original definition of the FLI given in Froeschlé et al. 1997a, to define the OFLI, we later included the definition of the FLI given in Froeschlé & Lega 2000, where the authors use only one deviation vector. As they claim, the results do not vary, but the CPU-time is obviously reduced. Finally, the OFLI can be also computed using just one deviation vector, following the definition given in Fouchard et al. 2002.

The OFLI$^2_{TT}$. In Barrio (2005) the author defines the OFLI$^2_{TT}$, a second order variational indicator, as follows:

$$OFLI^2_{TT}(t) = \sup_{0 < t < t_f} \hat{w}(t)^\perp,$$

$\hat{w}(t)^\perp$ being the orthogonal component of the flow of $\hat{\vec{w}}(t)$, where $\hat{\vec{w}}(t)$

$$\hat{\vec{w}}(t) = \vec{w}(t) + \frac{1}{2}\vec{w}^{(2)}(t)$$

$\vec{w}(t)$ and $\vec{w}^{(2)}(t)$ being the solutions of the first and second order variational equations at time t, respectively.

Finally, in the `LP-VIcode` we take the superior $\hat{w}(t)^\perp$ in the interval $(0, t_f]$ for a given total time t_f. For further details, we refer to Barrio (2005), Barrio et al. (2009), and Barrio et al. (2010).

The Average Power Law Exponent (APLE). This method is based on the concept of Tsallis Entropy, thoroughly explained in Lukes-Gerakopoulos et al. (2008). Thus, here we limit ourselves to show the formula we use to compute the indicator in the `LP-VIcode`.

For an N–dimensional Hamiltonian, these authors consider a partitioning of the $2N$–dimensional phase space $\mathcal{S}$ into a large number of volume elements of size δ^{2N} for some small δ and let $\vec{x}(0)$ be the initial condition of an orbit located in a particular volume element. Thus, they introduce the APLE as follows:

$$APLE = \frac{\ln\left(\frac{|\vec{w}(t)|^2}{|\vec{w}(t_1)|^2}\right)}{2\ln\left(\frac{t}{t_1}\right)},$$

where $|\vec{w}(t)|^2 = \sum_{k=1}^m \|\vec{w}_k(t)\|^2$, and $\vec{w}_k(t)$ is one of the m deviation vectors of an orthogonal basis $\{\vec{w}_k(t)\}$ of the tangent space to $\mathcal{S}$ at the initial point $\vec{x}(0)$. Every $\vec{w}_k(t)$ has a length greater or equal to δ, and t_1 is a transient initial time of the evolution of the orbits.

2.2. The arrangement of CIs in units

As the dynamics of a given Hamiltonian should be analysed using different techniques, providing a variety of CIs in the `LP-VIcode` proves advantageous.

Nearly all the CIs mentioned in Section 2.1. are completely independent from each other and can be computed separately. The only exceptions are the SElLCE, which it strictly depends on the MEGNO values to do the least square fit, and the RLI (which is the difference between the LI of two close orbits). Nevertheless, the main drawback of computing them separately is the huge amount of CPU time required. An alternative to reduce such a time-consuming process is to arrange the CIs according to (the similarity in) their computation. That is, although nearly all indicators can be computed independently, some of them share some basic routines. For instance, both the FLI and the OFLI use the evolution of the same $2N$ deviation vectors, N being the number of degrees of freedom of the system[2]. Therefore, we decided to group the CIs in units the indicators sharing part of their computing processes.

This grouping is ordered as follows: the RLI is implemented in the same unit as the LI because it is the difference between the LI for two close initial conditions. Furthermore, the SALI is implemented in the same unit as the LI and the RLI, since it uses the evolution of the length of 2 deviation vectors. This is done in order to share the routine that computes the evolution of the deviation vectors which uses a renormalization process. Because the SElLCE (as mentioned before) uses the MEGNO to estimate the LI of the orbit, both CIs, the MEGNO and the SElLCE belong to the same unit. The FLI, the OFLI and the APLE can be computed using the same deviation vectors ($2N$ or just one, depending on the definition). Moreover, all of them use a routine that computes the evolution of the deviation vectors without the renormalization process previously used with the LI, RLI or SALI. The SSNs are basically built on the computation of histograms, and the D uses the difference of the SSNs for two different deviation vectors of a given orbit. Then, both of them are included in the same unit. The $GALI_k$ is computed in a different unit, due to the fact that it is the only CI using the SVD routine (see Section 2.1.). The $OFLI_2^{TT}$ is in a separate unit because it is the only CI that needs the computation of the second order variational equations, which requires the evaluation of third order derivatives.

Finally, the CIs implemented so far are arranged in the `LP-VIcode` as follows:

- *Unit 1:* LI, RLI and SALI

- *Unit 2:* MEGNO and SElLCE

- *Unit 3:* FLI, OFLI and APLE

- *Unit 4:* SSN and D

- *Unit 5:* GALI$_k$

- *Unit 6:* OFLI$_2^{TT}$

---

[2]According to the orginal definition given in Froeschlé & Lega (1997a), they can use the same deviation vector, according to the actual definition given in Froeschlé & Lega (2000), see Section 2.1. for further details.

2.3. The input files

The `LP-VIcode` needs two input files in order to work. One of these files is a parameter file, in which all the information about the calibration of the indicators as well as the format of the output files is introduced. The other one is an input data file, which indicates which orbits we would have to compute.

The parameter file. Nearly all parameters can be set from this parameter file and only a few are still remain in the main program (mainly related with the specific problem). The main structure of this file consists of two parts. The first one arranges the input and output filenames and the second one arranges the parameters themselves. These parameters are the following:

- Physical parameter: the energy of the system (in case the user does not specify all the initial condition coordinates of the phase space).

- RLI parameter: the initial separation of both orbits (see Section 2.1).

- Output parameter: a binary value to set which output is preferred in the computation, i.e. "0" only the final value of the CI (i.e. the value of the indicator at the end of the computing process) and "1" the time evolution of the CI.

- Trajectory parameter: a binary value to print the phase space coordinates of the orbit ("1"), or skip the time-consuming writing process ("0").

- CIs' selection parameter: a set of integer values to specify which units or CIs are to be computed.

- Formatting parameters.

The CIs' selection parameter is one of the key parameters of the code, because it allows us to compute several CIs (with the efficient grouping mentioned before) to reduce the CPU time (see Section 2.2.), or to compute them separately. The parameter is a horizontal array of 6 integers (one for each unit) which indicates the program if a given unit should be computed (value set equal to "1") or not (value set equal to "0"). In the case of the second unit (the MEGNO and the SElLCE), "1" is adopted so as to compute the MEGNO alone and "2" to compute the MEGNO and the SElLCE together.

The data file. The data file has a very simple format. In the first commented line, the order in which the program will read the data values is specified, i.e. the cartesian coordinates, the conjugate momentums and, the total integration time.

2.4. The Integrator

There are a lot of Ordinary Differential Equations (ODEs) integrators which can be implemented in the `LP-VIcode`, and the independency from the integrator routine is part of a future implementation. On the other hand, all the indicators already implemented in the code must integrate not only the equations of motion

but also the first (and second) variational equations. Therefore, we need to count on an efficient integrator for these tasks. For instance, according to D12, a suitable integrator routine is the Prince & Dormand implementation of a Runge-Kutta method of order $7-8$ called DOPRI8 (for more information see Prince & Dormand 1981). Thus, the DOPRI8 routine is the ODEs integrator selected for the current version of the `LP-VIcode`.

3.   Applications

In order to compare both types of chaos detection techniques (i.e. the variational and the frequency-based ones), we apply the MEGNO/SElLCE and the LI as representatives of the variational indicators implemented in the `LP-VIcode` and the FMFT, which is the selected spectral analysis method. Thus, in this section we are going to use the `LP-VIcode` and a spectral analysis method as complementary tools to study two regions of the stationary and the $x_0 - z_0$ start spaces (Schwarzschild, 1993) of the model introduced in Muzzio et al. (2005) which will be briefly described in the next section.

3.1.   The potential

The self–consistent triaxial Hamiltonian model of an elliptical galaxy is obtained after the virialization of an N–body self–consistent system composed of 10^5 particles (Muzzio et al. 2005). The model reproduces many dynamical characteristics of real elliptical galaxies, such as mass distribution, flattening, triaxiality and rotation (Muzzio 2006). Therefore, it seems to provide a useful realistic scenario to apply the `LP-VIcode` and the FMFT as well.

The equation that reproduces the potential is

$$V(x,y,z) = -f_0(x,y,z) - f_x(x,y,z) \cdot (x^2 - y^2) - f_z(x,y,z) \cdot (z^2 - y^2),$$

where

$$f_n(x,y,z) = \frac{\alpha_n}{\left[p_n^{a_n} + \delta_n^{a_n}\right]^{\frac{ac_n}{a_n}}}, \tag{14}$$

α_n, δ_n, a_n, ac_n are constants and p_n^2 is the square of the softened radius given by $p_n^2 = x^2 + y^2 + z^2 + \varepsilon^2$ when $n = 0$, or $p_n^2 = x^2 + y^2 + z^2 + 2 \cdot \varepsilon^2$ for $n = x, z$.

The adopted value for the softening parameter is $\varepsilon \simeq 0.01$ for any n. The functions $f_n(x,y,z)$ were computed through a quadrupolar N–body code for 10^5 particles, which allowed the authors to write them in a general fashion given by Eq. (14). The adopted values for the constants α_n, δ_n, a_n and ac_n are given in Table 1. For further references, see Muzzio et al. (2005) and Cincotta et al. (2008).

The stationary character of the parameters given in Table 1 were tested by performing several fits at different times after virialization, resulting in a precision of 0.1%.

After the system had relaxed, there remained 86818 particles resembling an elliptical galaxy (the system obeying a de Vaucouleurs' law, as shown in Fig. 2

Table 1.   Adopted values for the coefficients of the functions f_n given by Eq. (14).

	α	a	δ	ac
$n = 0$	0.92012657	1.15	0.1340	1.03766579
$n = x$	0.08526504	0.97	0.1283	4.61571581
$n = z$	−0.05871011	1.05	0.1239	4.42030943

in Muzzio et al. 2005) with a strong triaxiality and a flattening that increases from the border of the system to its center (see Table I in the same paper).

The obtained triaxial potential has semi–axes X, Y, Z satisfying the condition $X > Y > Z$, and its minimum, which is close to −7, matches the origin. As expected, the potential is less flattened than the mass distribution (see Table I in Muzzio et al. 2005).

As we mentioned at the beginning of this section, the potential seems to provide a useful realistic scenario to test the `LP-VIcode` and the FMFT. Thus, in the next subsection, we describe how we proceed (in order) to compare both techniques as chaos detection tools.

3.2.   Comparative evaluation of the FMFT and the SElLCE as global chaos detection techniques

Preliminaries.   Herein we apply a spectral analysis method, the FMFT, and a CI, the SElLCE (one of the indicators in the library of the `LP-VIcode`) to two regions on the energy surface −0.7 of the potential described in Section 3.1.. In order to compare both techniques as chaos detection tools we apply the FMFT and the SElLCE to a few samples of initial conditions in the stationary space and in the $x_0 - z_0$ start space of the self-consistent triaxial stellar model.

There are several ways to compare chaos detection tools. Our choice is to determine which technique offers the most detailed phase space portrait using the same integration time. Therefore, we must first determine an integration time by which the techniques (at least for most of the initial conditions of the samples) are out of a transient regime; otherwise, we will obtain unreliable phase space portraits.

We will consider a time of 10^3 characteristic times[3] to keep the LI out of the transient interval, as the authors did in Maffione et al. (2011a). That is, a convergent LI is the criterion used (in order) to yield reliable values of the CIs, in particular of the MEGNO/SElLCE indicators. From Maffione et al. (2011a) we know that for the energy surface −0.7, the characteristic time is ~ 7 u.t. So, (in order) to obtain reliable values for the CIs previously mentioned, the integration time must be 7×10^3 u.t.

---

[3]We approximate this time-scale as the period of the axial orbit on the semi-major axis X of the model.

The FMFT indicates regular motion when the frequencies do not change in time, i.e. the orbit is confined within a torus, which is well defined by a set of N frequencies (where N is the number of the d.o.f. of the system). However, if the orbit is chaotic, there is a variation in the frequencies. This lack of constancy in time means that the orbit is not confined within a torus[4]. The precision on the computation of the frequencies is a key parameter, because the FMFT might not show a variation of the frequencies in time and thus, it might not be able to distinguish chaotic from regular orbits. Then, to determine such a precision and the efficiency in describing the phase space portraits of the triaxial model for the FMFT, we use the same final integration time used with the SElLCE, i.e. 10^3 periods, which is enough to stabilize the LI for most of the initial conditions of the samples.

Finally, the equations of motion and their first variationals are integrated for a final integration time of 7×10^3 u.t. in the case of the SElLCE. For the FMFT, we compare the computation of the fundamental frequencies in two 50% (Wachlin & Ferraz-Mello 1998) overlapping time intervals, (in order) to estimate possible variations in the frequencies. The first interval goes from 0 u.t. to 7×10^3 u.t., and the second one, from 3.5×10^3 u.t. to 1.05×10^4 u.t.

We apply the SElLCE and the FMFT to 624100 orbits in the region of the stationary space and to 596258 orbits in the region of the $x_0 - z_0$ start space.

The integration of the equations of motion, which are necessary to compute the frequencies with the FMFT, was carried out with the `taylor` package (Jorba & Zou 2005), which proved to be a very convenient tool for the model under analysis (see D12). The precision required for the phase space coordinates was of 10^{-15}.

On the other hand, the integrations for the LI and the MEGNO/SElLCE were carried out with the DOPRI8 routine (see Section 2.4.), which it is more efficient than `taylor` in the case of the simultaneous integration of both the equations of motion and their variational equations for the self-consistent triaxial stellar model (we refer to D12 for further details). The energy preservation with DOPRI8 was of the order of $\sim 10^{-13}, 10^{-14}$.

The following configuration was used for all the computations included in this paper: *a)* Hardware: CPU, 2 x Dual XEON 5450, Dual Core 3.00GHz; M.B., Intel S5000VSA; RAM, 4GB(4x1GB), Kingston DDR–2, 667MHz, Dual Channel. *b)* Software: gfortran 4.2.3.

The experiment. In order to use the FMFT as a chaos detection tool, we compute the quantity $\log(\Delta F)$ (Wachlin & Ferraz-Mello 1998). The $\log(\Delta F)$ is defined as $\Delta F \equiv |\nu_x^{(1)} - \nu_x^{(2)}| + |\nu_y^{(1)} - \nu_y^{(2)}| + |\nu_z^{(1)} - \nu_z^{(2)}|$, where $\nu_j^{(i)}$ is the fundamental frequency computed with the FMFT and associated with the degree of freedom j ($j = x, y, z$) for the interval (i), with $i = 1, 2$ (the two overlapping time intervals). Besides, we must have all the fundamental frequencies computed for every orbit on both intervals and this is not the general case for every orbit.

---

[4]For further details on the FMFT, refer to Sidlichovský & Nesvorný (1996). Herein, we simply describe how the indicator distinguishes between chaotic and regular motion, because it is needed (in order) to compare its performance with the variational tool, the SElLCE.

Thus, the phase space portraits of the $\log(\Delta F)$ finally consist of 622521 orbits on the stationary space and 594690 orbits on the $x_0 - z_0$ start space.

In Fig. 1 we present the SElLCE (left panels) and the $\log(\Delta F)$ (right panels) values for the region on the stationary space (top panels) and for the region on the $x_0 - z_0$ start space (bottom panels) of the triaxial model under analysis.
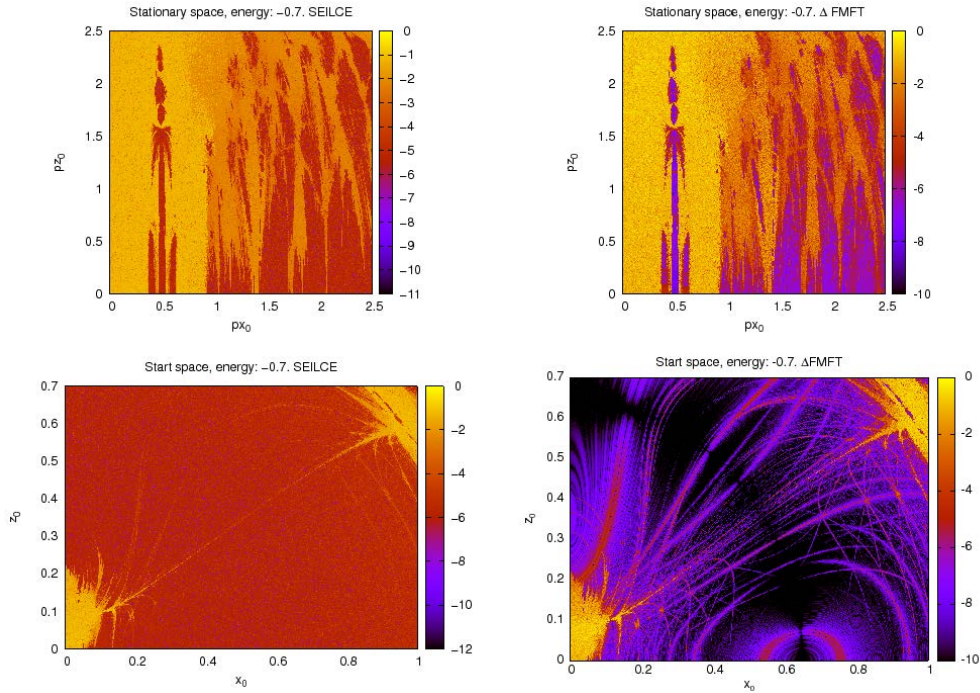

Figure 1. Phase space portraits of the stationary space with 624100 initial conditions (top left panel) and of the $x_0 - z_0$ start space with 596258 initial conditions (bottom left panel), using the values of the SElLCE integrated for 7×10^3 u.t. Right panels, idem but with 622521 and 594690 initial conditions (top and bottom right panels, respectively), using the $\log(\Delta F)$ integrated on two overlapping time intervals of 7×10^3 u.t. each. The values of the SElLCE and the $\log(\Delta F)$ are in logarithmic scale.

Although the SElLCE and the FMFT show similar results on the stationary space (top panels of Fig. 1), the latter includes a high amount of spurious structures[5] on the $x_0 - z_0$ start space (bottom right panel of Fig. 1). This spurious structures jeopardize the choice of a threshold value in order to identify regular and chaotic orbits due to an unclear separation of the different kind of motions. On the contrary, with a variational indicator as the SElLCE (bottom left panel of Fig. 1), this classification into regular and chaotic motion seems to

---

[5]Some of them due to the Moiré phenomenon, which is common with methods using the discrete Fourier transform (Barrio et al. 2009).

be more natural and thus, more efficient. Nevertheless, if in the $x_0 - z_0$ start space we take as chaotic orbits those which preserve 4 decimal digits or fewer in their computed fundamental frequencies with the FMFT, we recover the phase space portrait obtained by the SElLCE.

This kind of results, where the distinction between chaotic and regular motion is not as clear as the one given by variational indicators like the SElLCE, makes the FMFT a less reliable indicator when we study the global dynamics of a divided phase space.

The process used to determine the chaoticity or regularity of the orbits by means of the FMFT is standard. Then, the somehow inaccurate descriptions of the portraits of divided phase spaces might be basically due to a high sensitivity of the method with its parameters.

As regards the computing times, the SElLCE (one of the fastest CIs, together with the FLI and the MEGNO), took ~ 670 hs for an integration time of 7×10^3 u.t. and for 624100 orbits on the stationary space. For the 594690 orbits on the $x_0 - z_0$ start space, the CI took ~ 330 hs.

Although the computing of the fundamental frequencies with the FMFT is quite fast, the determination of the $\log(\Delta F)$ is time consuming. For instance, for the generation of the right panels of Fig. 1, two 50% overlapping time intervals of 7×10^3 u.t. each were necessary. In other words, the integration of the equations of motion was performed for a total time interval of 1.05×10^4 u.t. in order to have the frequencies computed after 10^3 periods in both intervals, the same amount of periods used with the SElLCE. Finally, the time taken by the $\log(\Delta F)$ was ~ 885 hs. for the stationary space and ~ 450 hs. for the $x_0 - z_0$ start space. Therefore, the computing speed of the fundamental frequencies by the FMFT is lost against the whole process involved in the determination of the variation of the frequencies with the $\log(\Delta F)$. In fact, the computing of the $\log(\Delta F)$ is necessary to distinguish between regular and chaotic orbits with the FMFT and thus, the FMFT as a global chaos indicator turns out to be slower than the SElLCE (and other similar fast CIs).

In the next section we consider the FMFT as an appropriate method for the determination of the fundamental frequencies only of the regular orbits as well as an efficient tool to determine the resonant map of the system. Furthermore, we use the FMFT as an complement to the `LP-VIcode` in order to depict the global picture of the stationary phase space.

4. Complementary use of the `LP-VIcode` and the FMFT

The analysis by means of the LI, the MEGNO/SElLCE and the FMFT of the stationary space of the triaxial potential of Muzzio et al. (2005) for different energy surfaces gives us enough information to shortly discuss the advantages of using both types of techniques together, i.e. CIs (within the `LP-VIcode`) and spectral analysis methods (with the FMFT).

4.1. The contribution of the `LP-VIcode`

We consider samples of 1000444 initial conditions for the energy surfaces defined by the constant values -0.1 and -0.7; the integration times are of 1.17×10^5 u.t. (for the energy surface -0.1 the period of the semi-major axis orbit is ~ 117

u.t.) and 7×10^3 (see Section 3.2.), respectively . As the computing times become critical in the experiment, we select, from the `LP-VIcode`, CIs of low computational cost: the LI and the MEGNO/SElLCE (the FLI/OFLI could be an equal efficient alternative).

The corresponding phase space portraits are presented in Fig. 2, left panel for the energy surface -0.1 and right panel for the energy surface -0.7.
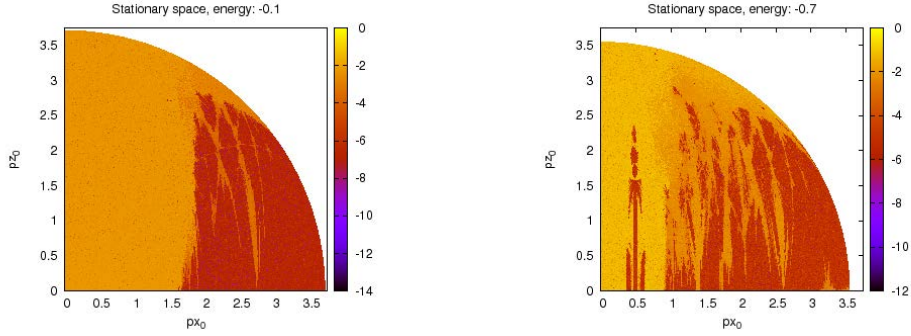

Figure 2.   Phase space portraits by means of the SElLCE for the stationary space of the triaxial model, integrating 1000444 orbits for 10^3 periods on two energy surfaces. On the left panel, for the energy surface -0.1 within a time interval of 1.17×10^5 u.t and on the right panel, for the energy surface -0.7 within a time interval of 7×10^3 u.t. The values of the SElLCE are in logarithmic scale.

The SElLCE shows a very good performance in describing the phase space portraits corresponding to a strongly divided phase space like the present one. Nevertheless, as it has not a natural way to determine a threshold value to distinguish chaotic from regular orbits, we have to estimate it (in order) to study the phase space portraits presented by the SElLCE in Fig. 2.

To determine such threshold for the SElLCE, we can calibrate the CI in order to obtain similar percentages of chaotic and regular orbits than those obtained with other confident CI with a defined threshold. The MEGNO is the first alternative due to the fact that the SElLCE needs its computation. However, as the MEGNO shows a high sensitivity with its asymptotically theoretical threshold (see e.g. M11; D12), it is not reliable to be used to calibrate other indicators. Therefore, we use other indicator already implemented in the `LP-VIcode`.

As aforementioned, the computing time is a key variable in the experiment, and thus, the LI is the CI of least computational cost given a fixed total integration time. Moreover, the indicator has a theoretical threshold value to start with: $\ln(T)/T$, with T being the total integration time. Starting with the theoretical approximation of the threshold, we calibrate it by inspection and find appropriate threshold values for the LI for both energy surfaces. Finally, the threshold of the SElLCE is estimated by an iterative process, which is stopped when the percentage of chaotic orbits best approximates the percentage of the chaotic component yielded by the LI.

In Table 2 we present, from left to right, the energy surface, the threshold value (V_c) estimated for the LI, the corresponding percentage of chaotic orbits,

the estimated threshold value for the SElLCE and the corresponding percentage of chaotic orbits.

Table 2.   For both energy surfaces considered on the stationary phase space (i.e. -0.1 and -0.7): the threshold value used for the LI, the percentage of the chaotic component given by the LI, the estimated threshold value for the SElLCE and the corresponding percentage of the chaotic component are detailed.

Energy	V_c (LI)	Chaos (%) -LI	V_c (SElLCE)	Chaos (%) -SElLCE
-0.1	1.2×10^{-4}	$\sim 71.46\%$	1.4×10^{-5}	$\sim 66.83\%$
-0.7	1.7×10^{-3}	$\sim 65.09\%$	2.7×10^{-4}	$\sim 62.87\%$

Independently of the energy surface considered, the chaotic component dominates the phase space portraits (columns 3 and 5 of Table 2). However, as we move to more negative energy surfaces (-0.7), the regular component increases. The variation is not important along the energies considered, though.

On the left panel of Fig. 2, we observe that the chaotic and regular components are almost separated. On the one hand, we have the chaotic component fully connected for values of $p_{x_0} \lesssim 1.7$, and the region of regular orbits for values of $p_{x_0} \gtrsim 1.7$, except for some structures which arise from the border of the energy surface and enter the regular component. Furthermore, these structures multiply themselves as we go to more negative energies (more bonded regions of the potential). These structures are resonances that overlap with each other, and start to populate the regular component. We can also observe a division inside the chaotic component, where the connected chaotic domain move back to lower values of p_{x_0}, giving place to another chaotic domain characterized by a regime of resonance overlap and by a lower Lyapunov exponent (notice the different colours). On the right panel of Fig 2, for an energy surface of -0.7, the resonances fill the regular component; strong resonances in the chaotic domains are also observed. The most remarkable is the one which lies around $p_{x_0} \sim 0.5$.

Given a global portrait as the one shown by means of the SElLCE, with a variational indicator we can visualize many phenomena such as how the chaotic and regular components interact, where the resonances appear and how they overlap to generate chaotic regions. However, we should make a great effort with the CIs if we need more detailed information because we need the resonant map to understand many of the causes of such phenomena.

In order to obtain the resonant map, the time evolution of the CIs which can provide information about the dimensionality of the torus on which the regular orbits lie (like the GALIs) can be analyzed, and thus, infer the resonances to which they belong. There is another way to identify the periodic orbits (e.g. using the OFLI) and analyze their stability (as done in Cincotta et al. 2008 with the MEGNO) in order to search for orbital families generated by perturbations to such parent periodic orbits in nearby regions. However, this process could be very slow because the CIs are not the best suited for the task.

It remains to apply the method of spectral analysis, i.e. the FMFT, to improve the study of the regular component with the associated resonant map,

and thus, complement the information given by the SElLCE (also assisted by the LI).

4.2. The contribution of the FMFT

We use the FMFT to compute the frequencies of a representative sample of regular orbits and determine the resonant map of the stationary space studied in Section 4.1. by means of the SElLCE.

We first need to identify the sample of regular orbits to apply the FMFT. Such a sample consists of orbits that are simultaneously classified as regular orbits by the MEGNO and the LI. We call this sample "A".

Having the representative sample "A" of regular orbits, we integrate the equations of motion for 3×10^2 characteristic times (which gives us enough precision) in order to compute the frequencies with the FMFT. We are not able to compute the three fundamental frequencies for all the orbits of sample "A" but for most of them. We call this reduced sample of regular orbits with the three fundamental frequencies computed sample "B". Finally, we use sample "B" to compute the commensurabilities and identify the resonance web.

We consider as resonant orbits those whose resonant vector $\vec{m} \in \mathbb{Z} - \{\vec{0}\}$ satisfies the relation: $\vec{m} \cdot \vec{\nu} < 10^{-6}$ with $\vec{\nu}$ being the frequency vector. The value 10^{-6} is an estimated value according to the best fit between the resonance web and the description of the phase space previously given by the CIs (Section 4.1.). We separate the resonances according to the d.o.f. involved, i.e. into resonances between 2 and 3 d.o.f. In the case of the resonances between 2 d.o.f., we only searched for those of highest order, i.e. iterating until 2×10^2 on each d.o.f. Lastly, we sorted them by their resonant vector's absolute value which yields information about the width and importance of the resonance (Reichl 2004).

In Table 3 we show, for each energy surface and a total of 1000444 initial conditions, the number of orbits in the representative sample "A", the total time used with the FMFT to compute the fundamental frequencies, the number of orbits in the reduced sample "B" and the number of orbits in resonance (with 2 or 3 d.o.f. involved).

Table 3. For each energy surface (i.e. -0.1 and -0.7) on the stationary space, the following information is provided: the number of orbits in the representative sample "A", the integration time used by the FMFT to determine the fundamental frequencies, the number of orbits in the reduced sample "B" and lastly, the number of orbits in resonance between 2 or 3 d.o.f.

Energy	Sample "A"	Time interval	Sample "B"	In resonance
-0.1	7.2426×10^4	3.51×10^4	7.2295×10^4	1.315×10^3
-0.7	7.2948×10^4	2.1×10^3	7.2781×10^4	1.773×10^3

Notice in Table 3 that the percentage of orbits in resonances is not high and is similar for both surfaces: $\sim 1.82\%$ for -0.1 and $\sim 2.43\%$ for -0.7 in the stationary space.

In Figure 3 we show the resonant maps corresponding to the energy surfaces -0.1 (left panel) and -0.7 (right panel) of the stationary space of the
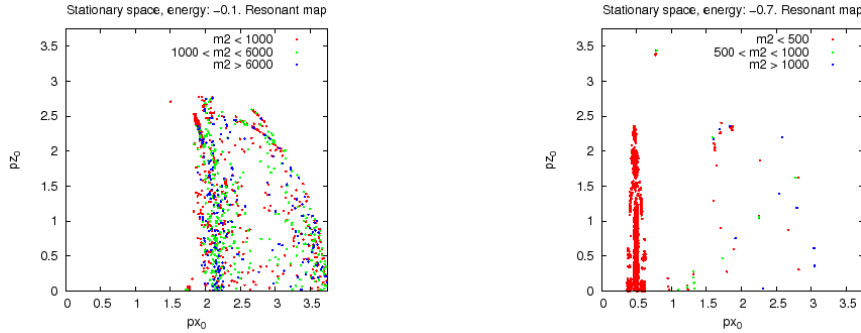
Figure 3.   Resonant maps for both energy surfaces on the stationary space. We sort the resonances by the value $m2 = |\vec{m}|^2$. Left panel for the energy surface -0.1 and right panel for the energy surface -0.7.

self-consistent triaxial stellar model of elliptical galaxy under study. The resonant map corresponding to the energy surface -0.1 shows populated regions of resonant orbits near the chaotic component and close to the border of the corresponding energy surface. However, the resonant orbits are of very low order (i.e. $|\vec{m}|^2 \gtrsim 10^3$). On the other hand, on the right panel of Figure 3, we show the results for the energy surface -0.7 and we observe a highly compact resonance of high order in the chaotic domain. In order to identify such a resonace we compute the rotational numbers with the fundamental frequencies given by the FMFT and find that the resonance is between 2 d.o.f., the $4 : 3$ (x:y) resonance.

    As we can see from the experiment, the variational indicators such as the LI and the couple MEGNO/SElLCE (computed with the `LP-VIcode`) and the spectral analysis methods such as the FMFT work remarkably well as complementary methods.

5.   Discussion

From all the studies carried on in this paper, we might conclude that having a large indicators diversity is essential to have a precise description of a dynamical system. Thus, in the first part of this work we presented the alpha version of the `LP-VIcode`. A code that is on a developing stage but which has already proved its value. The purpose of the `LP-VIcode` is to efficiently arrange together a great variety of CIs in order to have at hand several dynamical tools to study a given dynamical system. The arrangement has not considered spectral analysis methods yet, since the original idea was to reduce the CPU time when computing several variational indicators. Nevertheless, this is not discarded for future implementations.

    The CIs included and full functioning in this alpha version of the `LP-VIcode` are the following: the LI, the RLI, the SALI, the $GALI_k$, the MEGNO, the SElLCE, the FLI, the OFLI, the $OFLI_2^{TT}$, the D, the SSNs and the APLE (Section 2.1.). They can be computed separately or within units, i.e., in order to reduce the CPU time economizing similar processes they have in their comput-

ing routines (Section 2.2.).

In the second part of this work, we use the availability of the CIs given by the `LP-VIcode` to compare different sort of tools for dynamical analysis. On the one hand, we have the CIs which are based on the concept of local exponential divergence and follow the evolution of the deviation vectors. On the other hand, we have the spectral analysis methods, which require the integration of the equations of motion to compute, e.g., the frequencies of regular orbits.

According to previous papers such as Barrio et al. (2009), we find that the complementary use of such techniques, the CIs and the spectral analysis methods, is a very efficient way to gather dynamical information (Section 4.).

Here, we show that the SElLCE (a CI) works better than the FMFT (a spectral analysis method) as a global chaos detection tool (Section 3.2.) to describe the divided phase space of the self-consistent triaxial stellar dynamical model resembling and elliptical galaxy (Section 4.1.). The fundamental frequencies of the regular orbits easily provided by the FMFT allow a fast building of the resonance web and thus, a quick understanding of many phenomena described in the phase space portraits given by the CI (Section 4.2.).

Finally, in view of the present succesful applications of the `LP-VIcode`, there are still many improvements to be made to the code. Among the main goals, we can mention the following:

- To continue increasing the record of CIs in the library of the code.

- To incorporate routines to compute diffusion rates, in order to take advantage of the many computations done by the code.

- To recode it in `FORTRAN 90`.

- To make the code independent of the integrator routine.

- To make the code independent of the model, using symbolic manipulation programs to decode the differential equations and implement them automatically.

Our aim is to release a stable version of the `LP-VIcode` with all those goals implemented and offer the code to public domain, so that the interested community may collaborate including their own chaos detection tools and/or improve the ones already implemented.

The alpha version of the `LP-VIcode` is available directly by the authors upon request.

References

Antonopoulos, Ch., Vasileios, B. & Bountis, T. 2010, Phys. Rev. E., 81, 016211

Barrio, R. 2005, Chaos, Solitons and Fractals, 25, 711-726

Barrio, R., Blesa, F. & Serrano, S. 2009, Physica D, 238, 1087-1100

Barrio, R., Blesa, F. & Serrano, S. 2010, Int. J. Bifurcations and Chaos, 20, 1293-1319

Benettin, G., Galgani, L. & Strelcyn, J. 1976, Phys. Rev. A, 14(6), 2338-2345

Benettin, G., Galgani, L., Giorgilli, A. & Strelcyn, J. 1980, Meccanica 15 Part I, 9-20; Part II, 21-30

Binney, J. & Spergel, D. 1982, Ap. J., 252, 308-321

Bountis, T.C. & Skokos, Ch. 2006, 561, 173-179

Carpintero, D. 2008, MNRAS, 388, 1293-1304

Cincotta, P. & Simó, C. 2000, Astron. Astrophys, 147, 205-228

Cincotta, P., Giordano, C. & Simó, C. 2003, Physica D., 182, 151-178

Cincotta, P., Giordano, C. & Muzzio, J.C. 2008, Discrete and Continuous Dynamical Systems B., 10, 439-454

Compère, A., Lemaĩ®tre, A. & Delsate, N. 2011, CeMDA, 112, 75-98

Contopoulos, G. & Voglis, N. 1996, CeMDA, 64, 1-20

Contopoulos, G. & Voglis, N. 1996, Astron. & Astrophys, 64, 1-20

Contopoulos, G., Voglis, N., Efthymiopoulos, Ch., Froeschlé, Cl., Gonczi, R., Lega, E., Dvorak, R. & Lohinger, E. 1997, CeMDA, 67, 293-317

Darriba, L., Maffione, N., Cincotta, P. & Giordano, C. (accepted for publication), IJBC (D12)

Fouchard, M., Lega, E., Froeschlé, Ch. & Froeschlé, Cl 2002, 83, 205-222

Froeschlé, Cl. 1984, CeMDA, 34, 65-115

Froeschlé, Cl., Gonczi, R. & , Lega, E. 1997a, Planet. Space Sci., 45, 881-886

Froeschlé, Cl., Lega, E. & Gonczi, R. 1997b, CeMDA, 67, 41-62

Froeschlé, Cl. & , Lega, E. 1998, Astron. Astrophys., 334, 355-362

Froeschlé, Cl. & , Lega, E. 2000, CeMDA, 78, 167-195

Froeschlé, Cl. & Lega, E. 2006, B.A. Steves et al (eds) Chaotic Worlds: from Order to Disorder in Gravitational N-Body Dynamical Systems, 131-165

Froeschlé, Cl., Lega, E. & Guzzo, M. 2006, CeMDA, 95, 141-153

Gayon, J. & Bois, E. 2008, Astron. & Astrophys., 423, 745-753

Giordano, C. & Cincotta, P. 2004, Astron. & Astrophys., 423, 745-753

Goździewsku, K., Konacki, M. & Meciejewski, A. 2005, Ap. J., 619, 1084-1097

Guzzo, M., Lega, E. & Froeschlé, Cl. 2002, Physica D., 163, 1-25

Hénon, M. & Heiles, C. 1964, Astronom. J., 1, 73-79

Hinse, T., Christou, A., Alvarellosn, J. & Gózdziewski, K. 2010, MNRAS, 404, 837-857

Jorba, À. & Zou, M. 2005, Experimental Mathematics, 14, 99-117

Kaneko, K. & Konishi, T. 1994, Physica D., 71, 146-167

Laskar, J. 1990, Icarus, 88, 266-291

Laskar, J., Froeschlé, Cl. & Celetti, A. 1992, Physica D, 56, 253-269

Lega, E. & Froeschlé 2001, CeMDA, 81, 129-147

Lega, E., Guzzo, M. & Froeschlé, Cl. 2010, CeMDA, 107, 115-127

Lemaître, A., Delsate, N. & Valk, S. 2009, CeMDA, 104, 383-402

Lukes-Gerakopoulos, G., Voglis, N. & Efthymiopoulos, Ch. 2008, Physica A, 387, 1907-1925

Maffione, N., Giordano, C. & Cincotta, P. 2011a, Int. J. of Nonlinear Mech., 46, 23-34

Maffione, N., Darriba, L., Cincotta, P. & Giordano, C. 2011b, CeMDA, 111, 285-307. (M11)

Manos, T. & Athanassoula, E. 2011, MNRAS, 415, 629-642

Muzzio, J. 2006, CeMDA, 96, 85-97

Muzzio, J., Navone, H. & Zorzi, A. 2009, CeMDA 105, 379-395

Paleari, S., Froeschlé, Cl. & Lega, E. 2008, CeMDA, 102, 241-254

Papaphilippou, Y. & Laskar, J. 1996, Astron. & Astrophys., 307, 427-449

Papaphilippou, Y. & Laskar, J. 1996, Astron. & Astrophys., 329, 451-481

Prince, P. & Dormand, J. 1981, J. Comput. Appl. Math., 35, 67-75

Reichl, L. 2004, The transition to Chaos. Conservative classical systems and quantum manifestations. Second Edition. Springer-Verlag

Sándor, Z., Érdi, B. & Efthyimiopoulos, C. 2000, CeMDA, 78, 113-123

Sándor, Z., Érdi, B., Széll, A. & Funk, B. 2004, CeMDA, 90, 127-138

Sándor, Z., Süli, Á., Érdi, B., Pilat-Lohinger, E. & Dvorak, R. 2007, MNRAS, 375, 1495-1502

Schwarzschild, M. 1993, Ap. J., 409, 563-577

Sidlichovský, M. & Nesvorný, D. 1996, CeMDA, 65, 137-148

Skokos, Ch. 2001, J Phys A-Math Gen., 34, 10029-10043

Skokos, Ch. 2010, Lect. Notes Phys., 790, 63-135

Skokos, Ch., Bountis, T. & Antonopoulos, Ch. 2007, Physica D., 231, 30-54

Skokos, Ch., Bountis, T. & Antonopoulos, Ch. 2008, The European Physical J. Special Topics, 165, 5-14

Széll, Érdi, B., Sándor, Z. & Steves, B. 2004, MNRAS, 347, 380-388

Tancredi, G., Sánchez, A. & Roig, F. 2001, Astronom. J., 121, 1171-1179

Todorović, N., Lega, E. & Froeschlé, Cl. 2008, CeMDA, 102, 13-27

Voglis, N. & Contopoulos, G. 1994, J. Phys. A: Math. Gen., 27, 4899-4909

Voglis, N., Contopoulos, G. & Efthymiopoulos, Ch. 1998, Phys. Rev. E., 57, 372-377

Voglis, N., Contopoulos, G. & Efthymiopoulos, C. 1999, CeMDA, 73, 211, 220

Wachlin, F. & Ferraz-Mello, S. 1998, MNRAS, 298, 22

ORAL COMMUNICATION

Simple Instability in a 3D Autonomous Hamiltonian system of galactic type

M. Katsanikas[1,2]

(1) Research Center for Astronomy, Academy of Athens
Soranou Efessiou 4, GR-11527 Athens, Greece
mkatsan@academyofathens.gr
(2) Section of Astrophysics, Astronomy and Mechanics
Department of Physics, University of Athens, Greece

Abstract.   In this paper we study the orbital behavior in the neighborhood of simple unstable periodic orbits in a 3D rotating galactic potential. We use the method of color and rotation to visualize the 4D spaces of section. We found four types of structures in the 4D space of section that correspond to four types of orbits. The first three types are sticky chaotic orbits and in the last one the orbit visits all available phase space.

1.   Introduction

In this paper we give a short review of recent results from studies about the structure of the phase space in the neighborhood of simple unstable periodic orbits in a 3D autonomous hamiltonian system of galactic type (Katsanikas *et al.* 2011b). In 3D autonomous hamiltonian systems the phase space is six dimensional and the space of section is 4D. For this reason we use the method of color and rotation to visualize this space (Patsis & Zachilas 1994). By using this method we first consider 3D projections and rotate the 3D figures on a computer screen to observe the figure from all its sides. Every point is colored according its value in the 4th dimension. The smooth color variation corresponds to ordered behavior and the mixing of colors to chaotic behavior in the 4th dimensional space.

For the study of stability of periodic orbits we use the method of Broucke (1969) and Hadjidemetriou (1975). According to this method we compute firstly the elements of the monodromy matrix of our 4D Poincaré map. Using the elements of the monodromy matrix, we find four types of periodic orbits according to their stability: (1) stable (2) simple unstable (3) double unstable and (4) complex unstable (Contopoulos $ Magnenat 1985, Contopoulos 2002 p.286). We study here the case of simple instability. In this case the periodic orbits have two complex eigenvalues of the monodromy matrix of our 4D Poincaré map that are on the unit circle and two real eigenvalues that are off the unit circle.

The system that we used for our applications is the one used by Katsanikas $ Patsis (2011) with $\Omega_b = 0.045$ in our units. Our potential in its axisymmetric form can be considered as an approximation of the potential for the Milky Way (Miyamoto $ Nagai 1975).

2.   4D surfaces of section

We examine first the different types of orbital behavior in the neighborhood of simple unstable periodic orbits. In previous works Magnenat (1982) found orbits that are represented with double loops in the 2D projections of the 4D space of section in the neighborhood of simple unstable periodic orbits. Here we study the morphology of these double loops in the 4D space. Besides this type of orbits we found totally four types of orbital behavior in the neighborhood of simple unstable periodic orbits. The first three types are associated with the phenomenon of stickiness (Contopoulos $ Harsoula 2008). In these three types the chaotic orbits are guided by the manifolds close to the regions that are occupied by the rotational tori (Vrahatis *et al.* 1997, Katsanikas and Patsis 2011) associated with the stable periodic orbits existing in the region. All of them are associated with motion close to a transition of a mother family from stability to simple instability or vice versa. The fourth type of orbital behavior (last type) is associated with orbits that visit very fast all available phase space.

The four types are represented in the surface of section from four different types of structures in the 4D space of section:

1. The first type is represented by a double loop with smooth color variation (as in the example of Fig. 1). The smooth color variation means that the 4th dimension supports this 3D topological surface (double loop in the 3D subspace of our 4D space of section) in the 4D space. This means that the double loop is a 4D object. At the intersection of the double loop we can see the meeting of different colors (red with violet). This indicates that this intersection does not exist in the 4D space. The previous structure is encountered for a specific number of intersections. For a larger number of intersections we observe that the consequents form a Θ-like structure with mixing of colors (like the Fig. 2). For even larger number of intersections the consequents leave this structure and form clouds of points that correspond to motion in a chaotic sea.

2. The second type is represented only by a Θ-like structure with mixing of colors (like the Fig. 2) for a number of intersections. For larger number of intersections we have again as in the first type of orbits the formation of clouds of points.

3. The third type of orbits (Fig. 3) is represented by a double loop with smooth color variation as the double loop of the first type of orbits. However, in this case we have a difference. The region of the intersection of the double loop has now the same color (green - Fig. 3). This means that this intersection is a real 4D intersection. This difference reflects properties of the family of periodic orbits under consideration (e.g. lack of a symmetric family with respect to the equatorial plane).

4. The fourth type of orbits correspond to motion in a chaotic sea. The chaotic sea is represented in the 4D surface of section with clouds of points with mixing of colors. The mixing of colors means that the scattered points are not only in the 3D subspaces of the space of section but are scattered in the 4D space. This means that these clouds are 4D.
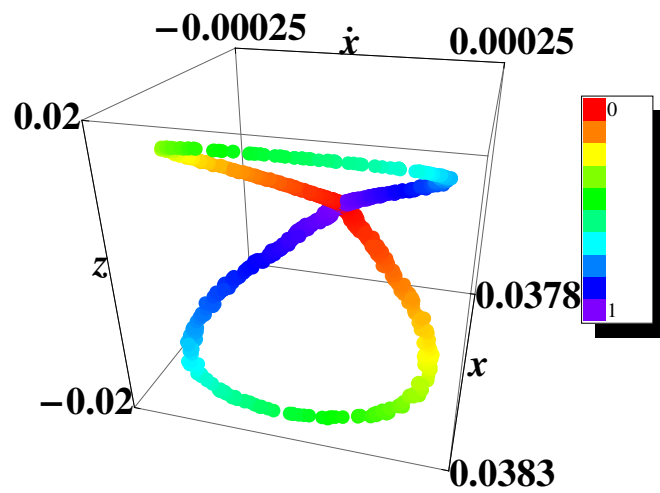
Figure 1. An orbit that is represented by a double loop with no real intersection for 2580 consequents in the 4D space of section. This orbit is in the neighborhood of a simple unstable periodic orbit of x1v2 for E_j=-4.66 (Katsanikas *et al.* 2011b). The consequents are given in the $(x, \dot{x}, z)$ projection, while every point is colored according to its $\dot{z}$ value. We observe that a double loop ribbon is formed with a smooth color succession on it (cf. with the color bar on the right). Our point of view in spherical coordinates is $(\theta, \varphi) = (30^o, 108^o)$.
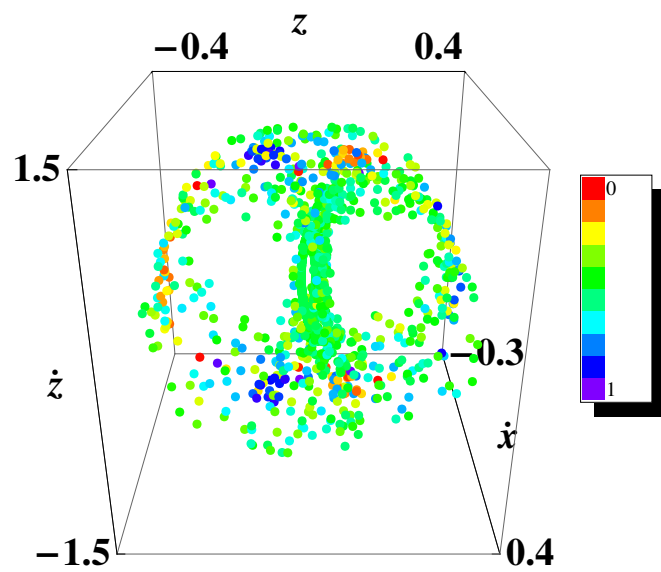
Figure 2.   Mixing of colors for the first 1000 consequents of the orbit in the Fig. 1 (beyond the first 2580 consequents) that deviate from the 4D ribbon surface. The 3580 points totally are depicted in the 3D subspace $(\dot{x}, z, \dot{z})$ and the color represents the 4th dimension x of the points. Our point of view in spherical coordinates is $(\theta, \varphi) = (120^o, 0^o)$.

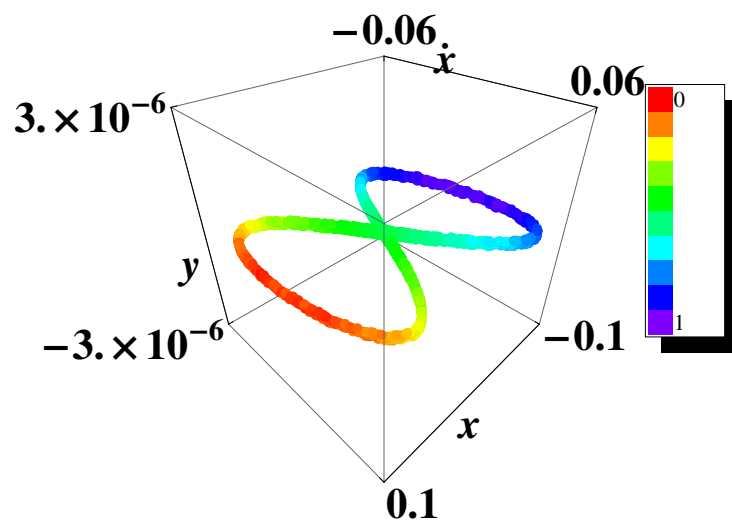Figure 3.   An orbit that is represented by a double loop with real intersection in the 4D space of section. This orbit is in the neighborhood of a simple unstable periodic orbit of z-axis family for a slow pattern model of our system $\Omega_b = 5 \times 10^{-6}$ for E_j=-6.6342 (Katsanikas *et al.* 2011b). Thus, we consider now the $z = 0$ surface as surface of section with $\dot{z} > 0$. The consequents are given in the $(x, \dot{x}, y)$ projection, while every point is colored according to its $\dot{y}$ value. We observe that a double loop ribbon is formed with a smooth color succession on it. Our point of view in spherical coordinates is $(\theta, \varphi) = (60^o, 60^o)$.

The results about the 4D structure of phase space close to a periodic orbit have been presented in a recent series of papers (Katsanikas & Patsis 2011, Katsanikas *et al.* 2011a,b, Katsanikas *et al.* 2011c, Katsanikas 2011).

References

Broucke R. 1969, NASA Tech. Rep. 32-1360, 1

Contopoulos G. 2002, *Order and Chaos in Dynamical Astronomy*, Springer-Verlag, New York Berlin Heidelberg

Contopoulos G., & Magnenat P. 1985, Celest. Mech. & Dynam. Astron., 37, 387

Contopoulos G., & Harsoula M. 2008, Int. J. Bif. Chaos, 18, 2929

Hadjidemetriou J. 1975, Celest. Mech. & Dynam. Astron., 12, 255

Katsanikas M., & Patsis P.A. 2011, Int. J. Bif. Chaos, 21, 467

Katsanikas M., Patsis P.A., & Contopoulos G. 2011a, Int. J. Bif. Chaos, 21, 2321

Katsanikas M., Patsis P.A., & Contopoulos G. 2011b, Int. J. Bif. Chaos (is submitted).

Katsanikas M., Patsis P.A., & Pinotsis A.D. 2011c, Int. J. Bif. Chaos, 21, 2331

Katsanikas M. 2011, Ph.D thesis, University of Athens (in greek - to be submitted).

Magnenat P. 1982, Celest. Mech. & Dynam. Astron., 28, 319

Miyamoto M., & Nagai R. 1975, Publ. Astron. Soc. Japan, 27, 533

Patsis P.A., & Zachilas L. 1994, Int. J. Bif. Chaos, 4, 1399

Vrahatis M.N., Isliker H., & Bountis T.C. 1997, Int. J. Bif. Chaos, 7, 2707

Author Index